Medical Engineering and Physics 37 (2015) 885–891

Contents lists available at ScienceDirect

Medical Engineering and Physics

journal homepage: www.elsevier.com/locate/medengphy

Surrogate modeling of deformable joint contact using artificial neural networks

Ilan Eskinazi [a], Benjamin J. Fregly [a,b,c,*]

[a] Department of Mechanical & Aerospace Engineering, University of Florida, Gainesville, FL, USA
[b] Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA
[c] Department of Orthopaedics and Rehabilitation, University of Florida, Gainesville, FL, USA

ARTICLE INFO

Article history:
Received 10 February 2015
Revised 9 June 2015
Accepted 24 June 2015

Keywords:
Elastic contact
Surrogate modeling
Neural networks
Biomechanics
Response surface
Meta-model
Knee contact
Tibiofemoral joint

ABSTRACT

Deformable joint contact models can be used to estimate loading conditions for cartilage–cartilage, implant–implant, human–orthotic, and foot–ground interactions. However, contact evaluations are often so expensive computationally that they can be prohibitive for simulations or optimizations requiring thousands or even millions of contact evaluations. To overcome this limitation, we developed a novel surrogate contact modeling method based on artificial neural networks (ANNs). The method uses special sampling techniques to gather input–output data points from an original (slow) contact model in multiple domains of input space, where each domain represents a different physical situation likely to be encountered. For each contact force and torque output by the original contact model, a multi-layer feed-forward ANN is defined, trained, and incorporated into a surrogate contact model. As an evaluation problem, we created an ANN-based surrogate contact model of an artificial tibiofemoral joint using over 75,000 evaluations of a fine-grid elastic foundation (EF) contact model. The surrogate contact model computed contact forces and torques about 1000 times faster than a less accurate coarse grid EF contact model. Furthermore, the surrogate contact model was seven times more accurate than the coarse grid EF contact model within the input domain of a walking motion. For larger input domains, the surrogate contact model showed the expected trend of increasing error with increasing domain size. In addition, the surrogate contact model was able to identify out-of-contact situations with high accuracy. Computational contact models created using our proposed ANN approach may remove an important computational bottleneck from musculoskeletal simulations or optimizations incorporating deformable joint contact models.

1. Introduction

Deformable contact models can be incorporated into multi-body dynamic simulations to compute the loads resulting from surface–surface interactions. In musculoskeletal biomechanics, the need to model deformable contact typically arises when cartilage–cartilage [1–3], implant–implant [4,5], human–orthotic [6,7], and foot–ground [8–11] interactions occur. However, deformable contact models are computationally expensive and thus can be prohibitive for studies that require large numbers of repeated contact evaluations such as optimizations and forward dynamic simulations.

Surrogate contact models can provide one solution to this problem. Surrogate models, also known as meta-models or response surface approximations, fit or interpolate input–output relationships sampled from a "slow" computational model (e.g., a finite element or elastic foundation contact model). The simplest example of a surrogate contact model is a response surface or multiple linear regression model, which has been used to calculate cartilage–cartilage contact forces in a natural tibiofemoral joint [12]. Kriging is a more complex surrogate modeling technique that has also been used to model contact forces and torques in the knee [13]. Kriging-based contact models have been used in an optimization approach that predicted muscle forces, tibiofemoral contact forces, and patellofemoral contact forces simultaneously in the knee during walking [14]. Other efforts to create surrogate knee contact models include a Hammerstein–Wiener model, a nonlinear autoregressive model with exogenous input, and a time delay artificial neural network [15]. In addition, a surrogate foot–ground contact model has been created using a lazy learning interpolation method [16].

While each of these surrogate contact modeling methods improves computational speed, each also suffers from important limitations. Kriging-based models suffer from two disadvantages. First, only a relatively low number of sample points can be interpolated given

* Corresponding author at: Department of Mechanical & Aerospace Engineering, 231 MAE-A Building, PO Box 116250, University of Florida, Gainesville, FL 32611-6250, USA. Tel.: +1 352 392 8157; fax: +1 352 392 7303.
E-mail address: fregly@ufl.edu (B.J. Fregly).

http://dx.doi.org/10.1016/j.medengphy.2015.06.006
1350-4533/

a computer's memory resources. As the number of sample points increases, so does the necessary memory and computation time required for model construction and use. Second, the most common implementation of Kriging interpolates the data instead of regressing it. This property becomes a disadvantage when the data contain noise, such as with element-based contact models whenever a proportionately small number of elements are loaded (e.g., at low loads with a coarse mesh). Hammerstein–Weinner, nonlinear autoregressive and time delay ANN models require knowledge of past configurations in time to evaluate the current configuration of the contacting bodies. This requirement implies that these methods can be used only in time-incremented analyses. Lazy learning models have the advantage of bounding the prediction error and adapting to a changing domain, but these benefits come at the cost of requiring additional "slow" contact model evaluations during surrogate model use.

