Computational modeling and simulation of neuromusculoskeletal (NMS) systems enables researchers and clinicians to study the complex dynamics underlying human and animal movement. NMS models use equations derived from physical laws and biology to help solve challenging real-world problems, from designing prosthetics that maximize running speed to developing exoskeletal devices that enable walking after a stroke. NMS modeling and simulation has proliferated in the biomechanics research community over the past 25 years, but the lack of verification and validation standards remains a major barrier to wider adoption and impact. The goal of this paper is to establish practical guidelines for verification and validation of NMS models and simulations that researchers, clinicians, reviewers, and others can adopt to evaluate the accuracy and credibility of modeling studies. In particular, we review a general process for verification and validation applied to NMS models and simulations, including careful formulation of a research question and methods, traditional verification and validation steps, and documentation and sharing of results for use and testing by other researchers. Modeling the NMS system and simulating its motion involves methods to represent neural control, musculoskeletal geometry, muscle–tendon dynamics, contact forces, and multibody dynamics. For each of these components, we review modeling choices and software verification guidelines; discuss variability, errors, uncertainty, and sensitivity relationships; and provide recommendations for verification and validation by comparing experimental data and testing robustness. We present a series of case studies to illustrate key principles. In closing, we discuss challenges the community must overcome to ensure that modeling and simulation are successfully used to solve the broad spectrum of problems that limit human mobility. [DOI: 10.1115/1.4029304]
modeling studies to have a broad impact, and is good scientific and engineering practice, helping to prevent the proliferation of erroneous conclusions.

1.1 Definitions. Several terms must be defined before reviewing verification and validation best practices for NMS modeling and simulation.

Model: We define a model as a set of mathematical equations that describe a physical system, along with the computational expression of these equations. For our purposes, the physical system is the human or animal neural and/or muscular system acting on a rigid multibody skeletal structure, possibly in interaction with external devices and/or the environment (e.g., the ground).

Simulation: A simulation is the result of using a model to study a specific motion or biophysical event. A kinematic simulation involves analyzing the motion of a system without regard to the forces causing or produced by this motion. A dynamic simulation can be forward or inverse. In a forward dynamic simulation, the model’s equations of motion are integrated forward in time to predict the motion resulting from the application of forces. In an inverse dynamic simulation, the motion of a model is used to predict the forces that caused that motion.

Verification: We follow the ASME definition, where verification is “the process of determining that a computational model accurately represents the underlying mathematical model and its solution,” [2] or “are we solving the equations correctly?”

Validation: We also follow the ASME definition, where validation is “the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model,” [2] or “are we solving the correct equations?”

Calibration: Calibration is the process of choosing model and simulation parameters that provide the best match to experimental or other reference data. Models and simulations must be carefully calibrated before verification and validation are performed. Data used to calibrate a model cannot be used to validate a model or simulation.

Error, Accuracy, Uncertainty, Sensitivity, Credibility: Error is the difference between a measured or estimated value of a parameter and its true value. Conversely, accuracy quantifies the agreement between measured or estimated values and their true values. Uncertainty is a potential source of error, which can arise from a gap in knowledge about the biological or physical system under study, or from inherent variability in the subject or phenomenon being measured. Sensitivity is a measure of the effect of a change in a particular variable on the simulation outputs of interest. We define credibility as the degree of trust placed in a particular model or simulation for answering a specific research question. A simulation is most credible when the outputs of interest are insensitive to variables with high uncertainty. Anderson et al. [3] provide an excellent review of these concepts.

1.2 Objective. Most biomechanics researchers are aware of the importance of validation, but the field lacks best practices for the challenging process of verifying and validating NMS models and simulations. Several papers [3-5] have laid the groundwork, identifying principles and considerations, but these papers stop short of providing specific guidelines for NMS modeling and simulation. The knowledge and practices of how to best validate a biomechanical model and verify modeling software used in past research studies have not been adequately synthesized. The goal of this paper, therefore, is to establish practical guidelines for verification and validation of NMS models and simulations that researchers, clinicians, reviewers, and others can adopt to evaluate the accuracy and credibility of modeling studies.

In Sec. 2, we provide an overview of the verification and validation process in NMS modeling and simulation based on practices that have evolved in our research group and a review of the literature. Then in Sec. 3 we detail best practices for verification and validation of each common component of an NMS modeling and simulation framework, from multibody dynamics to neural control modeling. We discuss: (1) recommended modeling choices (e.g., appropriate level of modeling complexity) and verification and validation best practices for a given model or simulation output of interest; (2) how close a match is expected for comparisons between different types of simulation data and the corresponding experimental data; and (3) which variables have the greatest error, uncertainty, or influence and must be most rigorously tested in the validation process.

We illustrate the best practices through case studies. While these studies use the OPENSIM [6] software platform for modeling and simulation, our guidelines are broadly applicable across the NMS modeling and simulation field, regardless of the software used. We close the paper with a discussion of key challenges for the field.

2 Verification and Validation Process

We break the verification and validation process for modeling and simulation into seven stages:

1. Formulate a research question that a model and simulation can answer.

2. Prototype your methods and create a verification and validation plan.

3. Verify your software.

4. Validate your results by comparing your model and simulation to independent experiments and other models.

5. Test the robustness of the study by evaluating the sensitivity of your results to model parameters and other modeling choices.


7. Generate predictions and hypotheses that can be tested in the real world.

The verification and validation process begins with the definition of a suitable research question, continues through the process of designing and conducting your study, and extends beyond study completion, since sharing and documenting your results allows others to reproduce, extend, and test your models and simulations (Fig. 2). Although we have enumerated a sequence of stages, the verification and validation process typically requires iteration. For
2.2 Prototype Your Methods and Create a Verification and Validation Plan. The next step is to design your methods, including the modeling and simulation framework you will use and any experimental data you will collect, and create a verification and validation plan to ensure confidence in the analysis of your results and conclusions drawn (credibility). You should be able to answer “yes” to each of these questions before collecting data or generating models and simulations:

1. Do you understand how your modeling framework maps to the physical system of interest and how the framework is implemented?
2. Does your research question lie within the scope of intended uses of the model and simulation framework?
3. Have you eliminated model complexity not required to answer your research question?
4. Have you identified modeling assumptions and their implications?
5. Have you collected or obtained experimental data to use as inputs to calibrate your model and generate simulations, or will you collect experimental data? Do you know the variability, error, and uncertainty in these measurements?
6. Do you have independent data to assess the variability, errors, and uncertainty identified above?
7. Can you address remaining variability, error, and uncertainty with sensitivity testing?

In Sec. 3, we review the common components of a modeling and simulation pipeline and for each component, provide information about the questions above, including making modeling choices to fit your intended use, common modeling assumptions and limitations, typical inputs along with their errors or uncertainties, and best practices for independent validation and sensitivity testing.

2.3 Verify Your Software. Before generating any results or conducting validation, you must verify your software to ensure that the computational model and underlying algorithms used to simulate the physiological or physical phenomenon are implemented correctly (i.e., the results match known standards). One cannot overstate the difficulty and importance of this step: if an algorithm has been implemented incorrectly, all resulting simulations will be incorrect and any conclusions drawn from these data will be contaminated. Paradoxically, it is not the most substantial errors that are most problematic, since the erroneous results they produce are likely to be immediately apparent. Rather, it is the small errors that can be most pernicious, as they can produce results that are plausible and can, therefore, easily go unnoticed.

If the software has been designed in a modular fashion, the software engineer can perform verification for each component on its own (called “unit testing”), and can design higher-level tests...
We will use two case studies to illustrate verification and validation best practices, included as pop-out boxes throughout the paper. These case studies provide example questions that are well suited to modeling and simulation. Consider the question, “Does switching from a fast walk to a slow run at a similar pace decrease muscle contraction velocities, enabling muscles to generate greater forces?” This question, tackled by Arnold and colleagues [8], had previously been unanswered—though experimental ultrasound data of plantarflexor muscle fascicles provided supporting evidence. This question is important for understanding human locomotion and also provided a valuable set of benchmark data describing muscle fiber velocities for both walking and running. Modeling and simulation were required to answer this question, since one cannot determine normalized velocities for a comprehensive set of lower extremity muscles without them. Finally, the question was testable using current modeling and simulation tools, though careful validation and sensitivity testing were necessary. Thus, this study satisfied each of the considerations described in Section 2.1.

This first research question, part of a series of investigations published by Arnold and colleagues [7–9], will serve as one of the case studies. The experimental protocol included collecting three-dimensional optical marker data, ground reaction forces, and electromyography (EMG) signals for 11 lower limb muscles during walking and running at a range of speeds, as shown in (a), including a fast walk and slow run at similar speeds. The outputs of the simulation were the fiber length and velocity of each muscle throughout the walking and running cycle, at each speed, which were used to answer the study’s research question. The significant change in the force-generating capacity of the soleus muscle due to velocity effects across the walk-to-run transition is shown in (b).

A second research question suitable for modeling and simulation is, “Which lower extremity muscles brake, propel, and support the body mass center during running?” [11–12]. The answer to this question was unknown. Although examining joint powers and EMG provides some information, a model and simulation are required to quantify the complex skeletal and muscle–tendon dynamics during running, and to discover the contributions of individual muscles to center-of-mass (COM) accelerations. Hamner and colleagues were able to build on an existing modeling and simulation framework for analyzing accelerations induced by muscle forces, improving contact modeling to provide accurate answers for running at different speeds [10]. Finally, this study provided a valuable dataset for future analysis of running performance (e.g., to explore injury mechanisms and performance enhancement).

This research question is the basis for our second case study. The aim of this second investigation [11–12] was to determine the contribution of each lower extremity muscle to body weight support, COM braking, and COM propulsion. The investigators shared the experimental data used by Arnold and colleagues, focusing on the running trials. For this study, the outputs of the simulation were the forces generated by each lower extremity muscle and the induced COM acceleration generated by each of these muscles across the running cycle and across speeds. As an example, the upward and fore/aft COM accelerations generated by the soleus muscle are shown in (c).

Fig. 3 Introduction to verification and validation case studies

for functionalities that depend on aggregates of components. A comprehensive verification suite ensures that inputs are bounded, only the state and unknowns can change, and computations—preferably for benchmark problems with known solutions—are repeatable as the code is updated. If modeling a physical system, the verification suite should test for physical principles such as energy conservation, including targeted tests where errors and miscalculations are likely.
The most effective technique for software verification is to employ existing software modules that have been independently verified and widely used for a range of applications, whether commercial or open source. If you choose open source modules, you should verify and widely used for a range of applications, whether commercial or open source tools. Some that we have used successfully include LAPACK, GCVSPKLINE [13], and PAPI [14]. The OPENMP [6,15] and SIBODY [16] software packages are also open source to promote community review and contributions. Publicly available modeling and simulation platforms should include a suite of verification tests for their modeling libraries and simulation algorithms. Even if you are not conducting the verification yourself and instead relying on existing testing suites, you should be aware of the key requirements for verification.

