Two-Dimensional Surrogate Contact Modeling for Computationally Efficient Dynamic Simulation of Total Knee Replacements

Yi-Chung Lin

Raphael T. Haftka
Ph.D.

Department of Mechanical and Aerospace Engineering,
University of Florida,
231 MAE-A Building,
P.O. Box 116250,
Gainesville, FL 32611-6250

Nestor V. Queipo
Ph.D.
Applied Computing Institute,
University of Zulia,
Maracaibo,
Zulia 4005, Venezuela

Benjamin J. Fregly[1]
Ph.D.
Department of Mechanical and Aerospace Engineering,
Department of Biomedical Engineering,
and Department of Orthopaedics and Rehabilitation,
University of Florida,
231 MAE-A Building,
P.O. Box 116250,
Gainesville, FL 32611-6250
e-mail: fregly@ufl.edu

Computational speed is a major limiting factor for performing design sensitivity and optimization studies of total knee replacements. Much of this limitation arises from extensive geometry calculations required by contact analyses. This study presents a novel surrogate contact modeling approach to address this limitation. The approach involves fitting contact forces from a computationally expensive contact model (e.g., a finite element model) as a function of the relative pose between the contacting bodies. Because contact forces are much more sensitive to displacements in some directions than others, standard surrogate sampling and modeling techniques do not work well, necessitating the development of special techniques for contact problems. We present a computational evaluation and practical application of the approach using dynamic wear simulation of a total knee replacement constrained to planar motion in a Stanmore machine. The sample points needed for surrogate model fitting were generated by an elastic foundation (EF) contact model. For the computational evaluation, we performed nine different dynamic wear simulations with both the surrogate contact model and the EF contact model. In all cases, the surrogate contact model accurately reproduced the contact force, motion, and wear volume results from the EF model, with computation time being reduced from 13 min to 13 s. For the practical application, we performed a series of Monte Carlo analyses to determine the sensitivity of predicted wear volume to Stanmore machine setup issues. Wear volume was highly sensitive to small variations in motion and load inputs, especially femoral flexion angle, but not to small variations in component placements. Computational speed was reduced from an estimated 230 h to 4 h per analysis. Surrogate contact modeling can significantly improve the computational speed of dynamic contact and wear simulations of total knee replacements and is appropriate for use in design sensitivity and optimization studies. [DOI: 10.1115/1.3005152]

Introduction

For more than three decades, total knee replacement (TKR) surgery has been performed on patients with severe osteoarthritis. Though modern TKR surgery has a high success rate, some patients still experience substantial functional impairment postsurgery compared to healthy individuals [1]. Furthermore, patients today desire more than just pain relief, seeking a high level and broad variety of daily activities following TKR surgery (e.g., tennis, hiking, gardening, and even jogging) [2,3]. Thus, maximization of durability and minimization of functional limitations have become two primary design goals.

One way to pursue these goals is through the development of computational technology that permits sensitivity and optimization studies of new TKR designs. A recent step in this direction has been the development of computational models of knee wear simulator machines [4–8]. Such models permit dynamic contact and wear simulations of knee implant designs to be performed in a time- and cost-efficient manner, with model predictions being validated against experimental motion and wear measurements [4,6,8,9]. Though finite element models can be used to perform contact calculations for these simulations, recent work has shown that elastic foundation contact models produce similar contact forces and pressures in a fraction of the CPU time [10]. Even so, CPU times on the order of 5–10 min for a one-cycle dynamic contact simulation can still be prohibitive for design sensitivity and optimization studies requiring thousands of simulations.

Computational limitations in other engineering disciplines have been overcome through the use of surrogate modeling approaches. Surrogate modeling (also known as "metamodeling" or "response surface approximation") involves replacing a computationally costly model with a computationally cheap model constructed using data points sampled from the original model. Once the surrogate model is constructed, it is used in place of the original computationally costly model when subsequent engineering analyses are performed. A variety of surrogate model fitting methods have been proposed, including polynomial response surfaces [11–13], Kriging [14,15], radial basis functions [16,17], splines [18], and support vector machines [19,20]. While surrogate models have been utilized successfully for a variety of engineering applications [21–26], only a few studies have used surrogate models to fit input-output relationships from a computational contact model [27–29]. Furthermore, none of these studies has proposed a surrogate modeling technique to improve the computational efficiency of dynamic contact simulations.

Unlike traditional engineering applications, joint contact analyses pose unique challenges to existing surrogate modeling ap-

[1]Corresponding author.

Contributed by the Bioengineering Division of ASME for publication of the JOURNAL OF BIOMECHANICAL ENGINEERING. Manuscript received November 19, 2007; final manuscript received August 22, 2008; published online February 4, 2009. Review conducted by Noshir Langrana.

 Copyright © 2009 by ASME

proaches. Sampling data points within a hypercube defined by the upper and lower bounds of the relative pose parameters will result in few physically realistic data points for surrogate model creation. Most sample points will be either out of contact (i.e., zero contact force) or in excessive penetration (i.e., unrealistically high contact force) since contact forces are highly sensitive to the small displacements in the contact normal direction. Consequently, traditional surrogate-based modeling approaches fail to provide the accuracy needed to perform dynamic simulations. A special surrogate-based modeling approach adapted to the specific needs of joint contact analyses is therefore required.

This study proposes a novel surrogate contact modeling approach for performing computationally efficient dynamic contact and wear simulations of total knee replacements. The approach addresses the unique challenges involved in applying surrogate modeling techniques to joint contact problems. The primary objectives of the study are twofold. The first is to evaluate the computational speed and accuracy of the proposed surrogate contact modeling approach by comparing its results with those generated by an elastic foundation contact model. Sample points for the surrogate contact model are generated by the same elastic foundation contact model used in the comparative dynamic wear simulations. The second objective is to demonstrate the practical applicability of the proposed approach by analyzing the sensitivity of knee replacement wear predictions to realistic variations in input motions and loads and component placements. A Monte Carlo approach requiring thousands of dynamic wear simulations is used to perform the sensitivity study. Both objectives are pursued using a dynamic contact model of a total knee replacement constrained to planar motion in a Stanmore machine, and both demonstrate the significant computational benefits that surrogate contact modeling can provide for the analysis of TKR designs.

Methods

Surrogate Contact Model Development. We have developed a novel surrogate modeling approach to replace computationally costly contact calculations in dynamic simulations of total knee replacements. To develop the approach, we utilized a three-dimensional (3D) elastic foundation contact model [30–34] of a cruciate-retaining commercial knee implant (Osteonics 7000, Stryker Howmedica Osteonics, Inc., Allendale, NJ) constrained to sagittal plane motion in a Stanmore simulator machine [35]. The model possessed three degrees of freedom (DOFs) relative to ground: tibial anterior-posterior translation, femoral superior-inferior translation, and femoral flexion-extension. Similar to a Stanmore machine, femoral flexion-extension rotation was prescribed while the remaining two DOFs were loaded using ISO standard input curves [36]. A pair of parallel spring bumper loads was also applied to the tibia in the anterior-posterior direction to approximate the effect of ligament forces [35]. The EF contact model was treated as an applied load on the femoral component and tibial insert and used linear elastic material properties [4]. The dynamic equations for the model were derived via Kane's method using AUTOLEV symbolic manipulation software (OnLine Dynamics, Sunnyvale, CA). The equations were incorporated into a MATLAB program (The Mathworks, Natick, MA) that was used to perform forward dynamic simulations with the stiff solver ode15s.