To address the need for fast, accurate, and multi-purpose surrogate contact models for musculoskeletal simulations and optimizations, this study explores the use of multi-layer feed-forward ANN models. ANNs can be formulated as time-independent regression problems capable of fitting arbitrary observable functions [17]. Feed-forward ANNs have been used as black box models in previous non-contact biomechanical simulations to determine multi-dimensional input–output relationships [18,19] but have not been explored for deformable contact applications. Our proposed ANN contact modeling approach includes a special sampling technique that seeks to improve computational speed and accuracy over existing Kriging-based schemes. The method also permits fitting many more sample points than would be possible with Kriging, allowing for varying levels of accuracy across multiple domains of input space. Furthermore, the approach fits the sample points via regression rather than interpolation, effectively smoothing noise in the sampled data points [20]. The computational speed and accuracy of our ANN contact modeling approach are evaluated using an elastic foundation (EF) contact model of an artificial tibiofemoral joint.

2. Methods

2.1. Surrogate contact modeling background

The goal of surrogate contact modeling is to replace a computationally "slow" contact model with a computationally "fast" contact model that exhibits the same input–output characteristics. This process involves sampling the computationally expensive model, henceforth called the *original model*, using a sampling plan or design of experiments. Sampling yields a series of *sample points* that relate original model inputs to original model outputs. For multi-body applications, the inputs to the original model are *pose parameters* consisting of three translations and three rotations defining the position and orientation of one contacting body with respect to the other. The corresponding outputs of the original model are *contact loads* consisting of three contact forces and three torques calculated with respect to a selected point on one of the contacting bodies. Finally, testing takes place to evaluate the discrepancy between the original model and the resulting surrogate model using sample points not included in the surrogate model construction process.

The method presented in this paper builds upon a previous study to which the reader is referred for further details [13]. A brief summary of the most relevant concepts is introduced next. The first concept is that of fixed and moving bodies. The fixed body is the contacting body that is conceptually considered to remain fixed in space, while the moving body is the contacting body conceptually considered to move. The position and orientation of the moving body with respect to the fixed body are defined by pose parameters consisting of three translations and three rotations. The second concept is that of sensitive directions. A sensitive direction is a degree of freedom (DOF) which when perturbed causes a relatively large change in associated contact loads. Every sensitive direction possesses an associated sensitive pose parameter and one or more associated sensitive contact loads. For example, if changing the y-translation by a small amount yields a large contact force change in the y-direction, then y-translation is a sensitive pose parameter and y-force is a sensitive load. The third concept is that of a sample point. A sample point is defined as a set of model inputs and corresponding outputs. While one would normally expect to use pose parameters as inputs and loads as outputs, sample point inputs and outputs are permitted to be any combination of pose parameters and loads. For example, sample point inputs could be defined as three rotations and three forces. The fourth concept is use of sample point definitions that contain sensitive pose parameters as outputs and the corresponding sensitive loads as inputs. Sampling in this manner results in a more desirable distribution of sample points since deeply interpenetrating and out-of-contact situations can be avoided. Use of such a sample point definition requires the original model to be sampled via repeated static analyses.

2.2. Multiple domains

In contrast to surrogate modeling methods such as Kriging, feed-forward ANNs can fit tens of thousands of sample points. Therefore, ANNs provide the ability to approximate contact models sampled in a variety of configurations that have not been previously considered. This capability motivates a new sampling approach.

Our sampling strategy consists of combining multiple domains of input space, each with a different span and sample point density. A large span minimizes the likelihood of evaluating the surrogate model outside the sampled domain, a situation that would lead to large prediction errors. A high sample point density leads to low prediction errors within the sampled domain. To maximize model accuracy, we combine sparsely sampled domains having large spans with densely sampled domains having limited spans that cover regions of input space likely to be encountered during the activity being simulated.

To define both types of domains, we introduce the concept of a reference envelope, which we will use to define the upper and lower bound of a domain. We obtain multiple time-histories of a pose parameter or load of interest corresponding to the activity to be simulated. Upper and lower bounds for these curves are defined for each time frame. These bounds comprise the reference envelopes that describe estimated variations in pose parameters and loads of interest across the entire motion.

To define a sparsely sampled domain having a large span, we expand the maximum and minimum values of the reference envelopes across all time frames by user-specified amounts. The resulting time-invariant upper and lower bounds define a domain that forms a large six-dimensional (6-D) hypercuboid input space. This space is filled using a Hammersley quasirandom sequence [21]. The domain should exclude physically unrealistic sections of input space corresponding to contacting surfaces "pulling" on each other.

To define a densely sampled domain having a limited span, we expand the maximum and minimum bounds of the reference envelopes at each time frame by user-specified amounts. The resulting time-varying upper and lower bounds define the domain. For each time frame of the reference motion, we define a 6-D hypercuboid where each dimension corresponds to a sample point input. The 64 vertices of each 6-D hypercuboid are sampled, and the interior of each hypervolume is filled using a 6-D Hammersley sequence (Fig. 1). The resulting sample points will be closely packed and will cluster around the reference envelopes.

Additional domains are added to obtain sample point inputs that place the contact surfaces in an "almost unloaded" condition (contact boundary points) and in an out-of-contact condition. In this way, we include domains to capture specific configurations.