As mentioned above, using modular or object-oriented software design is a particularly useful strategy for efficient and robust software verification. In object-oriented design, complex code is built up from simpler atomic components, each of which can function (and, therefore, can be tested) in isolation. Each component has a clear interface that defines its inputs, the modifiable quantities (state), and its outputs, but hides its internal implementations from users (including other components) to help promote compartmentalization and facilitate unit testing. Complexity is engineered through hierarchy, with high-level components coordinating the activities of simpler subcomponents. Consider the following example, adapted from Reddy [17], of a multibody system consisting of a collection ("container") of rigid bodies. Suppose the container has two methods: it can return a body’s color given its index (e.g., calling the function getColor on "body #2" returns "blue") and it can return a body’s index given its name (e.g., calling the function getIcon on "femur" returns "2"). This design obviates the need to provide a third method that directly returns the body’s color given its name, since the two existing methods can be cascaded to do the same job (i.e., calling getIcon on femur and then calling getColor on the returned index, 2, will give the desired answer, blue). Providing only the first two methods is preferable: if naming or indexing conventions change in the future, the third method will create errors that will go unnoticed unless there is a separate verification test for this new method. Branching logic and circular relationships are also discouraged, even if they make the code shorter, since linear processing is easier to understand, test, and debug. Components should be reusable so that new and custom code can be minimized, isolated, and easily targeted by tests. The more a block of code is used, the more likely it is to be verified in a wide range of use cases. Much has been written on the topic of object-oriented design (e.g., Refs. [17] and [18]).

More details about verification testing for each part of the modeling and simulation pipeline are included in Sec. 3. Following the verification stage, you should be confident that your equations (mathematical models) are implemented correctly. You can then begin generating and analyzing your initial modeling and simulation results and proceed with formal validation, described in Secs. 2.4 and 2.5, below.

2.4 Validate Your Results by Comparing Your Model and Simulation to Independent Experiments and Other Models. The first step of formal validation is comparing the outputs of your model and simulation to as many independent datasets as possible. Validation against independent data is distinct from model calibration (the process of finding the best-fit model and simulation to available data), though both are key steps in any modeling and simulation study. Calibration should be conducted first, followed by comparison to any independent data not used to tune your model or generate your simulation.

A wide range of potential data are available for calibration and validation. These sources include data from a typical motion capture analysis like optical marker trajectories, ground reaction forces, and electromyography (EMG) signals, along with data collected from imaging or cadaver studies to help define musculo-skeletal dynamics and geometry. Additional experimental modalities, such as ultrasound to measure muscle fascicle and tendon dynamics in vivo or instrumented knee replacements to measure internal joint loads, also provide valuable data for calibration and validation. In one common workflow, for example, experimental kinematics, joint and ground reaction force data are used to generate a muscle-driven simulation of the observed motion. In this case, the experimental kinematics and ground reaction force data are used to validate the simulation. EMG data are reserved for independent validation to help researchers determine whether the muscle coordination predicted by the simulation is a good fit to experimentally observed timing of muscle activity. Comparing muscle–tendon dynamics to ultrasound data, or predicted joint loads to data from instrumented joint replacements, is additional potential data to aid validation in this example workflow.

We also recommend comparing the predictions of models and simulations to previously validated and published studies, when possible. For example, if you are creating a new simulation framework for studying pathological gait, your framework (minus any pathology-specific changes) should produce gait simulations of normal walking that are a good match to the many walking simulations available in the literature. Data you can compare include joint angles and moments, ground reaction forces, muscle activations and forces (timing and/or magnitude), internal joint loads, and muscle fiber and tendon velocities.

More details about using experimental and previously published data for calibration and validation of each component of a modeling and simulation framework, including how close a match is expected, are described in Sec. 3.

2.5 Test the Robustness of the Study by Evaluating the Sensitivity of Your Results to Model Parameters and Other Modeling Choices. The next step in the validation process is evaluating sensitivity, since validation by comparison to independent experimental data often leaves remaining uncertainty (e.g., one of the model’s input parameters may have known experimental measurement error that cannot be eliminated). Many quantities are impossible to adequately measure experimentally (otherwise, we would not need a model!). As a modeler, your key responsibilities are to determine the range of possible outcomes (e.g., confidence intervals) and how sensitive the outputs of interest (i.e., those used to test your hypothesis) are to the input parameters and data—particularly those with large known variability, uncertainty, or influence. Sensitivity analysis can also help determine the level of modeling complexity appropriate for your study by assessing whether your conclusions are robust to a particular modeling simplification or assumption (e.g., reducing the number of musculotendon actuators in your model or ignoring muscle activation dynamics).

Several approaches to sensitivity analysis are available, from simple to complex [19]. The direct approach is a differential analysis, which determines the analytic relationship between inputs and outputs [20]. Although generally not practical for complex, nonlinear models and simulations, differential analysis serves as the foundation upon which other approaches to sensitivity analysis are built. A straightforward and common approach to sensitivity testing is the parametric study, where the researcher sweeps through a range of input parameters, varying one parameter at a time or multiple parameters in a factorial design, to determine interaction effects. A parametric assessment can be sufficient when possible input ranges and sensitivities are known in advance; however, testing a large number of interactions quickly becomes intractable.

Sampling methods for sensitivity analysis, such as Monte Carlo analysis, are also common in biomechanics research (e.g., Refs. [21] and [22]). The researcher creates random samples of the input parameters given specific probability distributions (e.g., Gaussian with an estimated mean and standard deviation (SD) based on data from experiments). The corresponding distribution

of the output is then estimated by performing repeated runs of the simulation or analysis. Sensitivity factors are computed based on the relationship between inputs and the output distribution (using a correlation coefficient, for example). The Monte Carlo approach requires many (often thousands) runs of a simulation and, thus, can become computationally expensive. Various techniques are available to improve sampling efficiency (e.g., Latin hypercube sampling [23]). Modern methods for probabilistic analysis develop an approximate relationship between input and output distributions to substantially reduce the number of simulation iterations required (e.g., using prior knowledge in a Bayesian inference approach [24]). These methods must be validated against the “gold standard” Monte Carlo approach, though initial tests in biomechanics look promising [19]. These approximate methods can provide information about the relative importance of various input parameters, but do not include enough information to determine the direction of influence.

Several texts and review papers provide overviews of approaches to sensitivity analysis for engineering applications [25,26], including biomechanics [19]. A review of known sensitivities in common modeling and simulation frameworks is provided in Sec. 3.

2.6 Document and Share Your Model and Simulation. After careful validation through comparison to independent data and sensitivity analysis, you will have gained confidence in the ability of your model and simulation to answer the research question you posed; however, the validation process does not end when the last result is computed or figure is generated. An additional, vital step is documenting your modeling and simulation methods, results, and conclusions. In the documentation process, you should clearly indicate how your findings answer your original research question and how your validation process has adequately addressed known sources of error and uncertainty. In most cases, some uncertainty will remain, so you must also describe the known limitations and detail how these limitations might impact your conclusions.

In addition to standard publications in journals, we believe that sharing your models, simulation tools, and results with other researchers and clinicians is essential for validation and helps ensure your research has a broad impact. Allowing others to review your models, simulations, and software can help identify errors and improve your models and simulation tools. As others apply your models and simulation tools to new research questions and analyses, more information about both the strengths and limitations of your model and simulation will be established. Finally, sharing your simulation data helps expand the pool of available independent data for future researchers to use in the validation process, as shown in the feedback arrow from “Generate simulation” to “Validate your results” in Fig. 2.

2.7 Generate Predictions and Hypotheses That Can Be Tested in the Real World. The validation process can continue beyond the life of a single modeling and simulation study by generating hypotheses that you or other researchers test with experimental data. In other words, do the high-level predictions and analyses hold up to independent testing with clinical or experimental data? For example, if a model suggests that plantarflexor muscle strength is vital for maintaining adequate knee extension during stance in children with cerebral palsy, does a plantarflexion strength-training program improve patients’ gait? Does an assistive device to reduce the metabolic cost of uphill walking, designed with a simulation, work when you build the device and test it with human subjects? Establishing more links like these between modeling and clinical or experimental studies is essential for advancing the fields of biomechanics and rehabilitation research.

3 Best Practices for Verification and Validation of Neuromusculoskeletal Models and Simulations

In this section, we review best practices for verification and validation of each potential component of a study’s modeling and simulation framework (Fig. 4). These best practices are based on a review of the literature, standards employed by our research group, and experience working with many researchers using computer models to gain insights into human movement. Since the validation process starts with the formulation of a research question and study design, we begin each section by reviewing modeling choices and providing recommendations, such as adjusting the level of modeling complexity to match that required by the study.

Generation of human and animal movement, and thus computer modeling and simulation of these phenomena, is a complex, multistep process, as demonstrated in Fig. 4. Studies can include models or estimations of neural command, musculotendon dynamics, musculoskeletal geometry, multibody dynamics, and contact and other external forces. The dynamics equations describing these components can be integrated forward in time to generate a simulated movement, or a known motion can be used in an inverse analysis to gain a greater understanding of the muscle coordination and forces involved in generating that motion. A study may include all or only a small subset of the components shown in Fig. 4. For example, many studies consider only multibody dynamics, using marker data from a motion capture experiment to determine joint angles. Although this type of motion analysis is
often not considered to be modeling, an underlying model—even if the model is simply a set of bodies, each with six degrees of freedom—is used to formulate equations of motion and calculate accelerations, velocities, and positions over time. Since multibody dynamics is part of nearly all modeling studies, this is where we begin our review.

### 3.1 Multibody Dynamics

Multibody dynamics enables the description and analysis of interconnected rigid and flexible bodies that move relative to each other, in the presence of forces generated internally and applied externally. Multibody dynamics provides the mathematical formulation (model) of the physical system—the equations of motion—that enables the modeler to calculate accelerations, velocities, and positions over time. A multibody dynamics analysis comes in many forms. We may wish to focus on only motion or kinematics, as in a typical inverse kinematics analysis, or study the motion resulting from forces applied by muscles in a forward dynamic simulation. We can solve for joint reactions to estimate the forces required by ligaments and articulating surfaces in the knee joint to prevent separation and interpenetration. Or, with an induced acceleration analysis, we can determine how much a single force in the system (due to gravity or a muscle, for example) accelerates each body in the multibody system.

#### 3.1.1 Modeling Choices

The first step in performing a multibody dynamics analysis is determining the underlying model that will be used to guide the solution of kinematics and dynamics. The model determines how bodies are permitted to move relative to each other and, in the case of a dynamics analysis, the distribution of mass in the system (i.e., inertial properties). Every type of motion analysis requires a model, but in many cases, the user of a commercial motion analysis processing package is unaware of the details of the underlying model and whether that model is appropriate for the motion under study. For example, some packages impose no constraints on body motions, which means the model can undergo nonphysiological motions like separation or impetration of body segments [27,28]. We recommend performing both inverse kinematics and dynamics analyses using a model that represents physiological joints and is scaled to the anthropometry of the subject. Using a model of the underlying skeletal geometry prevents nonphysiological motions and typically makes inverse kinematics and dynamics computations more robust to noise [29]. More details about formulating and validating models of biological joints are included in Sec. 3.2.