To develop a surrogate contact model to replace the EF contact model in the larger dynamic model described above, we followed a four-step process: (1) design of experiments (DOE), (2) computational experiments, (3) surrogate model selection, and (4) surrogate model implementation (Fig. 1). The first step, DOE, is a statistical method for determining at which locations in design space computational experiments should be performed with the computationally costly model to sample the outputs of interest (Fig. 1(*a*)). For our sagittal plane knee implant model, locations in design space (or "sample points") were defined by the following three relative pose parameters: tibial anterior-posterior (AP) translation X, femoral superior-inferior (SI) translation Y, and femoral flexion-extension rotation θ. The associated outputs of interest calculated by the EF contact model were the AP contact force F_x and the SI contact force F_y. Calculated contact torque was not of interest since θ was prescribed. To minimize the total number of sample points while maximizing the quality of the resulting surrogate model fit, we chose a Hammersley quasirandom (HQ) sampling method [37]. This method uses an optimal design scheme for placing sample points within a multidimensional hypercube and provides better uniformity properties than do other sampling techniques [38,39].

The second step, computational experiments, involved performing simulations with the original contact model at the sample points defined in the first step. Because contact forces are highly sensitive to some pose parameters but not others, we made one modification to the computational experiments performed at the HQ sample points (Fig. 1(*b*)). A traditional HQ sampling method would sample values of the three inputs X, Y, and θ within their specified bounds and use the EF contact model to calculate the two outputs F_x and F_y at each sample point. However, since SI contact force F_y is highly sensitive to small variations in SI translation Y [29,40], this approach produced a large number of sample points that were either out of contact or deeply penetrating. Inclusion of these unrealistic points in the surrogate model fitting process reduced the accuracy of the resulting surrogate contact model. To resolve this issue, we inverted the HQ sampling method for superior-inferior translation by sampling forces in the sensitive direction (i.e., F_y) and displacements in the insensitive directions (i.e., X and θ). Specifically, we used the HQ method to sample 300 pairs of X and θ values. For each of these pairs, we generated five sample points by performing five static analyses with the EF contact model using F_y values ranging from 600 N to 2400 N (approximately three times body weight) in 600 N increments plus a 10 N value for determining contact initiation. The outputs of each static analysis were F_x and Y. Evaluating 1500 sample points (X, F_y, and θ) using static analyses required approximately 3 h of CPU time on a 3.4 GHz Pentium IV PC.

The third step, surrogate model selection, involved determining the most appropriate type of surrogate model for fitting the input-output relationships defined by the computational experiments in the second step. Based on its ability to interpolate multidimensional nonuniformly spaced sample points, Kriging [15,41] was selected as the surrogate modeling method to fit the input-output relationships defined by the 1500 sample points. Originally developed for geostatistics and spatial statistics, Kriging has been widely applied in many different fields [28,42]. In mathematical terms, Kriging utilizes a combination of a polynomial trend model and a systematic departure model as follows:

$$y(\mathbf{x}) = f(\mathbf{x}) + z(\mathbf{x}) \tag{1}$$

where $\mathbf{x}$ is a vector of input variables, $y(\mathbf{x})$ is the output to be fitted, $f(\mathbf{x})$ is a polynomial trend model, and $z(\mathbf{x})$ is a systematic departure model whose correlation structure is a function of the distances between sample points. The $f(\mathbf{x})$ term approximates the design space globally while the $z(\mathbf{x})$ term interpolates local deviations from $f(\mathbf{x})$. The necessary parameters for $z(\mathbf{x})$ were obtained through maximum likelihood estimation [15]. The MATLAB toolbox DACE [43] was used to construct all necessary Kriging models using a cubic polynomial trend function and a cubic spline correlation function based on preliminary tests.

The fourth step, surrogate model implementation, required the development of a special model inversion process to accommodate our modified HQ sampling approach (Fig. 1(*c*)). During a dynamic simulation, the EF contact model requires X, Y, and θ as inputs and provides F_x and F_y as outputs, whereas the surrogate contact model requires X, F_y, and θ as inputs and predicts F_x and Y as outputs. To address this inconsistency, we fitted F_y as a function of Y using the general form of a Hertzian contact model [44]

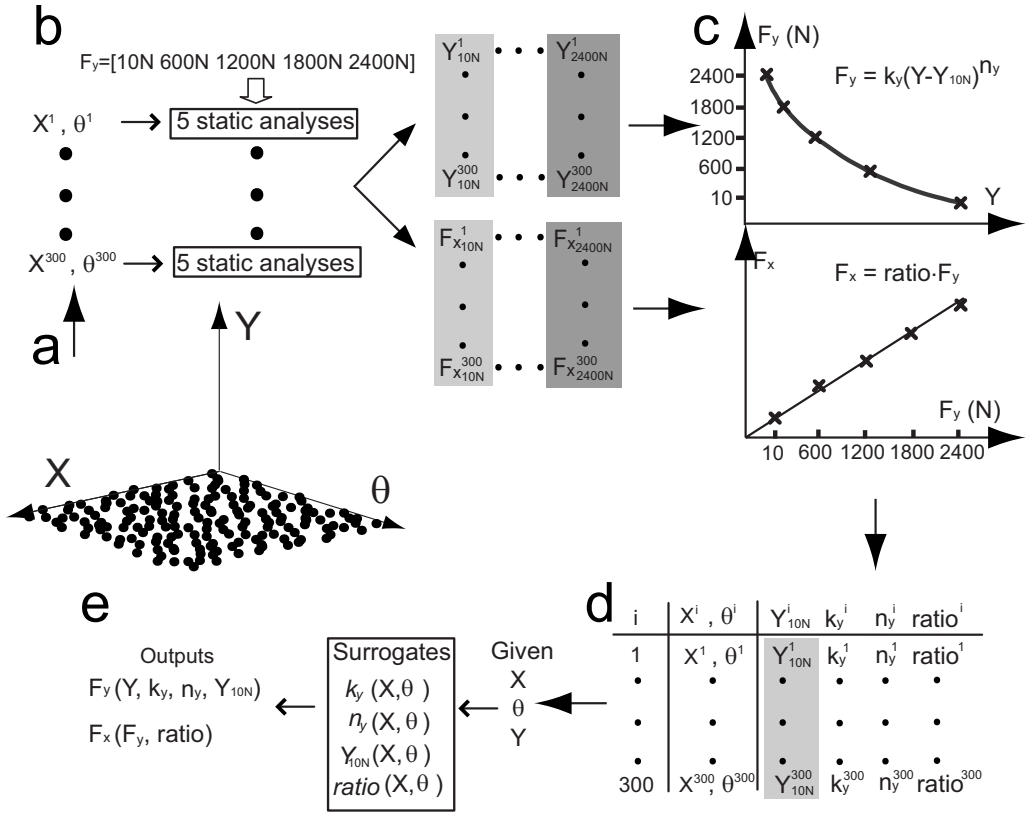

Fig. 1 Overview of surrogate contact model creation and implementation process. (*a*) 300 points (black dots) are sampled in the (X, θ) design space. (*b*) Five static analyses using five axial loads are performed at each sample point location using an elastic foundation contact model to calculate Y and F_x. (*c*) Generalized Hertzian contact theory is used to model F_y as a function of Y while F_x is modeled as proportional to F_y. (*d*) The values of $Y_{10\,\mathrm{N}}$, k_y, n_y, and ratio from the 300 sample points are fitted as functions of X and θ using Kriging. (*e*) During a dynamic simulation, the surrogate contact model calculates F_y and F_x from four Kriging models given the current values of X, Y, and θ.