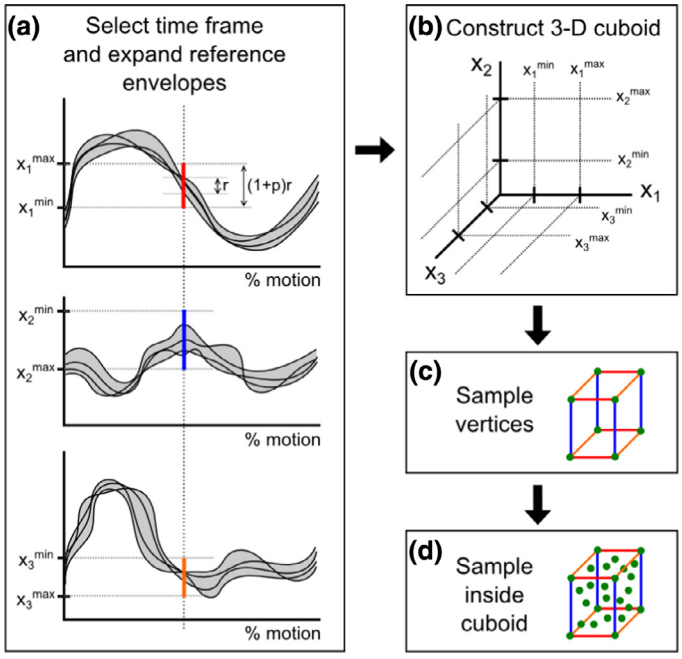

Fig. 1. Domain sampling using reference envelopes defined by curves of pose parameters and/or loads. The displayed three-dimensional sampling analogy describes how the reference envelopes (gray areas) are expanded and sampled. (a) Envelope magnitude "*r*" at the selected time frame is expanded by a percentage value "*p*", but more generally it could be expanded by a specified offset. (b) At the selected time frame, the expanded envelope forms a cuboid input space. (c) The vertices of the cuboid are sampled. (d) Finally the volume of the cuboid is sampled using a quasirandom space-filling method.

2.3. Special case for two contact patches

Contacting bodies possessing two contact patches that are not "close" to one another present a challenge for sampling contact boundary configurations. These scenarios result in a contact model having a rotational sensitive direction. A previous study [13] would define the sensitive torque and compressive force as inputs and the corresponding sensitive rotation and translation as outputs. However, this combination of inputs does not allow us to specify explicitly contact configurations where one contact patch is heavily loaded while the other one is at the contact boundary. Our solution to this problem is to replace the input force and torque with two forces acting within the area of their contact patches. For example, the knee has separate medial and lateral contact patches, making abduction–adduction rotation and superior–inferior translation sensitive directions. Instead of using the adduction torque and superior–inferior force as inputs, we use the vertical medial and lateral forces as inputs.

2.4. Sampling

For each combination of load and pose parameter inputs, we obtain corresponding outputs by performing a static analysis. Static analyses can be performed via optimization or forward dynamic simulation, both of which fix the moving body pose parameters to those from the sample point input. When using optimization, we attempt to match the loads from the sample point input by varying the pose parameters from the sample point output. When using forward dynamic simulation, we apply the loads from the sample point input to the moving body, add joint axis damping, and use implicit numerical integration until all accelerations are zero to within a specified tolerance. At that point, the model is static and we have found the poses corresponding to the applied loads.

2.5. Surrogate model structure

We construct a surrogate model by interconnecting multiple ANNs, each with multiple inputs and a single output. The inputs to the surrogate model are pose parameters – three translations and three rotations. The outputs of the surrogate model are any loads that we desire to compute. The proposed surrogate modeling structure contains two stages. The first stage consists of *k* ANNs that fit sensitive loads as a function of the pose parameters, where *k* is the number of sensitive directions in the contact model. The second stage is composed of ANNs that fit the remaining loads as a function of both pose parameters and sensitive loads calculated in the first stage. This approach is similar to a previous one [13] except that all pose parameters are inputs to all ANNs. Unlike in [13], we do not swap a sensitive pose for a sensitive load as a second stage ANN input, since some loads will be zero in out-of-contact configurations.

2.6. ANN architecture and training

Each feed-forward ANN within the surrogate model is fully interconnected, containing sigmoidal hidden neurons and a single linear output neuron. We train the ANNs using the batch-mode back-propagation Levenberg–Marquardt algorithm. The training process is programmed to stop when one of the following three criteria is met: (1) the training error is below a pre-specified error, (2) the allotted training time has been exceeded, or (3) the mean squared error of a validation data set has increased consecutively for a specified number of training iterations. The first criterion stops training when the ANN fit is considered to be "good enough," the second stops training when the ANN has exceeded a user-defined training time limit, and the third stops training when the model begins to overfit the data [22,23].

There are multiple ANN architectures that can yield a satisfactory function approximation. We achieve one such architecture by incrementally adding neurons and layers. We start by training a single hidden-layer network of a small number of neurons. We train the ANN while keeping track of how the fitting error diminishes with time. If we estimate that the fitting error after a certain training period will be above our desired error, then we stop training and change the ANN architecture by increasing the number of neurons in the layer and/or by increasing the number of hidden layers. Adding more neurons to each layer and adding more hidden layers increases the capacity of the ANN to encode the training data.