Given an underlying model, the next key decision is how to formulate and solve the corresponding equations of motion. We recommend, when possible, using existing multibody dynamics codes that have been designed and verified for engineering and scientific use. ADAMS® and SIMBODY [16] are, respectively, commercial and open source examples. Software that has originally been designed for other purposes, such as gaming and computer animation (e.g., ODE, Bullet, and Havok), must be used cautiously and verified independently, since these codes are designed to optimize speed and appearance rather than physical accuracy.

One common approach to formulating the equations of motion is to use Cartesian coordinates (as in ADAMS, ODE, and Bullet). Each body segment has six degrees of freedom (i.e., its position and orientation in space), and constraints are added where necessary to model the limitations imposed on the motion by joints. This formulation tends to be easier to implement in software, but requires many nonlinear constraints that are usually solved with approximate iterative methods to maintain tolerable execution times. We recommend, instead, using an internal coordinate approach [30], which provides an exact representation of idealized joint motion. The resulting system of equations is small and dense, and is generally solved with exact methods. The internal coordinate approach is more difficult to implement, but several existing dynamics packages such as SIMBODY [16], MOTIONGENESIS (formerly AUTOLEV) [31], and SD/Fast [32] contain well-tested implementations.

Once the equations of motion have been formulated, they are solved over time to analyze or predict the motion of a physical system. We recommend using error-controlled numerical methods to ensure accuracy and careful state handling, as described by Sherman et al. [16]. An important design consideration is the handling of the model’s state, which is the set of unknowns or variables that fully describes the system at an instant in time. Tentative state changes are required for error estimation, but can lead to incorrect computations if the state is not handled correctly. Thus, dynamic simulation software must be designed and verified to ensure that tentative state changes persist and desist appropriately.

Additional considerations when choosing a dynamics engine, such as the method for enforcing constraints, the handling of discrete and continuous values, and the user interface for defining accuracy, are included in publications on SIMBODY [16] and OpenSim [33]. The numerical methods employed by a dynamics engine are also relevant to verification and validation, and are discussed in Sec. 3.6.

#### 3.1.2 Verifying Multibody Dynamics

Multibody dynamics software can be verified in many ways [34]. We recommend ensuring that the physical system adheres to the principles of conservation (including mass, momentum, and energy) across a wide range of conditions. For example, a common verification task is ensuring that constraints and ideal joints do no mechanical work. In a forward dynamics simulation, all applied forces can be divided into their constitutive power generators (ideal actuators), storage components (e.g., potential energy due to gravity or elastic strain), and dissipative components (e.g., friction or viscous). In this case, verification should confirm that system energy minus work plus dissipated energy is conserved for each force and actuator, and also for the multibody system as a whole.

We can also represent the same physical system using different formulations, the comparison of which is another avenue for verification. For example, joints can be modeled with internal coordinates or free joints with constraints. These formulations should produce identical solutions (within numerical tolerance). When developing new multibody simulation software, we recommend verifying that a modeled system in your code produces the same results as the identical system modeled in existing code(s).

#### 3.1.3 Validating and Evaluating Robustness of Multibody Dynamics

Following verification, a multibody dynamics engine can be used to calculate motions and forces for the system of interest. The accuracy of your multibody dynamics solution is limited by the quality of the input data you provide—typically from motion capture and force plate recordings—and the quality of the underlying model. Traditional marker-based motion capture systems generally have system errors of 1–5 mm [35], which is combined with soft tissue artifacts of up to 10 mm for human movement [36]. Joint torques calculated from inverse dynamics have been shown to be sensitive to both force plate recordings and kinematics [37,38]. Thus, motion capture systems must be regularly and carefully calibrated and you should clearly document all steps of your experimental protocol (e.g., with written notes, photos, and videos). Errors in the underlying model can also create discrepancies, particularly when models of the joints are not included. Careful model selection and subject-specific scaling, as discussed in Sec. 3.2.1, are essential for obtaining multibody dynamics models that closely match to experimental measurements. We illustrate practical validation of multibody dynamics by returning to our case study on running. The case study was introduced in Fig. 3 and in Fig. 5, we review the process the
investigators used for validating multibody dynamics, the first step of which was careful data collection and model scaling. We recommend comparing calculated joint angles and moments to those reported by previous studies, when possible. For example, kinematics and kinetics for walking and running at multiple speeds for healthy individuals are readily available in the literature (e.g., Refs. [12] and [39]) and new simulation results should be within 2 SD of previously reported values. Researchers often use a model to reproduce experimental kinematics from external markers placed on a subject. We recommend, in this case, testing that model marker locations fall within the measurement accuracy (combing calibration accuracy [35] with skin movement [36] or marker placement [40] errors) for each marker.

An additional approach to validation is comparing the acceleration of the system center of mass (COM) to the net externally applied force (i.e., $F = ma$), as demonstrated in the case study of Fig. 5. In many motion-capture assessments of inverse dynamics, the size of the discrepancy between motion and forces is neither calculated nor reported—a dangerous omission, since a discrepancy nearly always exists [41]. In a motion capture experiment, kinematic data (obtained from optical markers or inertial measurement units) are measured independently from kinetic data.
(obtained from force plates and sensors) and errors in either of these data can lead to inconsistencies.

Dynamic inconsistencies should be calculated using the model’s equations of motion by comparing the net forces and moments acting on the defined COM from the measured ground reaction force to the product of the system’s mass matrix and COM accelerations calculated from inverse kinematics. When forward simulations are performed (e.g., in Computed Muscle Control [42]) or a force balance is calculated (e.g., in Static Optimization [43]), residual or “hand of god” forces are typically applied to achieve dynamic consistency between the motion and the applied forces. In general, we recommend ensuring that residual forces and moments are small enough that they do not affect the conclusions of the study (e.g., if a residual moment is contributing significantly to the accelerations of a variable of interest, the residual must be reduced). In particular, we recommend force discrepancies that are 5% or less (peak and RMS) than the magnitude of the experimentally measured net external force and residual moments that are less than 1% of COM height times the magnitude of the measured net external force.

The first key step in minimizing dynamic inconsistencies is well-calibrated data collection and careful processing, along with proper scaling of the model to the subject of interest, as described above. If $F = ma$ discrepancies are still large, additional adjustments can be made to the underlying model (e.g., changing inertial parameters) or the kinematics (e.g., eliminating motions that the model cannot achieve) to reduce or eliminate the dynamic inconsistencies [6]. If the recommended agreement cannot be achieved with reasonable adjustments to the model or kinematics, the assumed model may be of insufficient quality to capture the underlying system dynamics or to represent the observed experimental data. For an example of this process, see the case study of Fig. 5, where the investigators ensured that residual forces and moments were reduced to the recommended thresholds, while maintaining kinematics that were within measurement error.

### 3.2 Musculoskeletal Geometry

The next layer of detail that many studies include is a model of muscle and skeletal geometry. Musculoskeletal geometry includes properties of bodies, which for the purposes of this paper are rigid with specified mass and inertial properties. The definitions of joints, the connections between bodies, are also part of musculoskeletal geometry. Joints can range from generic mechanical connections, like pins and gimbals, to more specialized relationships that define the coupled motion of various structures. For example, one common model of the knee joint couples the translation of the cartilage and bone at the articular surface, along with other structures like menisci, the joint capsule, ligaments, and muscle (e.g., the rotator cuff muscles of the shoulder). The effects of these various structures can be represented in a joint model in several ways, from simple to complex, including less or more anatomical detail.

On the simple end of the joint modeling spectrum, skeletal joints can be built from basic mechanical joints (e.g., a ball-and-socket joint is a reasonable approximation of the hip), often with passive elements or other constraints to limit the range of motion and represent the net action of passive structures (e.g., to prevent excessive hip extension beyond physiological ranges). Constraints can also be used to represent coupled motion. For example, one common model of the knee joint couples the translation of the tibia with respect to the femur as a function of knee flexion angle $\theta$.

At the complex end of the spectrum, a joint can include explicit contact and ligament models that constrain the joint to follow physiological motion. This type of model allows more detailed analysis of loads, but incorporates many parameters that are potentially difficult to measure and must be calibrated and validated. For many applications, an explicit contact and ligament model at the joint is unnecessary. Models combining simple mechanical or custom biologically inspired joints with constraints and lumped passive properties are typically sufficient to reproduce gross motions like walking, reaching, running, cycling, etc. These model formulations can also provide net internal joint moments (e.g., Ref. [53]).

**Bodies.** The dimensions and inertial parameters of the rigid bodies that compose a human or animal model have been measured in cadavers and via imaging in living subjects to create a generic set of anthropometry (e.g., Refs. [44] and [45]) that is then scaled to represent individual subjects. The anthropometry of bodies can also be estimated on a subject-specific basis with dual-energy X-ray absorptiometry scans, magnetic resonance imaging (MRI), and 3D body scans (e.g., Refs. [46–48]). The dimensions of a subject’s body segments can also be estimated from manual measurement or from markers by finding the distance between functional joint centers (e.g., Refs. [49] and [50]) or between markers placed on anatomical landmarks.

We recommend, at minimum, obtaining marker-based or careful manual measurements of segment lengths and using these measurements to scale the dimensions and inertial properties of a generic model to an individual subject. If feasible, imaging techniques can provide valuable additional information in the scaling process. We recommend subject-specific modeling through MRI, measurements of bone alignment, or other means in cases where body segment geometry is known to deviate substantially from existing generic models [51].

**Joints.** In Sec. 3.1.1, we discussed how to map the description of a joint’s motion to a system of dynamic equations. In this section, we review how to define a joint’s motion based on experimental measurements. In the musculoskeletal system, the motion permitted by a particular joint depends on the geometry and properties of the cartilage and bone at the articular surface, along with other structures like menisci, the joint capsule, ligaments, and muscle (e.g., the rotator cuff muscles of the shoulder). The effects of these various structures can be represented in a joint model in several ways, from simple to complex, including less or more anatomical detail.

On the simple end of the joint modeling spectrum, skeletal joints can be built from basic mechanical joints (e.g., a ball-and-socket joint is a reasonable approximation of the hip), often with passive elements or other constraints to limit the range of motion and represent the net action of passive structures (e.g., to prevent excessive hip extension beyond physiological ranges). Constraints can also be used to represent coupled motion. For example, one common model of the knee joint couples the translation of the tibia with respect to the femur as a function of knee flexion angle $\theta$.

At the complex end of the spectrum, a joint can include explicit contact and ligament models that constrain the joint to follow physiological motion. This type of model allows more detailed analysis of loads, but incorporates many parameters that are potentially difficult to measure and must be calibrated and validated. For many applications, an explicit contact and ligament model at the joint is unnecessary. Models combining simple mechanical or custom biologically inspired joints with constraints and lumped passive properties are typically sufficient to reproduce gross motions like walking, reaching, running, cycling, etc. These model formulations can also provide net internal joint moments (e.g., Ref. [53]).