$$\hat{F}_y = k_y(X,\theta)[Y_{10\,\mathrm{N}}(X,\theta) - Y]^{n_y(X,\theta)} \tag{2}$$

where $\hat{F}_y$ is the predicted SI contact force, k_y is the contact stiffness, $Y_{10\,\mathrm{N}}$ is the SI translation at contact initiation, and n_y is the contact exponent. Use of Eq. (2) to calculate $\hat{F}_y$ required fitting k_y, $Y_{10\,\mathrm{N}}$, and n_y as a function of X and θ. Each of the 300 (X, θ) sample pairs had five sampled values of F_y (i.e., 10 N, 600 N, 1200 N, 1800 N, and 2400 N) and five associated output values of Y. Thus, $Y_{10\,\mathrm{N}}$ was already known for each (X, θ) sample pair. To find k_y and n_y for each (X, θ) sample pair, we linearized Eq. (2) by taking $\log_{10}$ of both sides and solved for $\log_{10}(k_y)$ (and hence k_y) and n_y using linear least squares (five equations in two unknowns). This process yielded one value of k_y, $Y_{10\,\mathrm{N}}$, and n_y for each (X, θ) sample pair.

To calculate F_x, we used the observation that F_x is close to a constant fraction of F_y at each (X, θ) sample pair as follows:

$$\hat{F}_x = \mathrm{ratio}(X,\theta)\hat{F}_y \tag{3}$$

where $\hat{F}_x$ is the predicted AP contact force, "ratio" is the average of the five (F_x, F_y) quotients obtained for each sample pair, and $\hat{F}_y$ is the SI contact force predicted by Eq. (2). Thus, the final surrogate contact model utilized four separate Kriging models to fit k_y, $Y_{10\,\mathrm{N}}$, n_y, and ratio as a function of X and θ (Fig. 1(*d*)), permitting the calculation of $\hat{F}_y = f(X, Y, \theta)$ and $\hat{F}_x = f(X, \theta, \hat{F}_y)$ from Eqs. (2) and (3) at any point during a dynamic simulation (Fig. 1(*e*)).

Surrogate Contact Model Evaluation. To evaluate the computational speed and accuracy of our surrogate contact modeling approach, we performed nine dynamic wear simulations with both the surrogate contact model and the EF contact model used to generate it. Each wear simulation utilized different nominal component placements that were variations away from the original placements (Fig. 2(*a*)). The specific variations simulated were all possible combinations of three SI locations of the femoral flexion axis (original$\pm$10 mm) and three AP slopes of the tibial insert (original$\pm$10 deg). To cover all possible dynamic simulation paths, we specified the ranges of the 300 HQ sampling points using the limits of motion obtained from dynamic simulations performed with the EF contact model when the SI location of the femoral flexion axis and the AP slope of the tibial insert were set to their extreme values. We quantified differences between the sets of results by calculating root mean square errors (RMSEs) in predicted contact forces, component translations, and wear volumes.

Generation of wear volume predictions with the surrogate contact model required the development of a separate surrogate modeling approach based on prediction of medial and lateral center of pressure (CoP) locations. This approach used Archard's wear law [45] to predict wear volume for any specified number of cycles from the time history of contact forces and CoP slip velocities on each side over a one-cycle dynamic simulation. Specifically,

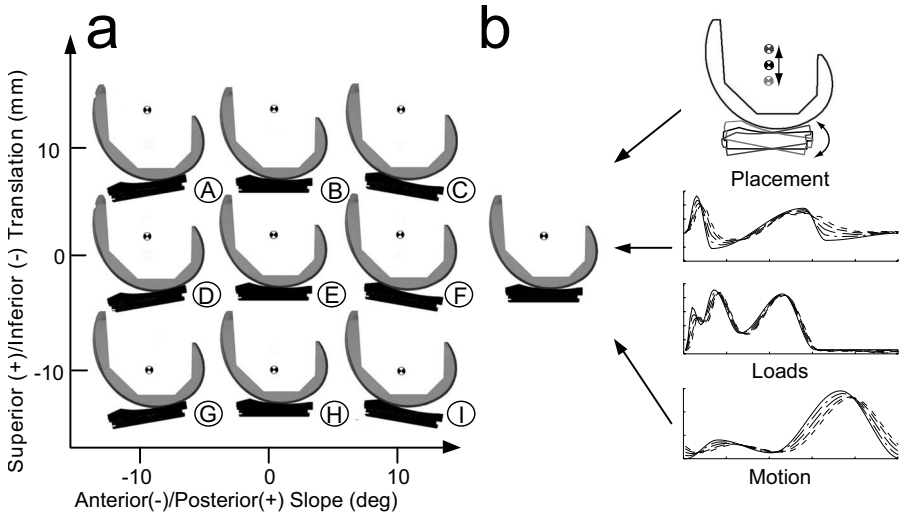

Fig. 2 Overview of global and local sensitivity analyses of predicted wear volume performed with the surrogate contact model. (*a*) Large variations in nominal component placements (labeled A–I) used for the global sensitivity analysis. (*b*) Small variations in component placements, input loads, and input motion used for local sensitivity analysis involving Monte Carlo methods applied to each nominal component placement.

$$V = N\sum_{i=1}^{n}\sum_{j=1}^{2} kF_{ij}d_{ij} = Nk\sum_{i=1}^{n}\sum_{j=1}^{2} F_{ij}|\mathbf{v}_{ij}|\Delta t \tag{4}$$

where V is the total wear volume of the medial and lateral sides over N cycles, N is the assumed total number of cycles (5×10^6), i is a discrete time instant during the one-cycle dynamic simulation (1 through n), j is the side (medial or lateral), k is the assumed material wear rate (1×10^{-7} mm^3/N m [46]), F_{ij} is the contact force magnitude at time instant i on side j, and d_{ij} is the sliding distance of the CoP at time instant i on side j. Due to symmetry of the implant geometry, the predicted value of F_{ij} on each side was taken as half the total contact force obtained from Eqs. (2) and (3). Prediction of d_{ij} on each side required determining the magnitude of the slip velocity $\mathbf{v}_{ij}$ of the CoP and multiplying by the time increment Δt [4]. To find $\mathbf{v}_{ij}$, we used the kinematic concept of two points fixed on a rigid body [47]

$$\mathbf{v}_{ij} = {}^{A}\mathbf{v}^{O} + {}^{A}\boldsymbol{\omega}^{B} \times \mathbf{p}^{O\mathrm{CoP}} \tag{5}$$