2.7. Surrogate model testing

We calculate the root-mean-square error and maximum absolute error for each of the surrogate model outputs using sample points not used for training that were kept segregated by domain. This approach allows us to evaluate accuracy within each domain. We assess computational speed by calculating the average computation time given a series of sample points. Furthermore, we compare the speed of the surrogate model and original model under simulation conditions.

3. Example application

To demonstrate how our ANN-based surrogate contact modeling approach works in practice, we applied the methodology to construct and evaluate a surrogate contact model of an artificial tibiofemoral joint. The geometry of the implant components, and the kinematics and contact loads needed to define reference envelopes, were taken from the First Grand Challenge Competition to Predict In Vivo Knee Loads [24].

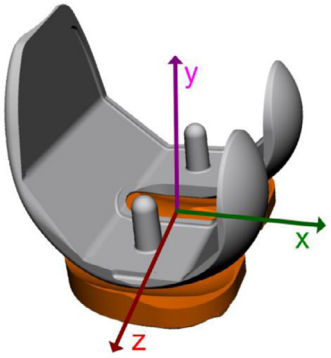

Fig. 2. Implant components for a right knee with coordinate system fixed in the tibial insert. *X* points posteriorly, *Y* points superiorly, and *Z* points medially. Tens of thousands of static analyses were performed using an elastic foundation contact model to obtain sample points for surrogate model training.

3.1. Original TF contact model

The original TF contact model was an EF contact model of the subject's implant components (Fig. 2). We treated the femoral component as a rigid moving body and the tibial insert as a deformable fixed body. We selected a fine grid of 80 × 40 elements on each side of the tibial insert to generate sample points for surrogate model creation [25]. In the original contact model, the $+x$ direction pointed posteriorly, the $+y$ direction superiorly, and the $+z$ direction medially (Fig. 2). Euler angles r_x, r_y, and r_z corresponding to an x–y–z body-fixed rotation sequence defined the orientation of the femoral component with respect to the tibial insert. Translations t_x, t_y, and t_z defined the position of the femoral component origin with respect to the tibial insert origin. These translations and rotations constituted the pose parameter inputs to the original model. The EF contact model outputs were the vertical contact loads F_y^{med} and F_y^{lat} for the medial and lateral compartments, respectively, along with total contact forces F_x^{tot}, F_z^{tot} and total contact torques $T_x^{tot}, T_y^{tot}, T_z^{tot}$ calculated with respect to the origin of the tibial insert. The vertical contact force was split into medial and lateral components because both quantities were relevant to our simulation purposes. All contact loads acted on the tibial insert and were expressed in the coordinate system of the tibial insert. For surrogate contact model use, contact loads of equal magnitude and opposite direction are applied to the femoral component.

3.2. Sample point definition for TF contact

We selected a sample point definition where the inputs were $\{t_x, F_y^{med}, t_z, F_y^{lat}, r_y, r_z\}$ and the outputs were $\{F_x^{tot}, t_y, F_z^{tot}, r_x, T_z^{tot}, T_z^{tot}\}$. Had we followed the methodology of [13], we would have defined the sample point inputs as $\{t_x, F_y^{tot}, t_z, T_x^{tot}, r_y, r_z\}$ because superior–inferior translation t_y and abduction–adduction rotation r_x are sensitive directions. This sample point input definition would not have allowed us to prescribe medial and lateral vertical loads independently. Our new definition allowed us to sample configurations where one contact patch is deeply in contact while the other one is at the contact boundary.

3.3. Sampling plan for TF contact

We sampled the EF contact model in seven domains labeled D1 through D7, each with physical significance as described below (Table 1). After sampling, we placed 90% of the sample points into a single pool used for surrogate model training. The remaining sample points were held back in segregated pools used for surrogate model testing.

We constructed reference envelopes for four pose parameters (t_x, t_z, r_y, r_z) and two contact loads (F_y^{med}, F_y^{lat}) using 14 gait cycles from the selected Knee Grand Challenge data set. For each time frame of each gait cycle, we used the four load cell measurements from the instrumented tibial tray to calculate medial F_y^{med} and lateral F_y^{lat} contact force along with anterior–posterior (AP) and medial–lateral (ML) center of pressure (CoP) location. We also calculated knee flexion angle r_z from a marker-based inverse kinematic analysis. After locking r_z to this value and r_y to the value measured fluoroscopically for a single gait cycle, we performed a pose estimation optimization that adjusted the four free pose parameters such that F_y^{med}, F_y^{lat}, AP CoP location, and ML CoP location matched their experimentally measured values as closely as possible for the selected time frame. These optimizations yielded 14 sets of curves for t_x, t_z, r_z, F_y^{med}, F_y^{lat} and a single curve for r_y. The time-varying upper and lower bounds for each curve were used to define reference envelopes.