We recommend, as discussed above, using internal coordinate formulations that permit motion in only the known degrees of freedom, rather than models with many degrees of freedom and
many constraints. Avoiding the use of constraints is particularly important for models that include body segments whose masses are small relative to the applied forces (as in the upper extremity, for example). Incorporating constraints in such models is computationally expensive due to the numerical singularities approached as segment masses approach zero.

Several types of experimental data are available to help formulate joint models. In cadaver experiments, the joint of interest can be isolated and its kinematic and dynamic properties can be measured externally with a loading rig (e.g., Refs. [54–56]. High-resolution imaging and bone pin data can also provide information about the relative motion of bodies in the skeletal system [57,58], as can less invasive—though less accurate—surface marker motion. We recommend basing new joint models on high-quality data from cadavers, bone-pins or high-resolution imaging, including cinefluoroscopy [59] which provides high-speed, high-resolution data of loaded joints.

When using existing joint models, we recommend that you acquire motion capture data to help determine subject-specific joint geometry (often called functional joint centers). Identifying joint axes based on anatomical landmarks can be problematic, since there is significant variability in marker placement between and within examiners [40]. Ehrig et al. [49] and MacWilliams [50] review methods for finding functional joint centers. In cases of pathology, MRI data can also provide detailed subject-specific information about joint center locations [51]. When choosing a joint model, you must also ensure the model has been validated for the entire range of motion you plan to study.

Muscle Geometry. The geometry of muscle–tendon units can include detailed path information, or the modeler can simplify muscle geometry with regression or other equations that directly define moment arms as functions of joint angles. For muscles with broad attachment sites (e.g., the gluteus medius), several muscle segments may be required to capture the varying moment arms of the muscle. Whether using moment arms or path geometry, a key modeling choice is the number of muscles used in the model; you should tune the model’s complexity to suit your research question. Studying the contribution of the abdominal muscles to stability during walking, for example, requires more detailed trunk musculature than studying the contributions of the ankle plantarflexor muscles.

Muscle moment arms can be measured in cadavers directly using tendon excursion experiments or using load experiments (e.g., Refs. [60–62]). Muscle moment arms can also be estimated experimentally using MRI, computerized tomography scans, or the digitization of cadavers to determine a muscle’s line of action and the perpendicular distance to the joint center of rotation (e.g., Refs. [61–63]). We recommend using moment arm data to calibrate and validate your musculoskeletal model, as discussed in Sec. 3.2.3.

Detailed models of muscle geometry include muscle paths defined by origin and insertion points, which are often combined with via points and wrapping surfaces to represent constraints from retinacula and to prevent muscles from passing through bones or reaching other nonphysiological configurations. We recommend modeling detailed muscle path geometry to accurately predict muscle–tendon dynamics and internal joint loads, although this approach is more computationally expensive than using precomputed moment arms. Muscle geometry can be determined by dissecting and digitizing muscles in cadavers (e.g., Refs. [64–66]) or can be estimated by segmenting MR images (e.g., Refs. [63] and [67]). Origin and insertion points are typically defined as the centroids of the muscle attachments in the cadaver or imaged subject, but via points and wrapping surfaces have been defined several ways. For example, Horsmann et al. fit geometric shapes to measured points on the surface of a cadaver [65]. We recommend using wrapping surfaces and via points to achieve physiological wrapping over bones and to match experimentally measured moment arms (e.g., Refs. [7] and [68]).

Modeling Assumptions and Limitations. Many phenomena are often ignored or simplified in models of musculoskeletal geometry. The geometry of a muscle is typically reduced to one or more line segments. This simplification makes simulation computationally tractable, but fails to represent the true underlying complexity of muscle geometry. Further, multiple muscles attach to a single tendon (e.g., the triceps surae), but models typically assume separate tendons exist for each muscle component. The deformability of bodies is also ignored. Some parts of the human musculoskeletal system are nearly always drastically simplified. For example, there is currently no widely used and well-validated detailed model of the foot and connective tissues (e.g., the iliotibial band) are often ignored or simplified. When documenting results, these limitations should be recognized and researchers should discuss potential impacts on the study conclusions. These assumptions are also important areas for future research to improve current models.

3.2.2 Verifying Musculoskeletal Geometry. The key steps for verification of skeletal models are described in Sec. 3.1.2. We recommend, in short, verifying that bodies and joints adhere to the laws of physics. Given the complexity of human and animal musculoskeletal geometry (e.g., joints may be complex and muscles may include via points and wrapping surfaces), we recommend using the effective torque method to calculate muscle moment arms, as described by Sherman et al. [69]. To verify muscle geometry code using the effective torque method, the model developer applies a unit tension to the muscle of interest and uses the system’s equations of motion to calculate the resulting (effective) moment about the joint; the moment arm is the moment divided by the unit muscle force. The moment arm computed using the effective torque method should be verified against the moment arm (r) calculated as the change in muscle–tendon length (df) resulting from a change in the joint angle (df), or r = d/dθ. As observed by Sherman et al. [69], this equation is a consequence of the assumption that all related constraints are workless and is commonly used to measure muscle moment arms experimentally (known as the “tendon-joint” excursion method) [62].

3.2.3 Validating and Evaluating Robustness of Musculoskeletal Geometry. Each component of musculoskeletal geometry should be validated, where possible, along with the musculoskeletal model as a whole.

Bodies. Anthropometry varies by age, sex, pathology, and natural variation (e.g., Refs. [44,45,70,71]). Errors in model dimensions and inertial parameters can affect joint angle and moment calculations [72–74], which will result in errors in estimated muscle forces. If performing a dynamic analysis, we recommend comparing the whole-body COM acceleration computed from kinematic data to the acceleration generated by measured ground reaction forces. If there is a large discrepancy, one possible source of error is missing or poorly modeled body segments (see Sec. 3.1.3 for more details).

Joints. Joint properties (e.g., centers of rotation and laxity) vary between individuals, particularly in the case of injury or pathology (e.g., Ref. [75]), and can have a significant effect on joint angles, moments, and internal loads (e.g., Ref. [76]), as well as muscle moment arms (e.g., Ref. [51]). In a Monte Carlo–based study of simulated maximum thumb tip forces, Valero-Cuevas et al. [22] identified the kinematic description of the joint as being the most influential parameter preventing the model from generating physiological maximum forces. When developing a joint model, we recommend ensuring that modeled kinematic are within measurement error of high-resolution experimental data (e.g., from cinefluoroscopy or bone pin data). When using the model in a simulation, the dynamic inconsistencies between the model, kinematic data, and force data should also be small as described in Sec. 3.1.3. Poor joint modeling is a common source of error. Further, we recommend comparing predictions of internal joint loads
to estimates from instrumented joint replacements; several data-
sets are available for the hip and knee joints [77–80].

**Muscle Geometry.** Muscle geometry and moment arms vary
between individuals as well. For example, Murray et al. [66]
found that moment arms could vary by a factor of two, although
these differences could be accounted for by using a simple scaling
technique. Herzog and Read [81] also observed considerable dif-
ference in moment arm magnitudes but found that the variation of
moment arms with joint angles were generally similar between
subjects. Duda et al. [82] digitized femoral muscle attachment
sites for six cadavers and found that the SD of centroid locations
was 80% of the mean. Attachments can also vary with pathology,
such as bone deformities common in children with cerebral
palsy [67].

Given this variability and uncertainty, careful validation is an
important step in any study that includes muscle geometry. When
using a model of muscle geometry, we recommend comparing the
model’s moment arms to experimental data to ensure that magni-
tudes and trends are similar (ideally within 2 SD) throughout the
range of motion that you plan to study. You should also compare
modeled muscle paths to imaging data (e.g., from MRI). Further,
we recommend that you combine musculoskeletal geometry with
muscle–tendon dynamics and perform indirect validation by com-
paring simulated joint moments to experimentally measured val-
ues, again throughout the range of motion of interest. Validation
data can include maximum moments from dynamometer testing,
experimentally measured passive moments through a range of
motion, and inverse dynamics–based moments from motions like
walking, running, or reaching. In our research group, we aim to
achieve moments within 2 SD of experimental values, unless
known experimental uncertainty or documented modeling
assumptions can account for additional discrepancies. In cases
where agreement is not achieved, you must clearly document the
limitations of your model and take these limitations into account
when interpreting the results of your simulation. In the case study
of Fig. 6, we demonstrate the process of comparing the outputs of
a musculoskeletal model to experimental moment arms and
passive and active net joint moments.

Muscle–tendon kinematics, dynamics, and forces (and, thus, in-
ternal joint loads) are all sensitive to a model’s muscle geometry
and moment arms. We recommend performing a sensitivity analy-
sis when your research question depends on muscle geometry or
moment arms, such as when the question demands estimating
joint loads or muscle force magnitudes. Several investigators have
explored the sensitivity of predicted muscle forces and activations
to moment arms and muscle geometry. In general, the magnitude
of a muscle force is more sensitive to geometry, while timing of
muscle activity is fairly robust to changing muscle geometry
[83–86]. Correa et al. [86] studied muscle contributions to COM
and joint accelerations using models with different moment arms,
and found the predicted muscle function to be similar, though
force magnitude varied. Grouping muscles with similar paths was
found to have little effect on the timing or coordination of esti-
mated activity in a muscle-driven simulation of walking [87].
There is evidence that joint force predictions are sensitive to how
muscles are discretized into paths in a given model [88], but more
work is needed to understand this relationship.

### 3.3 Muscle–Tendon Dynamics

While the muscle geometry described above can characterize the motion of the whole
muscle–tendon unit, a model of muscle–tendon dynamics is required to estimate muscle activations, muscle forces, and the
interplay between muscle and tendon during motion. In this sec-
tion, we focus on the most widely used computational model of
muscle, the Hill-type model, which includes a contractile fiber,
series and parallel elastic elements, and activation dynamics.

#### 3.3.1 Modeling Choices

Modeling muscle–tendon dynamics adds significant complexity as well as many additional parameters
to calibrate and outputs to validate. In some cases, it is reasonable
to ignore or simplify muscle–tendon dynamics. Examples where
muscle–tendon dynamics can be ignored include studies where
only net joint moments or powers are required, or in purely kine-
matic studies of joints or whole muscle–tendon units. Inverse
dynamics with static optimization to resolve muscle forces fre-
quently ignores tendon moment arms and enables estimation of
muscle forces and joint loads. We recommend including a full
computational model of muscle and tendon dynamics when your
research question depends on the interaction between muscle and
tendon during a motion or when elastic storage of energy in ten-
don is a known or likely contributor to the motion of interest (e.g.,
high-force motions like running). Further, recent work in predic-
tive simulation has shown that muscle–tendon dynamics are
required to achieve humanlike motion [89,90]. If computational
or other restrictions require making simplifications, we recom-
mand using sensitivity testing to determine the impact of omitting
complexity (e.g., modeling rigid versus compliant tendons) in
your model.