In this equation, A represents the tibial insert, B represents the femoral component, O is the femoral coordinate system origin located at the intersection of the flexion-extension and superior-inferior translation axes, CoP is the medial or lateral center of pressure location treated as a point fixed in body B, ${}^{A}\mathbf{v}^{O}$ is the velocity of point O with respect to reference frame A as determined from the relative translation of the femoral component origin, ${}^{A}\boldsymbol{\omega}^{\beta}$ is the angular velocity of body B with respect to reference frame A as determined from the flexion-extension of the femoral component, and $\mathbf{p}^{O\mathrm{CoP}}$ is the position vector from point O to point CoP. While the values of ${}^{A}\mathbf{v}^{O}$ and ${}^{A}\boldsymbol{\omega}^{B}$ were obtained directly from the dynamic simulations, calculation of the CoP location on each side required six additional Kriging models (i.e., three components per side). Each Kriging model was created by fitting the average of the five CoP locations obtained from each (X,θ) sample pair. Wear volumes calculated by the surrogate contact model using the CoP-based approach were compared with those calculated by the EF contact model using both the CoP-based approach and a previously published element-based approach [4,8] to verify that the surrogate model wear volume predictions were accurate.

Surrogate Contact Model Application. As a practical application of the proposed surrogate contact modeling approach, we performed two types of sensitivity analyses to investigate how machine setup issues affect predicted wear volume. The first was a global sensitivity analysis to investigate the effect of large variations in component placements within the ranges defined in Fig. 2(*a*) (i.e., ±10 mm/deg). Specifically, we generated 400 combinations of component placements using 20 evenly spaced values of femoral flexion axis SI location and tibial insert AP slope. For each combination, the surrogate-based contact model was used to perform a dynamic simulation and predict wear volume. Percent changes in predicted wear volume were calculated relative to the maximum predicted value from all global sensitivity analysis results. The second category was a local sensitivity analysis to investigate the effect of small variations in nominal component placements and motion and load inputs by using Monte Carlo techniques (Fig. 2(*b*)). For the local sensitivity analysis, allowable variations in motion and load input curves were defined using experimental observations from eight different implant designs tested in a Stanmore simulator machine [36]. Deviations of the experimental input curves from the ISO standard curves were calculated and analyzed using principal component analysis [48], with the first two principal components being sufficient to account for over 98% of the variability in each type of curve. These two components were used to generate new deviation curves by selecting component weights as uniformly distributed random numbers within the bounds of the experimental curves. These deviation curves were added to the ISO standard curves to create new input motion and load curves as needed for the Monte Carlo analyses.

To distinguish between the effects of small errors in motion and load inputs and small errors in component placements, the local sensitivity analysis performed four types of Monte Carlo analyses for each of the nine nominal component placements for a total of 36 Monte Carlo analyses in all. Each type of Monte Carlo analysis varied input profiles and component placements separately or together. The first type varied motion and load profiles within 100% of their allowable ranges and also varied component placements within ±1 mm/deg. The second type was similar but used smaller variations of only 10% of the allowable range for motion and load profiles and ±0.1 mm/deg for component placements. The third type varied only motion and load profiles within 100% of their allowable ranges and imposed no variations on component placements. Finally, the fourth type varied component placements within ±1 mm/deg but imposed no variations on motion and load

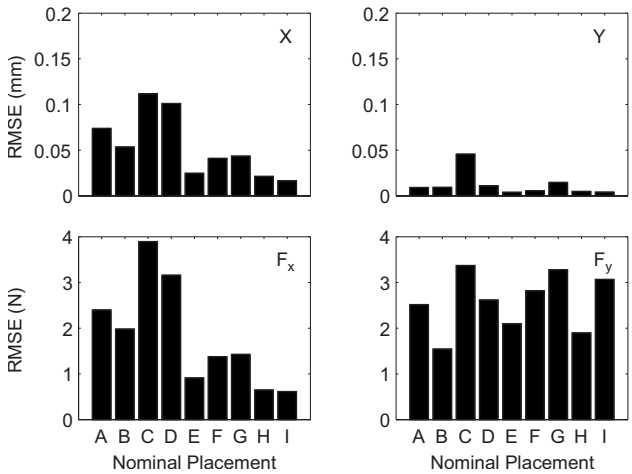

Fig. 3 Root-mean-square errors in joint motions and contact forces for each nominal component placement. Errors are computed by comparing dynamic simulation results generated with the elastic foundation contact model and the surrogate contact model.

Table 1 Comparison between wear volume predictions (mm^3) made with the elastic foundation contact model and the surrogate contact model for each of the nine nominal component placements. Each prediction was made for 5×10^6 motion cycles. Both contact models can predict wear volume using a simple center of pressure-based approach, while only the elastic foundation contact model can make predictions using a more detailed element-based approach.

Nominal placement	Elastic foundation contact model		Surrogate contact model
	Element-based	CoP-based	CoP-based
A	37.07	37.05	36.98
B	34.03	34.02	33.91
C	39.04	39.03	38.87
D	32.91	32.89	32.88
E	30.78	30.76	30.77
F	34.20	34.19	34.14
G	29.37	29.36	29.32
H	26.45	26.44	26.38
I	29.37	29.35	29.31

profiles. For all types, percent changes in predicted wear volume were calculated relative to values obtained from the nominal component placements using the ISO standard inputs. Each Monte Carlo analysis was performed on a 3.4 GHz Pentium IV PC and involved at least 1000 dynamic contact simulations. The convergence criterion for each analysis was met when the mean and coefficient of variance (i.e., 100 × standard deviation/mean) for the last 10% of the wear predictions were within 2% of the final mean and coefficient of variance [49].

Results

For the computational evaluation, the dynamic simulations performed with the surrogate contact model closely reproduced the planar motions, contact forces, and wear volumes predicted with the EF contact model for all nine component placements. On average, the RMSEs for planar motions and contact forces were less than 0.125 mm and 4 N, respectively (Fig. 3). While each dynamic simulation performed with the EF contact model required approximately 13 min of CPU time, each simulation performed using the surrogate contact model required 13 s. For both the surrogate contact model and the EF contact model, the CoP-based approach for predicting wear volumes reproduced the element-based EF results to within 0.5% error (Table 1).

For the practical application, the global and local sensitivity analyses revealed that large variations in component placements relative to the original locations and small variations in input motion relative to the ISO standard had a significant affect on predicted wear volume. For the global sensitivity analysis, predicted wear volume was sensitive to large variations in the SI location of the femoral flexion axis (maximum change of 26% when the AP slope of tibial insert held fixed), the AP slope of the tibial insert (maximum change of 13% when the SI location of femoral flexion axis held fixed), or both parameters simultaneously (maximum change of 33%) (Fig. 4). For the local sensitivity analysis, wear volume ranges were larger (mean of 18 mm^3 compared to 2 mm^3) and 50% percentile wear volume results were lower (mean of 23 mm^3 compared to 31 mm^3) when motion and load inputs and component placements were varied within 100% rather than 10% of their allowable ranges (Figs. 5(*a*) and 5(*b*)). When inputs and placements were varied separately within 100% of their allowable ranges, predicted wear volume was much more sensitive to motion and load inputs than to component placements (maximum change of 64% compared to 5%; Figs. 5(*c*) and 5(*d*)). Finally, when an additional global sensitivity analysis was performed with the original component placements to separate the influence of the different inputs, predicted wear volume varied approximately linearly with small changes in each input curve but was more sensitive to changes in flexion-extension motion (maximum change of 47%; Fig. 6) than to changes in input loads (maximum change of 20% for F_x and 14% for F_y; Figs. 6(*a*) and 6(*b*)). Each Monte Carlo analysis performed with the surrogate contact model required 4 h of CPU time compared to an estimated 230 h with the EF contact model.