Domains D1 and D2 defined sets of sample points that clustered around the reference kinematic and load envelopes. Domain D1 was smaller and denser than D2 and completely contained within D2. D1 and D2 were defined as the reference envelopes expanded by ± 20% and ± 100%, respectively. For a specific time frame, if the difference between the upper and lower bounds for a reference envelope was below the mean difference for all time frames, then rather than using a percentage of the envelope, we expanded with an offset of ± 0.5 times the mean difference for D1 and ± 2 times the mean difference for D2. This expansion method ensured that the domains were never too small for time frames that showed below average variability. Since there was only one curve for r_y, this curve was expanded by offsets corresponding to our estimates of variability.

Domain D3 was designed to cover a large input space and was formed by expanding the maximum and minimum of each reference envelope by ± 25% over the entire gait cycle. The resulting domain formed one large 6-D hypercuboid input space which was sampled using the Hammersley method.

Domains D4 through D7 addressed out-of-contact and contact boundary configurations (Fig. 3). Domain D4 corresponded to configurations where either the medial or lateral side was out of contact. This domain was obtained by taking points where one of the condyles was at the contact boundary, changing r_x such that one of the condyles lifted off, and adjusting t_y such that the vertical force

Table 1
Physical meaning and approximate number of sample points in each domain. Relatively few sample points are required to fit configurations where both the medial and lateral sides are out of contact (D5) or at the contact boundary (D7).

Domain	Physical meaning	Approximate No. of sample points
D1	Expand gait cycle envelope by ± 20% at each time frame	24,000
D2	Expand gait cycle envelope by ± 100% at each time frame	24,000
D3	Expand gait cycle maximum and minimum value by ± 25%	24,000
D4	One side out of contact	30,000
D5	Both sides out of contact	300
D6	One side at contact boundary	20,000
D7	Both sides at contact boundary	100

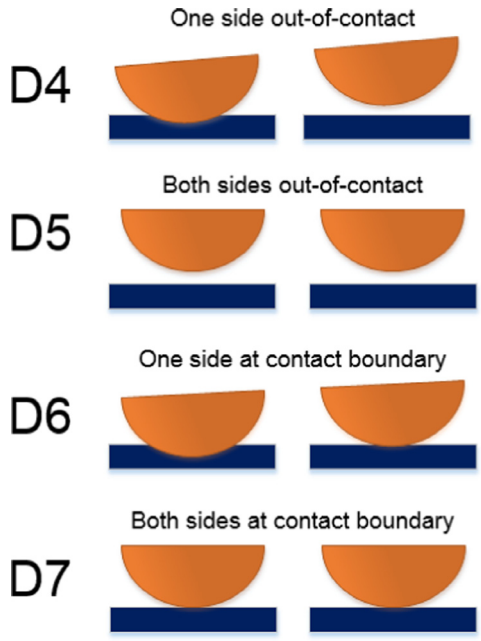

Fig. 3. Illustration of surface configurations for domains D4–D7 representing out-of-contact and contact boundary configurations. The flat surfaces represent the tibial insert while the curved surfaces represent the femoral component condyles. Sample points from these domains could be included due to the high encoding capacity of feed-forward artificial neural networks.

in the other condyle remained the same as it was previously. The change in r_x was performed in three equal increments such that the maximum liftoff was approximately 2 mm. Domain D5 accounted for points where both the medial and lateral sides where out of contact. This domain was created by taking points were both condyles were at the contact boundary and increasing the vertical translation t_y in three increments until achieving a maximum liftoff of approximately 2 mm. All contact loads were zero in this case. Domain D6 consisted of points where either the medial or lateral side was at the contact boundary (i.e., a vertical force of 5 N). Sample points were obtained by performing Hammersley sampling on the pose parameters and on one of the vertical contact forces while the other vertical contact force remained fixed at 5 N. Lastly, domain D7 consisted of points where both sides were at the contact boundary. Sample points were obtained by Hammersley sampling only pose parameters while both the medial and lateral contact forces were fixed at 5 N.

We performed sampling using the optimization approach. The design variables were t_y and r_x while the cost function minimized the normalized difference between the original model outputs and sample point inputs F_y^{med} and F_y^{lat}. To reduce computation time, we used a coarse grid EF contact model for each optimization and input the resulting pose into the fine grid EF contact model to obtain the corresponding outputs.

3.4. Surrogate model structure for TF contact

The TF surrogate contact model was composed of seven different ANN models, each with a single output. The first stage contained ANNs that fit the sensitive forces F_y^{med} and F_y^{lat} as functions of implant pose parameters (Eqs. (1) and (2)). The second stage contained ANNs that fit the remaining load components as a function of all six pose parameters and sensitive forces F_y^{med} and F_y^{lat} computed from the first stage (Eqs. (3)–(7)).

$$F_y^{med} = f(t_x, t_y, t_z, r_x, r_y, r_z) \quad (1)$$

$$F_y^{lat} = f(t_x, t_y, t_z, r_x, r_y, r_z) \quad (2)$$

$$F_x^{tot} = f\left(t_x, t_y, t_z, r_x, r_y, r_z, F_y^{med}, F_y^{lat}\right) \quad (3)$$

$$F_z^{tot} = f\left(t_x, t_y, t_z, r_x, r_y, r_z, F_y^{med}, F_y^{lat}\right) \quad (4)$$

$$T_x^{tot} = f\left(t_x, t_y, t_z, r_x, r_y, r_z, F_y^{med}, F_y^{lat}\right) \quad (5)$$

$$T_y^{tot} = f\left(t_x, t_y, t_z, r_x, r_y, r_z, F_y^{med}, F_y^{lat}\right) \quad (6)$$

$$T_z^{tot} = f\left(t_x, t_y, t_z, r_x, r_y, r_z, F_y^{med}, F_y^{lat}\right). \quad (7)$$

The vertical contact forces were fit separately for the medial and lateral sides to allow the evaluation of medial–lateral load split during simulations.