**Muscle–Tendon Dynamics Input Parameters.** Before valida-
tion, you must understand and calibrate the parameters of the
muscle–tendon dynamics model. Parameters for Hill-type muscle
models include the maximum isometric force of the muscle, deter-
mined by combining estimates of physiological cross sectional
area (PCSA) with muscle specific tension; pennation angle; the
force–length–velocity relationship; the passive muscle force–
length relationship; timing parameters for activation dynamics;
and the stiffness and slack length of the tendon. Architectural
parameters (e.g., pennation angle) are typically derived from
measurements in human cadavers, while the parameters that
define dynamic muscle force generation (e.g., the force–length–
velocity relationship) are typically derived from isolated muscle
experiments in animals. Additional data can also come from MR
images (to determine muscle volumes), laser diffraction (to deter-
mine sarcomere lengths), and ultrasound imaging (to determine
muscle fascicle lengths).

**We recommend** creating models with muscle–tendon dynamics
based on comprehensive muscle and tendon architecture data
from cadaver and MRI measurements of multiple subjects. Mod-
els using MRI-based estimates of muscle PCSAs (e.g., Ref. [91])
are preferred, since cadaver measurements are typically obtained
from older subjects whose PCSAs can be substantially smaller
than those of young, healthy subjects [92–94]. Many current mod-
els are built from the older Wickiewicz dataset [95], which was
obtained from a small number of cadavers without the aid of laser
diffraction to measure sarcomere lengths. The fiber lengths in this
study differ by 10–100% from the more recent Ward dataset [96].
The optimal fiber lengths in the dataset provided by Ward and col-
leagues are based on more subjects as well as measurements of
sarcomere lengths.

Hill-type muscle models rely on normalized curves for the
force–length, force–velocity, and passive properties of muscle
fibers based on experiments in rat, cat, and rabbit muscle, where
force output is measured for a range of lengths and activation
levels (see Ref. [97] for a review). We recommend using muscle
model curves that are a close match to available experimental data
(e.g., Ref. [98]).

**Simplifying Muscle–Tendon Dynamics.** It may be reasonable
to simplify muscle–tendon dynamics, depending on the research
question and outputs of interest. Some studies ignore the series
elastic element and assume the tendon is inextensible, which can
significantly decrease computation time [98]. Assuming a rigid
tendon is valid for analyzing muscles with short or very stiff ten-
dons (e.g., some of the upper extremity muscles and muscles
crossing the hip) or in slower motions with low forces in which
tendon stretch is minimal (e.g., walking [99]). The parameter
that affects the tendon modeling decision is the ratio of the mus-
cle optimal fiber length to the tendon slack length [64]. We
Case Study – Choosing and validating a musculoskeletal model

The investigation by Arnold and colleagues [8] into the force-generating capacity of lower extremity muscles depended on the interplay between muscle and tendon dynamics (i.e., fiber and tendon velocities) during motion. The researchers built a model [7] with high-fidelity musculoskeletal geometry and muscle–tendon dynamics based on a comprehensive muscle architecture dataset from twenty-one cadavers [95]. The investigators then validated the predictions of the model against a range of experimental data, including moment arm curves (a), passive moments (i.e., moments generated by muscles when they are inactive) (b), maximum isometric joint moments throughout the range of motion (c), and inverse dynamics moments from experimental studies of walking and running (d). The moment arms calculated with the model for the knee flexors (a), and other muscles, match experimental data measured in cadavers. The passive moments generated by inactive muscles crossing the knee (b) tend to generate larger forces than passive moments measured in experiments, especially at large knee flexion angles. This deviation occurs because the fibers of the vasti are stretched beyond their optimal lengths and generate passive forces. The maximum isometric moments generated by the knee flexors and extensors (c) have roughly the same magnitude as in experimental measurements, but the angle at which the moment peaks differs between experimentally measured maximum moments and the moments predicted by the model. These differences between the model and experimental results produce a discrepancy between the moments calculated from inverse dynamics and the moments generated in the simulation (d), and must be taken into account when interpreting the simulation results.

![Graphs showing different aspects of muscle behavior and validation](image)

Adapted from [7, 8].

**Fig. 6 Case study—choosing and validating a musculoskeletal model**

recommend including tendon compliance in motions with large forces like running or when analyzing elastic energy storage in tendons, where compliance plays a crucial role. For example, in an optimal control simulation of running, ignoring tendon compliance reduced the maximum sprinting speed achieved [100].

Additional simplifications include ignoring the force–length–velocity relationship or activation dynamics, although we recommend including these properties, in most cases. Excluding the force–length–velocity relationship did not significantly change the muscle forces during walking predicted by static optimization [99], although many muscles have been shown to have a wide
operating range on their force–length and force–velocity curves even during walking [9]. Ignoring activation dynamics also had a small effect on static optimization solutions of walking [99], although its exclusion reduced maximum simulated running speed in a dynamic optimization study by Miller et al. [100]. Activation dynamics must also be included when using experimental EMG measurements as the neural control input [8,9].

Assumptions and Limitations. Several phenomena are ignored in Hill-type muscle models. Experimental evidence exists for variable-gearing pennate muscles [101], force enhancement [102], short-range muscle stiffness [103], and variations in the force–length curve at submaximal activation [104], but these phenomena are not accounted for in typical Hill-type models. The mass of muscles and effects of fatigue are also typically ignored, as is the distribution of fast- and slow-twitch fibers within a muscle. Fiber lengths and velocities are assumed to be constant throughout the muscle, in contrast with recent imaging studies and finite element models [105,106]. In addition, properties of tendon (e.g., plasticity and creep) are typically omitted from musculotendon models. These limitations should be acknowledged when documenting your findings, and represent valuable areas for future work, combining experiments with models and simulations.

3.3.2 Verifying Muscle–Tendon Dynamics. Before performing validation, we recommend the following set of verification tests for muscle models. You should verify that the active and passive fiber force–length, fiber force–velocity, and tendon force–length curves are C2 continuous and that all force–length curves have no negative stiffness regions (other than the descending limb of the fiber force–length curve). These criteria prevent failure of numerical methods during simulation, such as encountering numerical singularities, and prevent nonphysical phenomena (e.g., collagen under strain does not exhibit negative stiffness). You should also check for self-consistency between the action of the whole muscle and that of the constitutive passive and active muscle components (Fig. 7). Further, energy should be conserved in the musculotendon unit (i.e., fiber work minus fiber strain potential energy should always equal tendon work minus tendon strain potential energy) during a wide range of conditions. To verify activation dynamics, you can input a step excitation and ensure the activation signal has the characteristic first-order rise or fall time (assuming a first-order activation dynamics model) and remains between the muscle model’s minimum and maximum activation values.

3.3.3 Validating and Evaluating the Robustness of Muscle–Tendon Dynamics. Validation of muscle–tendon dynamics can be accomplished through several means. First, when creating a new computational muscle model, we recommend comparing muscle forces predicted by a simulation to independent, experimentally measured forces and dynamics (i.e., length and velocity profiles) from isolated muscle experiments for a range of activation conditions (e.g., Refs. [98] and [107]). Millard et al. [98] reported mean absolute force errors of 9% for maximal activation tests and 16% for submaximal activation tests. A similar or better match for new models is desirable.

When using an existing muscle model to study a new subject or a new type of motion, we recommend combining the muscle–tendon dynamics model with a model of musculoskeletal geometry to transform calculated muscle forces into joint moments, then comparing simulated moments to experimental and inverse dynamics moments as described in Sec. 3.2.3 and illustrated in the case study of Fig. 6. Ultrasound data can also provide information to help validate muscle fascicle and tendon lengths and velocities for muscle–tendon units during motion (e.g., Ref. [108]).

Muscle–tendon dynamics predictions depend on the error, uncertainty, variability, and sensitivity of each parameter that defines the computational muscle–tendon model, as described below. When determining sensitivity of muscle–tendon models, defining the output of interest is essential. For example, single-muscle experiments are generally more sensitive to muscle model parameters than whole-body walking simulations, where forces are generated to track a specified trajectory or meet walking goals [109]. Timing of muscle activity is also generally more robust than predicted force magnitude, at least for walking and running simulations [83,110]. Little is known about the sensitivity of joint reaction forces to muscle–tendon model parameters, though joint forces are known to be sensitive to the distribution of muscle forces (e.g., Ref. [111]).

Maximum Isometric Force. The maximum isometric force of a muscle, derived from measurements of specific tension and PCSA, is a key parameter of computational muscle–tendon

![Fig. 7 Verification test to ensure power from active and passive muscle fiber and tendon is equal to whole muscle actuator power. We generated a simulation with a constant muscle excitation of 0.6 (u), an initial block position of 0 m (x), and an initial block speed of 1 m/s (x). We terminated the simulation after 0.5 s (t). The stacked area graph shows the summed power in the active muscle fiber (blue), passive muscle fiber (red), and tendon (green); the total muscle power is equal to the summation of these constituent powers (dashed black line).](http://biomechanical.asmedigitalcollection.asme.org/ on 01/25/2015 Terms of Use: http://asme.org/terms)
models. Specific tension has been estimated in several experiments for humans and animals [112,113] by determining the ratio between maximum measured muscle tension and a known PCSA, but a large range of values (11–47 N/cm²) has been reported.

PCSA can be estimated by measuring muscle mass in cadavers or measuring muscle volumes with MRI. In adults, PCSA typically scales uniformly with the product of height and mass in the lower extremity [114] and with overall muscle volume in the upper extremity [114]. In many past studies where muscle PCSAs were based on cadaver data, the specific tension parameter was selected to inflate muscle strengths. Using MRI-based PCSAs with known size relationships should avoid the need to artificially increase the specific tension parameter to inflate muscle strengths. The relative distribution of muscle strengths can change as a function of age in the upper extremity [115] and lower extremity [116], though linear scaling laws hold among individuals of similar age. MRI-derived volumes will allow development of improved age-matched musculoskeletal models. More work is needed to understand how muscle volumes and strengths are affected by pathology.

The predicted patterns of muscle force and activation for many simulations are not especially sensitive to the maximum isometric force [109,110,117], particularly when timing rather than magnitude is the most important simulation output [84] and the relative strength distribution between muscles matches experimental measurements. We recommend avoiding nonuniform muscle strength scaling, unless isometric force values are based on subject-specific experimental data.

**Pennation Angle.** Pennation angle is estimated by measuring muscle fiber orientation with a goniometer in dissected cadaver specimens or in vivo with ultrasound. The pennation angle of a muscle can vary within a muscle and throughout a motion. For example, the gastrocnemius pennation angle varies 20–30 deg during running [108]. Within the muscle, the SD of the pennation angle in gluteal and soleus muscles is large, nearly equal to the measured value [96]. Fortunately, muscle force solutions seem to be less sensitive to pennation angle than other muscle model parameters [117]. Further, in ultrasound experiments, tendon strain computations were not strongly sensitive to pennation angle [108]. More work is needed to understand how pennation angle affects predictions of muscle force and tendon lengths, velocities, and strains during motion.