Discussion

Computational speed is a major limiting factor for performing sensitivity and optimization studies of total knee replacement designs. Unlike constraint-based joint models, knee joint contact models require repeated geometry evaluations that consume the vast majority of the CPU time in a dynamic contact simulation [50]. This paper has presented a novel surrogate modeling approach for performing efficient dynamic contact simulations of human joints. The surrogate contact model was developed to replace the original EF contact model to avoid time-consuming repeated evaluation of the surface geometry during dynamic simu-

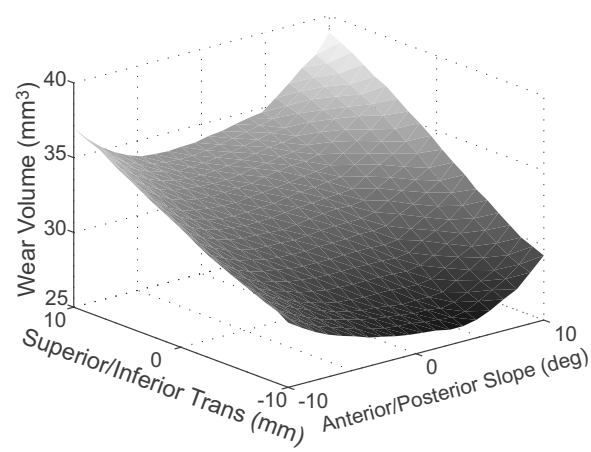

Fig. 4 Predicted wear volume as a function of large variations in component placements as calculated from 400 dynamic simulations performed with the surrogate contact model. Wear increases linearly when the femoral flexion axis is translated superiorly and quadratically when the tibial slope is increased anteriorly or posteriorly.

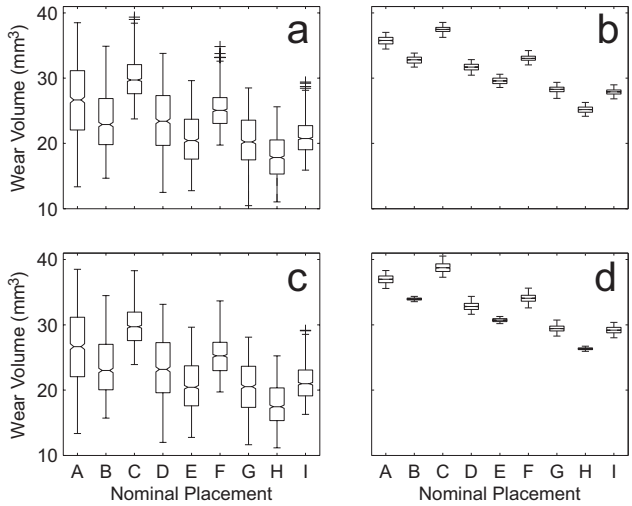

Fig. 5 Box plot distribution of predicted wear volume generated by four Monte Carlo analyses using the surrogate contact model. Each box has (from bottom to top) one whisker at the 10th percentile, three lines at the 25th, 50th, and 75th percentile, and another whisker at the 90th percentile. Outliers are indicated by black crosses located beyond the ends of the whiskers. For the first and second Monte Carlo analyses, motion and load inputs and component placements are varied together within (*a*) 100% or (*b*) 10% of their maximum ranges, respectively. For the third and fourth Monte Carlo analyses, (*c*) motion and load inputs or (*d*) component placements are varied separately within 100% of their maximum ranges, respectively.

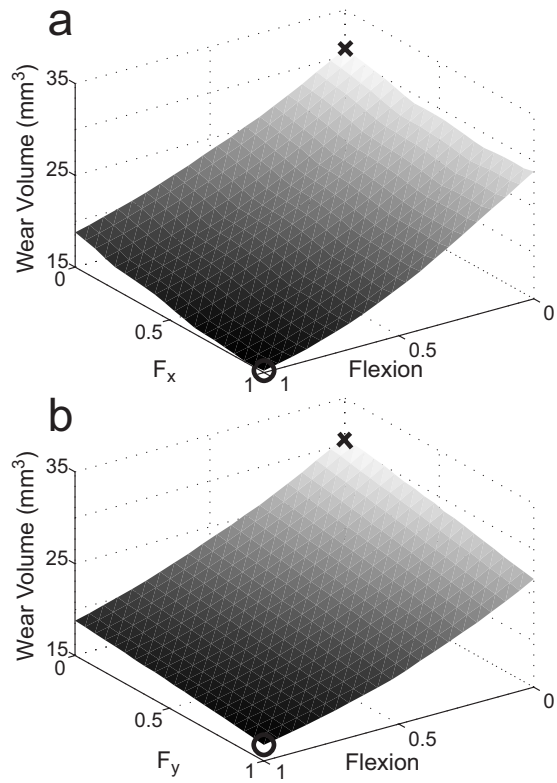

Fig. 6 Predicted wear volume as a function of variations in (*a*) anterior-posterior load and femoral flexion angle and (*b*) axial load and femoral flexion angle. Both plots were created by performing dynamic simulations with the surrogate model and changing the weight of the first principal component for each input profile variation away from the ISO standard. Each weight was normalized to be between 0 and 1. The cross marker (×) represents the ISO standard input with no superimposed input profile variations (i.e., weights of 0) while the circle marker (○) represents the ISO standard input with maximum superimposed input profile variations (i.e., weights of 1).

lation. The surrogate contact model produced dynamic simulation and wear results that accurately matched those from the EF contact model but with an order of magnitude improvement in computational speed. This speed improvement would have been even more significant if a finite element contact model had been the original model. The 36 Monte Carlo analyses demonstrated how analyses that would have been impractical (or at a minimum would have required extensive parallel processing) can be easily achieved on a single computer with the surrogate contact model. The sensitivity results also have practical value, suggesting that wear volume generated by simulator machines may be sensitive to large variations in component placements within the machine as well as small variations in the flexion-extension motion input to the machine.

The primary benefit of using surrogate contact models in dynamic simulations is improved computational efficiency. For elastic foundation or finite element contact models, the computation time per dynamic simulation is largely determined by the number of deformable contacts in the model. Thus, a musculoskeletal computer model utilizing deformable contact models for multiple joints could easily require hours or days of CPU time to complete a single dynamic simulation. Repeated dynamic simulations as part of a sensitivity or optimization study would become impractical. By using surrogate contact models instead, one could eliminate the computational cost of the contact solver as the limiting factor. The larger the number of deformable contacts in a multibody dynamic model, the greater the anticipated benefit from using surrogate contact models.