3.5. Surrogate TF model creation and training

We designed the ANN architectures and performed ANN training using a maximum training time stopping criterion of six hours. The testing error was monitored during training and no overfitting was observed. The resulting architecture had four hidden layers of 30 neurons each and a single linear output neuron. All seven ANN models were created using Matlab's Neural Network Toolbox™ (The Mathworks, Natick, MA).

Each ANN model was fitted to a random subset of approximately 80,000 sample points taken from all seven domains. The seven trained ANNs were assembled into one Simulink® model, where the outputs of Eqs. (1) and (2) were used as inputs to Eqs. (3)–(7). If either of the vertical forces was positive, corresponding to surfaces pulling on each other, then this force was set to zero. The Simulink model was exported as C++ code to maximize computational speed.

3.6. Surrogate TF model testing

The surrogate contact model was tested using two approaches. First, accuracy was tested using 10% of the sample points from within domains D1 through D7 that were withheld from training. Second, computational speed and accuracy were assessed under conditions similar to actual simulation of a walking motion. A sequence of poses, all within domain D1 and consistent with kinematic data from five gait trials, was generated for this purpose. Loads corresponding to these pose sequences were evaluated using the EF contact model with a fine element grid (80 × 40), the EF contact model with a faster but less accurate coarse element grid (40 × 30), and the surrogate contact model. For both tests, accuracy was assessed by comparing surrogate model predictions to fine grid EF contact model calculations, while for the second test, computational speed was assessed by comparing surrogate model predictions to coarse grid EF contact model calculations. All tests were performed on a PC workstation with an Intel® Xeon® Quad Core 3.1 GHz processor.

3. Results

For the accuracy test spanning all domains, root-mean-square (RMS) errors were all below 15 N for contact forces and below 500 N mm for contact torques. Maximum absolute contact force errors were between 4 N and 250 N, while absolute contact torque errors were between 123 N mm and 4929 N mm. RMS and maximum errors increased with domain size (Tables 2 and 3). RMS errors were small for D1 (under 5 N/105 N mm), larger for D2 (under 9 N/225 N mm), and even larger for D3 (under 15 N/435 N mm). RMS errors for out-of-contact points D4 and D5 were relatively small (under 11 N/265 N mm). These RMS errors in compressive force correspond to less than 1% of the peak compressive contact force reported for walking [26].

For the speed and accuracy test using only domain D1, the average computation time per input configuration using the surrogate model was 45.5 μs and the RMS error was 2.6 N for contact forces and 78 N mm for contact torques. On average, the surrogate contact model was approximately 19,000 and 960 times faster than the EF contact model with fine and coarse grid, respectively (Table 4). With

Table 2
Contact force (F) and torque (T) RMS errors for each domain. Errors were calculated between surrogate contact model and fine grid elastic foundation (EF) contact model. Medial is "med", lateral is "lat", and total is "tot", while x is anterior–posterior direction, y is superior–inferior direction, and z is medial–lateral direction.

Domain	F_y^{med}(N)	F_y^{lat}(N)	F_x^{tot} (N)	F_z^{tot}(N)	T_x^{tot} (N mm)	T_y^{tot} (N mm)	T_z^{tot} (N mm)
D1	4.9	1.9	1.7	1.7	104	43	78
D2	8.1	6.6	4.0	3.8	225	103	182
D3	14	9.6	14	14	433	334	405
D4	10	6.2	9.1	10	245	297	262
D5	3.1	1.2	3.2	5.0	197	63	249
D6	10	5.9	9.9	7.2	302	281	261
D7	5.1	9.3	2.9	4.4	199	102	111

Table 3
Contact force (F) and torque (T) maximum errors for each domain. Errors were calculated between surrogate contact model and fine grid EF contact model. Subscripts and superscripts are as noted in Table 2. Maximum errors are about an order of magnitude larger than RMS errors indicating that while most test points are well described by the surrogate model, outliers are also present.

Domain	F_y^{med} (N)	F_y^{lat} (N)	F_x^{tot} (N)	F_z^{tot} (N)	T_x^{tot} (N mm)	T_y^{tot} (N mm)	T_z^{tot} (N mm)
D1	40	14	18	14	837	427	845
D2	51	94	61	24	1587	901	1367
D3	126	124	114	249	4656	4582	3609
D4	92	63	184	115	1352	3787	3815
D5	10	4.0	7.6	13	451	123	865
D6	71	54	109	119	2859	4929	1401
D7	14	23	7.0	12	457	288	237

Table 4
Speed and accuracy comparison between surrogate and EF contact models. Errors were calculated by comparing surrogate model outputs to fine grid EF contact model outputs for sequences of poses corresponding to realistic walking kinematics from within domain D1.