**Optimal Fiber Length.** The optimal fiber length parameter is estimated by measuring fiber lengths in dissected cadaver muscles and, in more recent studies, by combining these fiber lengths with laser diffraction-based sarcomere length estimates. Optimal fiber lengths must be carefully examined in the calibration and validation process. Optimal fiber length varies by subject and within regions of each muscle [96] and muscle force predictions have been shown to be sensitive to the optimal fiber length during gait [109,117–119], and likely most other motions.

**F–L–V Properties and Passive Stiffness of Muscles.** The animal-based data that help define normative force–length and force–velocity curves are assumed to scale well to human muscles, though direct validation of this assumption is difficult. Imaging and finite element models indicate that not all fibers within a muscle operate at the same length or velocity for a motion [105,106], which could have the effect of making whole-muscle force–length and force–velocity curves broader than the single-fiber curves.

**Activation Dynamics.** Activation dynamics are typically modeled as a first-order differential equation that relates the rate of change of muscle activation to the current muscle excitation. Single-muscle and walking simulations are not particularly sensitive to changes in the activation/deactivation time constants [109].

**Tendon Properties.** The key properties of tendon that must be measured for use in simulation are its stiffness and slack length. Butler et al. [120] reported Young’s moduli of 362 MPa and 613 MPa for human cadaver gracilis and semitendinosus tendons; however, the physical properties of cadaver tissue can be affected by storage method, donor age, and other factors. Maganaris and Paul [121] measured both gastrocnemius tendon stiffness in vivo using real-time ultrasound imaging, and reported a Young’s modulus of 1.16 ± 0.15 GPa. These results agree with analogous in vivo measurements of the (less highly stressed) human tibialis anterior tendon [122], suggesting that tendon material properties are independent of function.

Tendon slack length cannot be easily measured in cadaver or imaging experiments. Researchers have typically measured fiber lengths and joint positions in cadaver experiments. Then, when building a model, slack lengths are set such that the fiber lengths in the model match the measured fiber lengths at the given position (e.g., Ref. [7]). It has also been shown that, of all the musculoskeletal model parameters, tendon slack length has the largest effect on predictions of muscle forces [85,109,118,119]. Changing the tendon slack length alters where muscles operate on the force–length curve, affects the joint angle where peak force is generated, and changes the range over which a muscle can generate force. Muscles with a high ratio of tendon slack length to optimal fiber length (e.g., greater than three) are most sensitive (e.g., gastrocnemius and rectus femoris) [64]. We recommend testing the sensitivity of your research question to tendon properties, particularly when your results depend on muscles whose tendon slack length is large relative to the muscle fiber length. An example sensitivity analysis for tendon compliance is shown in the case study of Fig. 8.

### 3.3.4 Energetic or Metabolic Modeling

Muscle activation is the primary driver of metabolic energy consumption during motion and is often a good surrogate for energetic cost. However, computational models of muscle energetics have the potential to provide more information about efficiency and energy use for the whole body and individual muscles during a range of motions. Several energetic models are available [123–126], which use muscle activation along with muscle length, muscle velocity, excitation and/or fast- and slow-twitch fiber composition to estimate the combined mechanical work done and heat released by muscles. These models are based on experiments with isolated fiber bundles from mouse and frog muscles. These models predict whole-body metabolic cost for walking and running with reasonable accuracy, but predictions are sensitive to the metabolic model, corresponding muscle model, and neural controller [90,127]. This is an important area for future work as interest in metabolic energy and efficiency grows in the community (e.g., in pathology or for designing assistive devices).

### 3.4 Contact and Other External Forces

In many simulations, experimentally measured contact forces and torques are applied to the model. These measurements can occur either before the simulation (e.g., ground reaction forces collected in a laboratory [11]) or during the simulation (e.g., in haptic feedback applications [128], where the simulation trajectory is not known a priori). In both cases, the quality of the contact force signal is limited only by the quality of the measurements; however, both strategies require experimental equipment and expertise. Contact models are useful when experimental data would be difficult or impossible to collect, such as when modeling internal joint loading [129].

### 3.4.1 Modeling Choices

We recommend choosing the contact model of least complexity and greatest computational efficiency that captures the phenomena of interest. In its simplest and most efficient form, contact can be idealized as a set of one or more kinematic constraints; alternatively, the modeler can employ a rigid or elastic contact model. Kinematic constraints represent interactions between infinitely stiff bodies, and are satisfied to...
within numerical tolerance by arbitrarily large contact forces. Ideal constraints such as those representing pin and slider joints do not exist in reality; however, computationally efficient algorithms exist for simulating the resulting differential-algebraic equations (DAEs). Thus, the constraint-based approach is useful when you don’t need information about detailed interactions, as is the case when studying muscle coordination in cycling, for example [130]. Constraint-based contact models can also reproduce ground reaction forces in walking and running when kinematics are known. In the case study of Fig. 9, we illustrate the process of validating a constraint-based contact model when performing an induced acceleration analysis.

An elastic foundation model represents an array of independent linear or nonlinear springs that generate force as a function of penetration depth of one surface into another. Such compliant contact models are convenient and can produce accurate results when properly tuned, but they can be computationally expensive when the interacting materials are stiff. Rigid contact models address this limitation by approximating the behavior of compliant models, treating impacts as instantaneous events and simulating persistent interaction using constraints. Although rigid contact models can be more efficient than their compliant counterparts, they are less realistic—in some cases, even nonphysical [131].

Assumptions and Limitations. As modeling and simulation of humans in interaction with external devices (e.g., prostheses or exoskeletons) become more common, additional models and experimental data are needed to understand and validate the complex contact that occurs at the device-soft tissue interface. More work is also needed to develop foot–ground contact models that are suitably fast for predictive simulation while still predicting contact forces that are in good agreement with experiments.
3.4.2 Verifying Contact Models. To verify the simulation of contact, we recommend checking that impacts are always energetically neutral or dissipative, normal forces are always repulsive, sliding friction is always dissipative, and other physical laws (such as those of Newton and Coulomb) are respected.

3.4.3 Validating Contact Models. Validation tasks are guided by the research question and application. If studying the upper extremity, for example, the simulation of foot-ground contact is likely irrelevant. In other cases where ground contact plays a significant role in the motion of interest (e.g., studying the triceps

---

**Case Study – Constraint-based contact modeling**

In the analysis of running by Hamner and colleagues [11,12], the investigation required calculating the contribution of each muscle to the acceleration of the center of mass of the body. When a muscle generates force, it induces a reaction force and moment at the ground; predicting these muscle-specific reaction forces and moments requires a foot–ground contact model, since the experimental ground reaction force measurement provides only the net values. In previous induced acceleration analyses for walking, it was assumed that the foot was welded to the ground during the stance phase of gait. Although reasonable for the single-limb stance phase of walking, assuming a weld constraint during running yielded unrealistic ground reaction force predictions. In particular, when Hamner and colleagues summed the contribution of each individual muscle to the ground reaction force, the resulting forces and moments were a poor match to the experimental net forces and moments (a). This result inspired Hamner and colleagues to implement a new foot–ground contact model (a rolling-without-slipping constraint), which yielded predictions much closer to experiments (b) [10]. Two of six potential reactions are presented: the vertical reaction force (first row) and the frontal plane reaction moment (second row) of the foot in contact with the ground. Model-generated reactions (thick black line) are compared to corresponding measured ground reactions averaged over 14 strides (shaded: mean ±1 standard deviation). Although the weld constraint matches most of the stance phase vertical ground reaction, it is evident that the weld applies a non-existent reaction moment in the frontal plane.

---

**Fig. 9** Case study—constraint-based contact modeling
surae using predictive simulations of running), we recommend validating the foot–ground contact model against experimentally measured ground reaction forces to be certain that appropriate stiffness and damping parameters or an appropriate constraint type (e.g., see the case study of Fig. 9) have been selected. Similarly, data collected from instrumented knee joint replacements can be used to tune or validate contact parameters for studying internal joint forces. Finally, if using a rigid contact model, we recommend validating its performance against the behavior of the compliant model it has been designed to approximate.

3.5 Neural Control. Neural control coordinates the muscle forces that drive motion and is arguably the most challenging component in Fig. 4 to model. Neural commands originate in the brain to execute planned motions, and combine with feedback from the peripheral nervous system (e.g., stretch feedback from muscle spindles). Thus, neural control of movement arises from a complex interaction of neural commands descending from the brain and reflex responses and sensory information delivered to the central nervous system from a vast set of sensors located in muscles, tendons, joints, and skin.

3.5.1 Modeling Choices. In some cases, a model of neural control is not required, since many biomechanics studies depend only on joint angles and net moments. Other studies use experimental input from EMG to estimate the neural command and drive motion or muscle dynamics. EMG-driven simulations do not generally produce a forward simulation that generates a coordinated movement unless additional, corrective forces are applied to maintain the desired trajectory. It is also difficult to acquire EMG for some muscles in the body. Nevertheless, the EMG-driven approach has been successfully used to understand muscle–tendon dynamics in walking and running (e.g., Refs. [8] and [9]) and to simulate various upper and lower extremity motions (e.g., Refs. [132] and [133]).

Modeling of neural control generally comes in two forms. The first, tracking control, is applicable when the dynamics of a motion are known (e.g., from a motion capture experiment) and the controller is responsible only for solving the muscle redundancy problem to resolve a net joint moment into a constituent moment from each muscle. The second form of neural control modeling, predictive simulation, is applicable when the motion is neither known nor assumed and the controller must represent the goals of a task to be performed (e.g., maximize jump height or minimize cost of transport) and optimize control signals to find the motion trajectory along with the moments or forces required to obtain that motion.

For both types of neural control modeling, an effort-based cost function is often used. When simulating walking and running, the specific muscle-based cost function chosen (e.g., energy

---

**Case Study – Comparing simulated muscle activations to EMG**

In the running study of Hamner and colleagues [11,12], EMG data (gray shaded bands) were used to validate the model of neural control and predicted muscle activity (solid and dashed lines). In this study, the Computed Muscle Control tool included with OpenSim [42] was used to generate the set of muscle activations necessary to match the measured dynamic trajectory from each of the experimental running trials. The investigators compared the onset and offset timing of muscle activity, which aligned well with EMG once electromechanical delays (estimated to be around 75 ms [146]) were taken into account. They also determined that the simulations and experiments showed similar features, such as large activation of the plantarflexors (e.g., gastrocnemius medialis (a)) and quadriceps (e.g., vastus lateralis (b)) and increases in activation with running speed. Capturing these features of the muscle activations was important because these muscles made large contributions to the accelerations of the mass center; thus, large differences in the activations of these muscles would affect the conclusions of the study.