While we have shown that surrogate contact models can be beneficial for sensitivity studies, the greatest computational benefit is likely to occur for optimization studies. For stochastic sensitivity studies, statistically based alternatives to Monte Carlo analyses exist that require fewer repeated simulations. For example, Laz et al. [51] used the mean value method [52] to perform probabilistic elastic foundation simulations of a TKR design and reduced the number of simulations by a factor of 4 compared to a traditional Monte Carlo approach. Since the CPU time per simulation was 6 min, the use of a surrogate contact model could still reduce the overall computation time by an additional factor of 28, assuming we could achieve comparable speed improvements for the 3D situation. In contrast, no good methods exist for reducing the number of simulations in optimization studies. For those situations, the factor of 60 reduction in computation time (i.e., 13 min−13 s) could mean the difference between an impossible and an achievable optimization study.

Other than computational efficiency, surrogate contact models possess the additional benefits of being adaptable and flexible. First, the surrogate-based contact modeling approach presented here is not limited to the knee joint. Theoretically, the approach could be applied to any joint contact model whose input-output relationships can be sampled. The modified HQ sampling method should be suitable for any joint contact model once the sensitive directions are identified. Therefore, neither type of joint (e.g., hip or knee or ankle) nor type of contact model (e.g., EF or finite element) nor type of material model (e.g., linear or nonlinear) is critical for developing a surrogate contact model. The benefit of using an EF contact model to generate the required sample points was that it is relatively fast computationally for performing repeated static analyses. Finite element models would require more time for each static analysis, plus the inputs are usually applied displacements rather than loads. For the finite element situation, it could be necessary to curve fit the force-displacement curve at each sample point to produce new points at fixed force incre-

ments. Second, the surrogate contact model represents the real surface geometry implicitly within the design space. Since the surrogate modeling approach uses the real surface geometry to generate the sampling points, no idealization of complex surface geometry is required. Third, the approach is not limited to a single surrogate model fitting method. While we used Kriging in the present study, any of the fitting methods mentioned previously could be investigated as well [24,26,53,54]. Though polynominal response surfaces are the most commonly used surrogate model, Kriging has the advantage of interpolating the sample point outputs, which is appealing for deterministic computer models [15,54].

The global sensitivity results demonstrate the importance of developing a standardized method for defining the location of the femoral flexion axis. For a given tibial slope, predicted wear volume varied approximately linearly with changes in the superior-inferior location of the femoral flexion axis, increasing as the axis was translated superiorly (Fig. 4). Superior translation of the flexion axis increases the distance to the CoP, thereby increasing the magnitude of the slip velocity in Eq. (4). Thus, to minimize wear for any given implant design, one can simply locate the femoral flexion axis close to the articular surfaces. In contrast, the predicted wear volume varied approximately quadratically with changes in tibial slope, increasing as the insert was tilted anteriorly or posteriorly away from its neutral position (Fig. 4). Increasing the tibial anterior or posterior slope increases the AP tibial sliding distance [55], thereby leading to an increase in predicted wear volume [56,57].

In contrast, the local sensitivity results demonstrate the importance of controlling machine inputs closely, especially the femoral flexion angle. The Monte Carlo analyses and the additional global sensitivity analysis revealed when small variations were imposed on the input motion, loads, and component placements, only variations in the femoral flexion angle resulted in large changes in predicted wear volume. When the flexion angle was varied within 100% of its allowable range, the drop in 50th percentile wear volume results (Figs. 5(*a*) and 5(*c*)) relative to the ISO standard inputs (i.e., 50% percentile results in Figs. 5(*b*) and 5(*d*)) was due to the use of imposed variations that always reduced the amplitude of the curve [36]. This finding is consistent with previous studies showing that predicted wear volume decreases when the amplitude of the input curves is reduced [58–60]. Thus, another way to minimize wear for any given implant design is to use a control system that systematically undershoots the peaks in the input flexion-extension curve. Combining this approach with an inferiorly located femoral flexion axis could cut wear volume by more than 50% compared to the original component placement with ISO standard inputs. The development of a standard for acceptable variations in input flexion-extension motion should be considered as well.

Despite its computational advantages, the proposed surrogate contact modeling approach possesses several important limitations. The biggest one is that new surrogate contact models must be generated if implant geometry, insert thickness, or material properties are changed, at least for the current formulation. The geometry limitation makes it impossible to use surrogate contact models for progressive wear simulations, where the surface geometry is changed gradually over an iterative sequence of wear simulations. However, this limitation is not serious if only wear volume is of interest, as recent progressive wear simulation studies have reported that predicted wear volume (but not wear area or depth) is relatively insensitive to whether or not the surface geometry is changed gradually [8,9]. Theoretically, changes in insert thickness or material model parameter values could be accommodated by adding more input variables to the surrogate model fitting process (e.g., make insert thickness an additional surrogate model input). For contact forces, a changing value of Young's modulus (E) in a linear elastic material model can be accommodated without adding another input variable. If all sample points are generated using a value of $E=1$, then the contact force output by the surrogate model can be scaled by the desired value of E [29].

A number of other limitations exist as well. First, the current surrogate contact model is restricted to sagittal plane motion, where contact force is dominated by SI translation of the femoral component. For 3D simulations, contact force will be highly sensitive to changes in varus-valgus rotation as well [40]. In addition to an increased number of sensitive directions, the number of sample points will also need to increase. However, the increased computational cost for evaluating more sample points is paid up front since the sampling process is only performed once. Second, the calculation of F_x is dependent on the value of F_y computed from Eq. (2). Thus, the quality of the F_y prediction will directly impact the F_x calculation. The current approach works fine for sagittal plane models with a single sensitive direction, but an independent calculation of F_x may be needed for 3D situations where more than one sensitive direction is present. Third, the CoP approach for calculating wear volume may not work well for some types of 3D motion. For example, for a sphere spinning about a vertical axis while in contact with a plane, the CoP would be located at the bottom of the sphere where the slip velocity is zero. Thus, no wear volume would be calculated, even though the element-based approach would calculate a nonzero wear volume based on the pressures and slip velocities throughout the contact area. Consequently, the CoP approach will work well for analysis of TKR designs undergoing 3D motion only if the amount of internal-external rotation is "small." Fourth, our Monte Carlo results are limited to cruciate-retaining designs with a nonconformal tibial insert. Whether or not our conclusions are generalizable to other implant designs and simulator machines would require further investigation.

In summary, this paper has demonstrated that surrogate contact modeling can significantly improve the computational speed of dynamic contact simulations and is appropriate for 2D sensitivity and optimization studies incorporating deformable contact. Refinement of the current 2D approach and extension to 3D problems are topics of ongoing research. Once surrogate contact models are developed for 3D situations, it should be possible to create dynamic musculoskeletal models with multiple deformable joints that can be simulated in a short amount of CPU time. Such models should improve our ability to analyze new knee implant designs rapidly and to predict muscle forces across joints where contact forces may have an important stabilizing effect.

Acknowledgment

This work was supported by an NSF CBET grant to B.J.F. and R.T.H.