Model	Average computation time (ms)	Average absolute error (N)/(N mm)
Fine grid EF contact	860.6	–
Coarse grid EF contact	43.8	20.4/574
Surrogate	4.55×10^{-2}	2.56/77.7

the fine grid EF contact model as a baseline, the surrogate contact model produced average RMS errors approximately 8 times lower for contact forces and 7 times lower for contact torques than those produced by the coarse grid EF contact model (Table 4).

4. Discussion

The goal of this study was to develop a surrogate contact modeling method that is fast, accurate, and does not require a pose history to calculate contact loads for the current pose. For the artificial tibiofemoral joint example application, our proposed ANN-based method performed extremely well. It achieved the rare combination of decreased computational speed and increased accuracy compared to our elastic foundation modeling methods [25]. In particular, it was approximately 1000 times faster and seven times more accurate than the EF contact model with coarse element grid, which is the model we have used most frequently in the past. We would expect our surrogate model to be about 50,000 times faster than a finite element contact model [27]. Thus, we anticipate that this new model formulation will perform extremely well when used in forward or inverse dynamic simulations and in optimizations.

The reported error trends were expected since the smaller domains were sampled more densely than were the larger domains. ANNs allowed for this type of clustering since all fitted points had the same weight in the training algorithm. Small errors for contact transition and out-of-contact points suggest that liftoff can be predicted accurately with this approach – a capability lacking in our previous Kriging-based method. This improvement will likely add robustness to future dynamic contact simulations and contact optimizations.

Our prediction errors should be interpreted in light of the number of sample points assigned to each domain. In the tibiofemoral joint example application, the number ranged from 20,000 to 30,000 except for D5 and D7, which had only 300 and 100 sample points, respectively. Few sample points were assigned to domain D7 because both vertical forces F_y^{med} and F_y^{lat} were predetermined to be 5 N (contact boundary), hence the six-dimensional sample space effectively become four-dimensional. For domain D5, where both the medial and lateral condyles were out of contact, our surrogate contact model was insensitive to the specified pose parameter values since that region of output space was flat (all loads were zero). Hence fewer points were needed to maintain good accuracy.

We believe that this method will work well for surrogate contact modeling of numerous anatomical joints such as the patellofemoral, hip, and glenohumeral joints. The key issue will be specification of the fixed and moving bodies and the corresponding sensitive directions. For the patellofemoral joint, one could define the patella as the fixed body with medial–lateral translation and anterior–posterior translation as sensitive pose parameters. For the hip joint, one could define the hip socket as the fixed body and all translations as sensitive pose parameters. Calculation of medial and lateral contact forces is not required while applying our method to other joints.

As with any computational method, our ANN-based method possesses important limitations. First, reference kinematics are needed to bound domains where high accuracy is desired. It would not be reasonable to expect D1-type accuracy when simulating kinematics or loading conditions outside the D1 domain. Second, sampling and fitting tens of thousands of sample points is time intensive. Depending on the speed of the contact model being sampled, it could

take days to generate the required data. Moreover, training ANNs consumes a considerable amount of computation time. Third, unlike local interpolation methods, ANNs do not provide error estimates. Consequently, if an ANN-based surrogate model is used outside the sampled domains, the user has no way to know that the outputs are invalid.

The feed-forward ANN-based method presented in this study resulted in a fast and accurate surrogate knee contact model that is flexible enough to be used in a variety of applications. Use of ANN-based surrogate contact models could allow researchers to eliminate the computational bottleneck caused by expensive contact evaluations in musculoskeletal simulations and optimizations. Future work will utilize this method to create surrogate contact models for use within simulations that estimate muscle and contact forces simultaneously in the lower limb.

Competing interests

The authors have no competing interests.

Ethical approval

Not required for this study.

Acknowledgments

Funding for this study was provided by NIH Grant R01EB009351 and by the University of Florida.