![Case Study – Comparing simulated muscle activations to EMG](http://biomechanical.asmedigitalcollection.asme.org/)

---

**Fig. 10 Case study—comparing simulated muscle activations to EMG**

Journal of Biomechanical Engineering

FEBRUARY 2015, Vol. 137 / 020905-17
expended, activation squared, and activation cubed) does not have a large effect on the overall muscle coordination strategy [90,134], though it affects predicted internal joint loads [111] and estimated cost of transport [90]. While the commonly used effort-based objectives show good results in many applications, their suitability in pathological or maximum performance motions (e.g., sprinting or jumping) may be less appropriate. For example, in individuals with osteoarthritis, muscle coordination might be tuned to minimize a combination of effort and joint loads. Tracking simulations have also been shown to overpredict the cost of transport due to co-contraction when precisely matching measured motions [90]. In these situations, validation against experimentally measured muscle activity and internal joint load measurements (from instrumented joint replacements) is important.

The majority of current studies that include a neural control model are concerned only with solving the muscle redundancy problem to track experimentally measured motion. An effort-based objective (e.g., minimizing the sum of squared muscle activations) is used to solve for a set of muscle excitation or forces that track measured motions within a specified tolerance. Muscle forces can be resolved at each time step during the motion of interest (i.e., using static optimization) or over the entire motion trajectory (i.e., using dynamic optimization). The effort minimization approach, whether in a static or dynamic optimization, has been used to understand muscle coordination in a range of motions, including walking, running, cycling, and reaching (see Ref. [135] for a review). In walking, static and dynamic optimization yield similar predictions of muscle and joint forces [99].

There is also growing interest in using predictive simulation to synthesize motions de novo. In predictive simulation approaches, a high-level motion task is specified and control signals or controller parameters are determined to achieve this task. This approach has been successfully used to simulate maximum-performance activities like a maximum-height jump and a standing long jump [136–138]. In these situations, the high-level task is relatively easy to specify. An explicit structure for the underlying controller can be assumed or a parameterized set of control signals can be used as design variables. Applying predictive simulation to locomotion is more challenging than tracking-based approaches and is also computationally expensive, but several recent studies show promise [89,139].

One proposed approach to understanding and simplifying neural control is the application of muscle synergies. In this growing area of research, muscle activity is decomposed using principal component analysis or related approaches to find groups of muscles that are activated together during movement, modulated by a single scaling parameter [140,141]. There is evidence that a consistent set of muscle synergies is used between tasks such as postural perturbations in a range of directions [142] or cycling at a range of speeds [143], and that the number/complexity of synergies is reduced in pathologies that affect the motor control system, such as poststroke [144]. Controversy remains, however, about whether synergies are truly representative of the underlying physiology or are merely a consequence of experimental data processing.

Assumptions and Limitations. Modeling neural control is a grand challenge. Fundamental questions about the nature of motor control remain unanswered, and as a result open questions remain about the form a model for neural control should take. How should we parameterize the neural control system in a model to capture the motions of interest with minimal complexity? What are the roles of reflexes and feedback? Can synergies help parameterize a neural control model? Initial results on the role of synergies or a central pattern generator (e.g., Ref. [145]) in driving motion show promise in helping to develop a simple model for the neural control of locomotion, but compelling evidence is still required to determine whether and how synergies should be incorporated into a model of physiological motor control. Much also remains to be learned about our internal “cost function” for various motion tasks (e.g., walking, running, and reaching). Outstanding questions include (1) how cost functions differ between individuals or with age and pathology; (2) whether the neural control system performs a real-time optimization when external or internal conditions change unexpectedly; and (3) if the human is optimizing in real time, what signals are the neural controller and is also computationally expensive, but several recent studies show promise [89,139].

3.5.2 Verifying Neural Control. Verification of neural control models is challenging because few laws or benchmark problems exist to characterize the underlying neural system. For tracking approaches, it is common to compare predicted muscle activations to a comparable optimal control solution to verify that coordination is similar when searching for local, rather than global, minimum effort (e.g., Ref. [99]). Investigators have also added noise to the input data to ensure tracking is robust [146]. For predictive simulation, the solution approach can be applied to simple models (e.g., a passive leg swing or simple maximal effort task) where the solution (i.e., the underlying control signal) is known. In the case of muscle synergies, verification should additionally confirm the eigenvalue decomposition has been performed correctly (e.g., by reducing dimensionality, reconstituting the matrix, and checking the error relative to the original matrix) and that the error decreases as the number of synergy groups increases.

3.5.3 Validating and Evaluating the Robustness of Neural Control. Validation depends heavily on whether and how neural control is modeled in your study. In the case where experimental EMG data are used to estimate neural control, the same EMG data cannot be used for independent validation. In this case, we recommend comparing the resulting joint kinematics (if not prescribed) and net joint moments from the EMG-driven simulation to experimental kinematics and net joint moments from inverse dynamics (see Fig. 6 for an example of joint moment comparison). Even with careful model calibration and EMG processing, it is generally unreasonable to expect that the EMG-driven moment will be an exact match to the inverse dynamics-based moments. For example, EMG is often unavailable for small or deep muscles. In our group, we check for similar timing of moment peaks and strive for simulated EMG-driven moments within 2 SD of inverse dynamics moments, unless known and documented modeling assumptions can account for additional differences (see the case study of Fig. 6 for an example of this process).

In tracking simulations, muscle activations can be determined without using EMG data. In this case, we recommend comparing predicted muscle activity to measured EMG from your own experiments and/or previous studies. Since EMG is difficult to normalize and is subject to measurement error [147], you generally cannot compare EMG signals directly to predicted muscle activations from a simulation. Instead, you should determine whether onset/offset timing in experiments and simulations are in good agreement, after accounting for electromechanical delays between EMG and activation or force [148]. The case study of Fig. 10 demonstrates the application of these principles to the study of running. Further, simulations should reproduce salient features of muscle coordination established in experimental studies (e.g., the vasti should show early stance phase activation in walking). If you test and simulate multiple conditions (e.g., walking with and without a carried load), your simulation should predict similar trends between EMG and simulated muscle activation (e.g., increased vasti activity [149]). If simulation results do not match experimental EMG data, researchers will sometimes constrain muscles to be on or off at specified times in the simulation. In this case, the EMG data for these muscles become calibration data rather than validation data. Many tracking simulations also apply extra forces or moments...
validation will become available. For example, previous just normal, unassisted walking, even more data for independent sophisticated and are able to synthesize a range of motions beyond jump). As predictive simulation frameworks become more so-
capital sources to aid human jumping should produce a humanlike unassisted motions (e.g., a predictive simulation framework to design devi-

We recommend making the following comparisons. First, onset and offset of muscle activity, as for tracking simulations, should be compared to experimental EMG. Kinematics, kinetics, and ground reaction forces should also be compared to experiments. Simulations should be within 2 SD of experimental data if they are claimed to replicate human motion. In the case where a pre-
dictive simulation is designed to simulate a new motion that has not or cannot be measured, you should at least determine whether your framework is able to reproduce similar, known motions (e.g., a predictive simulation framework to design devi-
tices to aid human jumping should produce a humanlike unassisted jump). As predictive simulation frameworks become more so-
plicated independent of EMG and motion capture data.

Predictive simulations allow for a larger set of independent validation data, since muscle activations and motions are typically resolved independently of EMG and motion capture data. We recommend making the following comparisons. First, onset and offset of muscle activity, as for tracking simulations, should be compared to experimental EMG. Kinematics, kinetics, and ground reaction forces should also be compared to experiments. Simulations should be within 2 SD of experimental data if they are claimed to replicate human motion. In the case where a predictive simulation is designed to simulate a new motion that has not or cannot be measured, you should at least determine whether your framework is able to reproduce similar, known motions (e.g., a predictive simulation framework to design devices to aid human jumping should produce a humanlike unassisted jump). As predictive simulation frameworks become more sophisticated and are able to synthesize a range of motions beyond just normal, unassisted walking, even more data for independent validation will become available. For example, previous experimental studies have characterized how metabolic cost and muscle coordination change with speed and load. Researchers can now determine whether their predictive models show similar adaptations.

3.6 Numerical Methods for Simulation and Analysis. Often overlooked are the numerical methods used to solve for the unknowns in a simulation. For example, the multibody dynamics code generally forms a set of coupled DAEs, and numerical inte-
gression is required to find a solution (i.e., a trajectory of time-
dependent states) that satisfies the system of DAEs. All numerical integrators are approximate, and their accuracy (how many digits of the solution are significant) or error tolerance (the absolute or relative difference with respect to the “true” answer) are param-
ters of the numerical integration algorithm. If the accuracy is too low, the solution may no longer be faithful to the modeled dynam-
ics; if too high, the computation time will become prohibitive.

In musculoskeletal models, we are primarily concerned with inaccuracies in the model’s state (e.g., joint angles and speeds, and muscle fiber lengths and activations). Errors in the state can result in erroneous forces, which lead to noisy acceleration esti-
mates, poor numerical performance, and untrustworthy future states. For example, consider the case where one muscle in a model has a short optimal fiber length, say 5 mm. If the integration accuracy is 0.001, then each state variable is accurate to approxi-
amately three significant figures, or about 1 mm if lengths are stored in meters. In this case, an error of 0.001 m amounts to 20% of the muscle’s fiber length, which is sufficient to cause the mus-
cle to switch from generating no passive forces to high passive forces, or from having a small force-generating capacity based on

Table 1 Summary of tests, dependencies, and validation best practices when building a musculoskeletal model or adapting a model to a new use. The first column lists tests to perform when building or adapting a musculoskeletal model. Each test should be performed throughout the range of motion for which the model will be analyzed. The second column lists the model parameters, experimental inputs, and antecedent output variables (the latter, in bold) being tested. Errors in each test can also stem from errors in each of the parameters and inputs listed in the second column (i.e., errors will propagate). Variables or inputs in italics are those to which an output is particularly sensitive. The third column lists best practices for validating the results of each test. These are recommended guidelines—if they cannot be achieved, the researcher should provide justification and perform sensitivity analyses to ensure that conclusions drawn from analysis of the model are robust given the errors or inaccuracies. The final column refers to sections of the article that provide more detail.