Appendix

Below we provide links to the MATLAB software tools used in this study to develop surrogate contact models.
Hammersley Quasirandom Sampling:
http://people.scs.fsu.edu/~burkardt/m_src/hammersley/hammersley.html
DACE Kriging Toolbox:
http://www2.imm.dtu.dk/~hbn/dace/

References

[1] Noble, P. C., Gordon, M. J., Weiss, J. M., Reddix, R. N., Conditt, M. A., and Mathis, K. B., 2005, "Does Total Knee Replacement Restore Normal Knee Function?," Clin. Orthop. Relat. Res., **431**, pp. 157–165.
[2] Mancuso, C. A., Sculco, T. P., Wickiewicz, T. L., Jones, E. C., Robbins, L., Warren, R. F., and Williams-Russo, P., 2001, "Patients' Expectations of Knee Surgery," J. Bone Jt. Surg., Am. Vol., **83A**(7), pp. 1005–1012.
[3] Weiss, J. M., Noble, P. C., Conditt, M. A., Kohl, H. W., Roberts, S., Cook, K. F., Gordon, M. J., and Mathis, K. B., 2002, "What Functional Activities Are Important to Patients With Knee Replacements?," Clin. Orthop. Relat. Res., **40**(4), pp. 172–188.
[4] Fregly, B. J., Sawyer, W. G., Harman, M. K., and Banks, S. A., 2005, "Com-