References

[1] Russell ME, Shivanna KH, Grosland NM, Pedersen DR. Cartilage contact pressure elevations in dysplastic hips: a chronic overload model. J Orthop Surg Res 2006;1:6. doi:10.1186/1749-799X-1-6.
[2] Farrokhi S, Keyak JH, Powers CM. Individuals with patellofemoral pain exhibit greater patellofemoral joint stress: a finite element analysis study. Osteoarthr Cartil 2011;19:287–94. doi:10.1016/j.joca.2010.12.001.
[3] Adouni M, Shirazi-Adl A, Shirazi R. Computational biodynamics of human knee joint in gait: from muscle forces to cartilage stresses. J Biomech 2012;45:2149–56. doi:10.1016/j.jbiomech.2012.05.040.
[4] Fitzpatrick CK, Baldwin MA, Clary CW, Wright A, Laz PJ, Rullkoetter PJ. Identifying alignment parameters affecting implanted patellofemoral mechanics. J Orthop Res 2012;30:1167–75. doi:10.1002/jor.22055.
[5] Reggiani B, Leardini A, Corazza F, Taylor M. Finite element analysis of a total ankle replacement during the stance phase of gait. J Biomech 2006;39:1435–43. doi:10.1016/j.jbiomech.2005.04.010.
[6] Liu X, Zhang M. Redistribution of knee stress using laterally wedged insole intervention: finite element analysis of knee–ankle–foot complex. Clin Biomech 2013;28:61–7. doi:10.1016/j.clinbiomech.2012.10.004.
[7] Actis RL, Ventura LB, Smith KE, Commean PK, Lott DJ, Pilgram TK, et al. Numerical simulation of the plantar pressure distribution in the diabetic foot during the push-off stance. Med Biol Eng Comput 2006;44:653–63. doi:10.1007/s11517-006-0078-5.
[8] Dorn TW, Lin Y-C, Pandy MG. Estimates of muscle function in human gait depend on how foot–ground contact is modelled. Comput Methods Biomech Biomed Eng 2012;15:657–68. doi:10.1080/10255842.2011.554413.
[9] Mahboobin A, Cham R, Piazza SJ. The impact of a systematic reduction in shoe–floor friction on heel contact walking kinematics – a gait simulation approach. J Biomech 2010;43:1532–9. doi:10.1016/j.jbiomech.2010.01.040.
[10] Neptune RR, Wright IC, Van Den Bogert AJ. A method for numerical simulation of single limb ground contact events: application to heel–toe running. Comput Methods Biomech Biomed Eng 2000;3:321–34. doi:10.1080/10255840008915275.
[11] To CS, Kirsch RF, Kobetic R, Triolo RJ. Simulation of a functional neuromuscular stimulation powered mechanical gait orthosis with coordinated joint locking. IEEE Trans Neural Syst Rehabil Eng 2005;13:227–35. doi:10.1109/TNSRE.2005.847384.
[12] Lin Y-C, Farr J, Carter K, Fregly BJ. Response surface optimization for joint contact model evaluation. J Appl Biomech 2006;22:120–30.
[13] Lin Y-C, Haftka RT, Queipo NV, Fregly BJ. Surrogate articular contact models for computationally efficient multibody dynamic simulations. Med Eng Phys 2010;32:584–94. doi:10.1016/j.medengphy.2010.02.008.
[14] Lin Y-C, Walter JP, Banks S, Pandy MG, Fregly BJ. Simultaneous prediction of muscle and contact forces in the knee during gait. J Biomech 2010;43:945–52. doi:10.1016/j.jbiomech.2009.10.048.
[15] Mishra M, Derakhshani R, Paiva GC, Guess TM. Nonlinear surrogate modeling of tibio-femoral joint interactions. Biomed Signal Process Control 2011;6:164–74. doi:10.1016/j.bspc.2010.08.005.
[16] Halloran JP, Erdemir A, van den Bogert AJ. Adaptive surrogate modeling for efficient coupling of musculoskeletal control and tissue deformation models. J Biomech Eng 2009;131:011014. doi:10.1115/1.3005333.
[17] Tinchcombe M, Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Netw 1989;2:359–66. doi:10.1016/0893-6080(89)90020-8.
[18] Arjmand N, Ekrami O, Shirazi-Adl A, Plamondon A, Parnianpour M. Relative performances of artificial neural network and regression mapping tools in evaluation of spinal loads and muscle forces during static lifting. J Biomech 2013;46:1454–62. doi:10.1016/j.jbiomech.2013.02.026.
[19] Favre J, Hayoz M, Erhart-Hledik JC, Andriacchi TP. A neural network model to predict knee adduction moment during walking based on ground reaction force and anthropometric measurements. J Biomech 2012;45:692–8. doi:10.1016/j.jbiomech.2011.11.057.
[20] Samarasinghe S. Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Boca Raton, FL: Auerbach Publications; 2007.
[21] Hammersley JM. Monte Carlo methods for solving multivariable problems. Ann NY Acad Sci 1960;86:844–74. doi:10.1111/j.1749-6632.1960.tb42846.x.
[22] Prechelt L. Automatic early stopping using cross validation: quantifying the criteria. Neural Netw 1998;11:761–7.
[23] Geman S, Bienenstock E, Doursat R. Neural networks and the bias/variance dilemma. Neural Comput 1992;4:1–58. doi:10.1162/neco.1992.4.1.1.
[24] Fregly BJ, Besier TF, Lloyd DG, Delp SL, Banks SA, Pandy MG, et al. Grand challenge competition to predict in vivo knee loads. J Orthop Res 2012;30:503–13. doi:10.1002/jor.22023.
[25] Bei Y, Fregly BJ. Multibody dynamic simulation of knee contact mechanics. Med Eng Phys 2004;26:777–89. doi:10.1016/j.medengphy.2004.07.004.
[26] D'Lima DD, Fregly BJ, Patil S, Steklov N, Colwell CW. Knee joint forces: prediction, measurement, and significance. Proc Inst Mech Eng H 2012;226:95–102. doi:10.1177/0954411911433372.
[27] Halloran JP, Easley SK, Petrella AJ, Rullkoetter PJ. Comparison of deformable and elastic foundation finite element simulations for predicting knee replacement mechanics. J Biomech Eng 2005;127:813. doi:10.1115/1.1992522.