<table>
<thead>
<tr>
<th>Musculoskeletal model test</th>
<th>Parameters and inputs tested</th>
<th>Validation best practices</th>
<th>Sec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model kinematics: compute model kinematics through each joint’s range of motion</td>
<td>Joint definitions (e.g., joint location, orientation, and type); body segment lengths</td>
<td>Joint ranges of motion match experimental data. Modeled joints reproduce experimental motion from bone pin, cadaver, or imaging data to within measurement error.</td>
<td>3.1, 3.2</td>
</tr>
<tr>
<td>Muscle moment arms: compute muscle moment arms throughout the model’s range of motion</td>
<td>Muscle geometry (attachment and via points, wrapping surfaces); mode kinematics</td>
<td>Moment arms are within 2 SD of experimental data measured by tendon excursion or MRI.</td>
<td>3.2</td>
</tr>
<tr>
<td>Forces from muscle–tendon dynamics model and constituent components, including active and passive muscle fiber force, tendon force and strain, muscle fiber lengths and velocities: use computational muscle–tendon dynamics model to replicate isolated muscle experiment</td>
<td>Muscle–tendon dynamics model and its parameters, including maximum isometric force, pennation angle, optimal fiber length, passive muscle stiffness, force–length–velocity relationships, tendon stiffness, and tendon slack length</td>
<td>Difference between predicted and experimental muscle forces are within 10% of measured muscle force.</td>
<td>3.3</td>
</tr>
<tr>
<td>Passive joint moments: calculate the net moment generated by muscles and other modeled forces (e.g., ligaments) throughout joint range of motion when muscles have zero activation</td>
<td>Model kinematics: muscle moment arm; muscle geometry; muscle–tendon dynamics model and its parameters, including passive force–length curve, tendon slack length, tendon stiffness, and maximum isometric force.</td>
<td>Passive moment curves are within 2 SD of experimental data.</td>
<td>3.2</td>
</tr>
<tr>
<td>Maximum net joint moments: calculate the net moment generated by muscles and other modeled forces (e.g., ligaments) throughout joint range of motion when agonist muscles have maximum activation input and all other muscles have zero activation</td>
<td>Model kinematics: muscle moment arm; muscle geometry; muscle–tendon dynamics model and all related parameters</td>
<td>Maximum moment curves are within 2 SD of experimental data. Joint moments generated by the model during submaximal activation should also be tested (see Table 3).</td>
<td>3.2</td>
</tr>
</tbody>
</table>
its force–length relationship to a capacity near its maximum. This behavior can cause the numerical solution to become unstable (i.e., chatter) and fail. We recommend choosing an accuracy corresponding to the largest change in a state variable that would result in an insignificant change in state derivatives.

Other commonly employed methods for determining unknowns in a model or simulation are root-solving and optimization algorithms. These methods are iterative and generally use derivatives of an objective function to steer the approach toward a solution. Optimization and root-solving methods are sensitive to several factors: derivative precision, function noise, and, the convergence criteria used to terminate the search. Derivatives are often calculated numerically by perturbing the objective function, with smaller perturbations yielding higher-precision estimates of the derivative. However, if there is noise or imprecision in the evaluation of the objective function, then seemingly high-precision derivatives from these small perturbations will predominantly be a measure of the local noise and, thus, will be grossly inaccurate. For these algorithms to proceed efficiently and reliably, the perturbation size must be large enough to capture functional behavior and not merely noise. We recommend choosing a perturbation size that is substantially (e.g., an order of magnitude) larger than the known or estimated accuracy of the objective function [150]. A convergence criterion is used to assess how close the optimization or root-finding algorithm is to a final solution and specifies when the algorithm will terminate. We recommend verifying that the required convergence has been achieved for any such computation by examining the plot of objective function value versus iteration number.

3.7 Summary of Best Practices for Verification and Validation. In Secs. 3.1 to 3.6, we reviewed the current state of the art for selecting a modeling and simulation framework, and for verifying and validating its many possible outputs. The majority of modeling and simulation research falls into three categories: (1) building or adapting a musculoskeletal model for a particular application; (2) generating an inverse dynamic simulation based on measured experimental kinematics and/or kinetics; and (3) generating a forward dynamic simulation, given a controlled or estimated control signals. We, therefore, also provide in Tables 1, 2, and 3 a synopsis of the typical tests or outputs, dependencies, and validation best practices for each of these research categories. These tables can be used to guide the development of new models and simulations. The guidelines they contain will continue to evolve as we create new tools for modeling and simulation and new techniques and datasets for verification and validation.

4 Verification and Validation Challenges

The resources and best practices for verification and validation of models and simulations must continue to improve as the field progresses. We have identified several key challenges for verification and validation.

4.1 Creating Gold Standard Datasets and Benchmark Problems. Gold standard datasets include comprehensive, high-fidelity experimental data collected using techniques inaccessible to many researchers (e.g., forces from instrumented joint replacements or motion from bone pin markers). These data should be made publicly available so others can use them to verify and validate both existing and new models. The Grand Challenge to Predict Knee Loads [78] is one example of a publicly available gold standard dataset. In addition to a comprehensive set of anthropometric and motion data, the Grand Challenge also includes a clear, validation-focused problem: predicting internal knee joint loads in

Table 2 Summary of dependencies and validation best practices for common outputs from an inverse simulation, assuming the model under study has already been validated. The first column lists typical simulation outputs. The second column lists the model parameters, experimental inputs, and antecedent output variables (the latter, in bold) upon which each output depends. Errors in each output can also stem from errors in each of the dependencies listed in the second column (i.e., errors will propagate) and from errors in the model (see Table 1). Variables or inputs in italics are those to which an output is particularly sensitive. The third column lists best practices for validation. These are recommended guidelines—if they cannot be achieved, the researcher should provide justification and perform sensitivity analyses to ensure that conclusions are robust given the errors or inaccuracies. The final column refers to sections of the article that provide more detail.

<table>
<thead>
<tr>
<th>Inverse simulation output</th>
<th>Dependencies</th>
<th>Validation best practices</th>
<th>Sec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kinematics of joints, bodies, and points of interest</td>
<td>Joint definitions; body segment lengths; sensor/motion capture marker placement; measurement noise; motion data processing</td>
<td>Difference between experimental and model sensors/markers is within measurement error. Kinematics are within 2 SD of published data for similar motion.</td>
<td>3.2, 3.1</td>
</tr>
<tr>
<td>Kinetics of joints and bodies, including net joint moments and residual forces and moments</td>
<td>Kinematics; body segment inertial parameters; joint definitions; measured forces and moments, including points of application, measurement noise, and data processing</td>
<td>Residual forces are &lt; 5% of the magnitude of net external force (peak and RMS). Residual moments are &lt; 1% COM height times the magnitude of the net external force (peak and RMS). Net joint moments are within 2 SD of published data for similar motion.</td>
<td>3.1</td>
</tr>
<tr>
<td>Muscle–tendon unit path lengths and lengthening speeds</td>
<td>Kinematics; muscle geometry (attachment and via points, wrapping surfaces); joint definitions.</td>
<td>Compare muscle paths and kinematics to imaging data (e.g., MRI).</td>
<td>3.2</td>
</tr>
<tr>
<td>Muscle forces and/or activations</td>
<td>Tracking requirements (e.g., kinematics and joint moments); optimization criterion (e.g., minimize sum of squared activations); muscle moment arms; muscle force–generating capacity as a function of path lengths and speeds; joint definitions (e.g., degrees of freedom and other passive structures modeled)</td>
<td>On and off timings of muscle activity are within electromechanical (EMG-to-force) delay (~100 ms) of experimental EMG. Muscle activity and EMG curves are qualitatively similar. Net joint moments from muscles and other modeled structures account for 95% of tracked joint moments. Indirect: joint reaction force criteria are satisfied (see below).</td>
<td>3.3</td>
</tr>
<tr>
<td>Joint reaction forces</td>
<td>Kinematics; kinetics; muscle forces; joint definitions; segment inertial properties; muscle geometry</td>
<td>Forces are within 2 SD of experimental joint forces (e.g., instrumented implants) for similar motion.</td>
<td>3.2, 3.3, 3.1.3</td>
</tr>
</tbody>
</table>
a blinded manner. We need additional experimental datasets for validation and similar Grand Challenges to engage the field. For example, comprehensive datasets from animal models could provide full musculoskeletal geometry data and experimentally measured internal muscle and joint forces during motion. Comprehensive human datasets could include traditional motion capture data for a range of activities, along with internal muscle–tendon dynamics from ultrasound (e.g., Ref. [108]) and sarcomere lengths from recently developed optical microendoscopy techniques [151]. Accompanying these datasets should be a clear set of problems, such as predicting internal joint and muscle forces in the animal models or predicting changes in neural control for different activities (e.g., walking with and without an assistive device) in humans. Modeling and simulation will benefit from a clear set of benchmark problems to help establish baselines and generate new tools that can immediately be used and extended by others.

4.2 Sharing Models, Data, and Simulation Tools. A related challenge for the field is further incentivizing and providing resources for sharing published models, data, and simulation tools with the wider research community. Sharing, although time intensive and not directly publishable, benefits the field by allowing others to reproduce, extend, and continue to verify and validate research products. Sharing can also benefit individual researchers by increasing awareness and subsequent citations of publications. Funding agencies have also increasingly pushed for open access to the results of sponsored research. We hope that the incentives and resources for sharing will continue to grow in the future, for example, by encouraging open access as part of grant applications and journal publications. Contribution to and use of open source software should also be encouraged as a way of disseminating new tools that can immediately be used and extended by others.

4.3 Developing Tools to Help Automate Verification and Validation. Verification tests are an integral part of current software packages for modeling and simulation. These test suites should continue to grow with the field, as new benchmark problems and gold standard datasets are made available and new modeling and simulation approaches are developed. The field also needs more tools to aid the model validation process, such as modern algorithms for efficient sensitivity testing (compatible with and/or embedded in NMS modeling packages) and automated tools to compare simulation results to established standards for kinetics, kinematics, and muscle activity and dynamics.

4.4 Learning and Teaching Others What’s Inside the “Black Box”. Knowledge of the physics and biology of the musculoskeletal system and its movement, as well as how existing tools transcribe real world systems into computational models,
the foundation for verification and validation. The verification and validation process demands an intimate knowledge of the capabilities and limitations of models and simulations. Members of the community must continue to expand their knowledge as new models, simulation techniques, and experimental modalities are developed. We also encourage the community to create and share teaching materials (e.g., hands-on tutorials and videos) for audiences from clinicians who are new to NMS modeling to advanced users and software developers.

4.5 Performing Excellent Research With Modeling and Simulation. Excellent research demonstrates the value of modeling and simulation by generating hypotheses and making predictions that impact real-world challenges. The set of available tools and datasets for building models, generating simulations, and validating predictions is rapidly growing. These advances are paving the way for modeling and simulation to be used as tools for planning musculoskeletal surgery, designing training programs to reduce sports injuries, and prototyping powered prostheses that aid locomotion. As a field, we must continue to identify and solve pressing real-world challenges where modeling and simulation are used to uncover phenomena not accessible via observation or experiment alone.

5 Conclusions
Musculoskeletal models and dynamic simulations of movement provide powerful tools to study neuromuscular coordination, analyze athletic performance, and estimate internal loading of the musculoskeletal system. Simulations can also be used to identify the essential physical phenomena, and that the conclusions we draw from these simulations are trustworthy. The guidelines provided here serve as a framework for achieving these goals.

Acknowledgment
We thank Michael Sherman, James Dunne, Edith Arnold, Samuel Hamner, and Bradley Davidson for their assistance in preparing this manuscript. This work was supported by NIH Grant Nos. U54 GM072970 and R24 HD06590, and the DARPA Warrior Web project.

References