putational Wear Prediction of a Total Knee Replacement From In Vivo Kinematics," J. Biomech., **38**(2), pp. 305–314.
[5] Giddings, V. L., Kurtz, S. M., and Edidin, A. A., 2001, "Total Knee Replacement Polyethylene Stresses During Loading in a Knee Simulator," ASME J. Tribol., **123**(4), pp. 842–847.
[6] Halloran, J. P., Petrella, A. J., and Rullkoetter, P. J., 2005, "Explicit Finite Element Modeling of Total Knee Replacement Mechanics," J. Biomech., **38**(2), pp. 323–331.
[7] Rawlinson, J. J., Furman, B. D., Li, S., Wright, T. M., and Bartel, D. L., 2006, "Retrieval, Experimental, and Computational Assessment of the Performance of Total Knee Replacements," J. Orthop. Res., **24**(7), pp. 1384–1394.
[8] Zhao, D., Sakoda, H., Sawyer, W. G., Banks, S. A., and Fregly, B. J., 2008, "Predicting Knee Replacement Damage in a Simulator Machine Using a Computational Model With a Consistent Wear Factor," ASME J. Biomech. Eng., **130**(1), p. 011004.
[9] Knight, L. A., Pal, S., Coleman, J. C., Bronson, F., Haider, H., Levine, D. L., Taylor, M., and Rullkoetter, P. J., 2007, "Comparison of Long-Term Numerical and Experimental Total Knee Replacement Wear During Simulated Gait Loading," J. Biomech., **40**(7), pp. 1550–1558.
[10] Halloran, J. P., Easley, S. K., Petrella, A. J., and Rullkoetter, P. J., 2005, "Comparison of Deformable and Elastic Foundation Finite Element Simulations for Predicting Knee Replacement Mechanics," ASME J. Biomech. Eng., **127**(5), pp. 813–818.
[11] Box, G. E. P., and Draper, N. R., 1987, "Empirical Model-Building and Response Surfaces," *Applied Probability and Statistics* (Wiley Series in Probability and Mathematical Statistics), Wiley, New York.
[12] Khuri, A. I., and Cornell, J. A., 1996, *Response Surfaces: Designs and Analyses, Statistics, Textbooks and Monographs*, Dekker, New York.
[13] Myers, R. H., and Montgomery, D. C., 1995, "Response Surface Methodology: Process and Product in Optimization Using Designed Experiments," *Applied Probability and Statistics* (Wiley Series in Probability and Statistics), Wiley, New York.
[14] Jones, D. R., Schonlau, M., and Welch, W. J., 1998, "Efficient Global Optimization of Expensive Black-Box Functions," J. Global Optim., **13**(4), pp. 455–492.
[15] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments," Stat. Sci., **4**(4), pp. 409–435.
[16] Chen, S., Chng, E. S., and Alkadhimi, K., 1996, "Regularized Orthogonal Least Squares Algorithm for Constructing Radial Basis Function Networks," Int. J. Control, **64**(5), pp. 829–837.
[17] Wendland, H., 1995, "Piecewise Polynomial, Positive Definite and Compactly Supported Radial Basis Functions of Minimal Degree," Adv. Comput. Math., **4**(1), pp. 389–396.
[18] Wahba, G., 1987, "Spline Models for Observational Data," *CBMS-NSF Regional Conference Series in Applied Mathematics, Based on a Series of Ten Lectures at Ohio State University at Columbus*, Mar. 23–27, Society for Industrial and Applied Mathematics, Philadelphia, PA.
[19] Girosi, F., 1998, "An Equivalence Between Sparse Approximation and Support Vector Machines," Neural Comput., **10**(6), pp. 1455–1480.
[20] Vapnik, V. N., *Statistical Learning Theory*, Wiley, New York.
[21] Cox, S. E., Haftka, R. T., Baker, C. A., Grossman, B., Mason, W. H., and Watson, L. T., 2001, "A Comparison of Global Optimization Methods for the Design of a High-Speed Civil Transport," J. Global Optim., **21**(4), pp. 415–433.
[22] Koch, P. N., Simpson, T. W., Allen, J. K., and Mistree, F., 1999, "Statistical Approximations for Multidisciplinary Design Optimization: The Problem of Size," J. Aircr., **36**(1), pp. 275–286.
[23] Liu, B., Haftka, R. T., and Akgun, M. A., 2000, "Two-Level Composite Wing Structural Optimization Using Response Surfaces," Struct. Multidiscip. Optim., **20**(2), pp. 87–96.
[24] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K., 2005, "Surrogate-Based Analysis and Optimization," Prog. Aerosp. Sci., **41**(1), pp. 1–28.
[25] Queipo, N. V., Pintos, S., Rincon, N., Contreras, N., and Colmenares, J., 2002, "Surrogate Modeling-Based Optimization for the Integration of Static and Dynamic Data Into a Reservoir Description," J. Pet. Sci. Eng., **35**(3–4), pp. 167–181.
[26] Wang, G. G., and Shan, S., 2007, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," ASME J. Mech. Des., **129**(4), pp. 370–380.
[27] Bouzid, A. H., De Technologie Superieure, E., and Champliaud, H., 1998, "On the Use of Dual Kriging Interpolation for the Evaluation of the Gasket Stress Distribution in Bolted Joints," *Proceedings of the ASME Pressure Vessels and Piping Conference*, Vol. 457, pp. 77–83.
[28] Chang, P. B., Williams, B. J., Santner, T. J., Notz, W. I., and Bartel, D. L., 1999, "Robust Optimization of Total Joint Replacements Incorporating Environmental Variables," ASME J. Biomech. Eng., **121**(3), pp. 304–310.
[29] Lin, Y. C., Farr, J., Carter, K., and Fregly, B. J., 2006, "Response Surface Optimization for Joint Contact Model Evaluation," J. Appl. Biomech., **22**(2), pp. 120–130.
[30] An, K. N., Himenso, S., Tsumura, H., Kawai, T., and Chao, E. Y. S., 1990, "Pressure Distribution on Articular Surfaces: Application to Joint Stability Analysis," J. Biomech., **23**(10), pp. 1013–1020.
[31] Blankevoort, L., Kuiper, J. H., Huiskes, R., and Grootenboer, H. J., 1991, "Articular Contact in a Three-Dimensional Model of the Knee," J. Biomech., **24**(11), pp. 1019–1031.
[32] Fregly, B. J., Bei, Y. H., and Sylvester, M. E., 2003, "Experimental Evaluation of an Elastic Foundation Model to Predict Contact Pressures in Knee Replacements," J. Biomech., **36**(11), pp. 1659–1668.
[33] Li, G. A., Sakamoto, M., and Chao, E. Y. S., 1997, "A Comparison of Different Methods in Predicting Static Pressure Distribution in Articulating Joints," J. Biomech., **30**(6), pp. 635–638.
[34] Pandy, M. G., Sasaki, K., and Kim, S., 1998, "A Three-Dimensional Musculoskeletal Model of the Human Knee Joint. Part 1: Theoretical Construct," Comput. Methods Biomech. Biomed. Eng., **1**(2), pp. 87–108.
[35] Walker, P. S., Blunn, G. W., Broome, D. R., Perry, J., Watkins, A., Sathasivam, S., Dewar, M. E., and Paul, J. P., 1997, "A Knee Simulating Machine for Performance Evaluation of Total Knee Replacements," J. Biomech., **30**(1), pp. 83–89.
[36] Desjardins, J. D., Walker, P. S., Haider, H., and Perry, J., 2000, "The Use of a Force-Controlled Dynamic Knee Simulator to Quantify the Mechanical Performance of Total Knee Replacement Designs During Functional Activity," J. Biomech., **33**(10), pp. 1231–1242.
[37] Hammersley, J. M., 1960, "Related Problems .3. Monte-Carlo Methods for Solving Multivariable Problems," Ann. N.Y. Acad. Sci., **86**(3), pp. 844–874.
[38] Diwekar, U. M., 2003, *Introduction to Applied Optimization*, Springer, New York.
[39] Kalagnanam, J. R., and Diwekar, U. M., 1997, "An Efficient Sampling Technique for Off-Line Quality Control," Technometrics, **39**(3), pp. 308–319.
[40] Fregly, B. J., Banks, S. A., D'lima, D. D., and Colwell, C. W., Jr., 2008, "Sensitivity of Knee Replacement Contact Calculations to Kinematic Measurement Errors," J. Orthop. Res., **26**(9), pp. 1173–1179.
[41] Cressie, N., 1992, "Statistics for Spatial Data," Terra Nova, **4**(5), pp. 613–617.
[42] Gupta, A., Ding, Y., Xu, L., and Reinikainen, T., 2006, "Optimal Parameter Selection for Electronic Packaging Using Sequential Computer Simulations," ASME J. Manuf. Sci. Eng., **128**(3), pp. 705–715.
[43] Lophaven, S. N., Nielsen, H. B., and Søndergaard, J., 2002, "Dace—A Matlab Kriging Toolbox," Technical Report No. IMM-TR-2002-13, Informatics and Mathematical Modelling, Technical University of Denmark.
[44] Johnson, K. L., 1987, *Contact Mechanics*, Cambridge University Press, Cambridge.
[45] Archard, J. F., and Hirst, W., 1956, "The Wear of Metals Under Unlubricated Conditions," Proc. R. Soc. London, Ser. A, **236**(1206), pp. 397–410.
[46] Fisher, J., Dowson, D., Hamdzah, H., and Lee, H. L., 1994, "The Effect of Sliding Velocity on the Friction and Wear of Uhmwpe for Use in Total Artificial Joints," Wear, **175**(1–2), pp. 219–225.
[47] Kane, T. R., and Levinson, D. A., 1985, *Dynamics, Theory and Applications* (Series in Mechanical Engineering), McGraw-Hill, New York.
[48] Daffertshofer, A., Lamoth, C. J. C., Meijer, O. G., and Beek, P. J., 2004, "PCA in Studying Coordination and Variability: A Tutorial," Clin. Biomech., **19**(4), pp. 415–428.
[49] Valero-Cuevas, F. J., Johanson, M. E., and Towles, J. D., 2003, "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., **36**(7), pp. 1019–1030.
[50] Bei, Y. H., and Fregly, B. J., 2004, "Multibody Dynamic Simulation of Knee Contact Mechanics," Med. Eng. Phys., **26**(9), pp. 777–789.
[51] Laz, P. J., Pal, S., Halloran, J. P., Petrella, A. J., and Rullkoetter, P. J., 2006, "Probabilistic Finite Element Prediction of Knee Wear Simulator Mechanics," J. Biomech., **39**(12), pp. 2303–2310.
[52] Wu, Y. T., Millwater, H. R., and Cruse, T. A., 1990, "Advanced Probabilistic Structural-Analysis Method for Implicit Performance Functions," AIAA J., **28**(9), pp. 1663–1669.
[53] Jin, R., Chen, W., and Simpson, T. W., 2001, "Comparative Studies of Metamodelling Techniques Under Multiple Modelling Criteria," Struct. Multidiscip. Optim., **23**(1), pp. 1–13.
[54] Simpson, T. W., Peplinski, J. D., Koch, P. N., and Allen, J. K., 2001, "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," Eng. Comput., **17**(2), pp. 129–150.
[55] Bai, B., Baez, J., Testa, N. N., and Kummer, F. J., 2000, "Effect of Posterior Cut Angle on Tibial Component Loading," J. Arthroplasty, **15**(7), pp. 916–920.
[56] Blunn, G. W., Walker, P. S., Joshi, A., and Hardinge, K., 1991, "The Dominance of Cyclic Sliding in Producing Wear in Total Knee Replacements," Clin. Orthop. Relat. Res., **273**, pp. 253–260.
[57] Kawanabe, K., Clarke, I. C., Tamura, J., Akagi, M., Good, V. D., Williams, P. A., and Yamamoto, K., 2001, "Effects of a-P Translation and Rotation on the Wear of Uhmwpe in a Total Knee Joint Simulator," J. Biomed. Mater. Res., **54**(3), pp. 400–406.
[58] Barnett, P. I., Fisher, J., Auger, D. D., Stone, M. H., and Ingham, E., 2001, "Comparison of Wear in a Total Knee Replacement Under Different Kinematic Conditions," J. Mater. Sci.: Mater. Med., **12**(10–12), 1039–1042.
[59] D'lima, D. D., Hermida, J. C., Chen, P. C., and Colwell, C. W., 2001, "Polyethylene Wear and Variations in Knee Kinematics," Clin. Orthop. Relat. Res., **392**, pp. 124–130.
[60] Johnson, T. S., Laurent, M. P., Yao, J. Q., and Gilbertson, L. N., 2001, "The Effect of Displacement Control Input Parameters on Tibiofemoral Prosthetic Knee Wear," Wear, **250**, pp. 222–226.