Credibility Assessment of Implicit Finite Element Models for Energy Return in a Prosthetic Foot per NASA-STD-7009A

Elena V. Markov, PhD[a], Jared L. Connors, MS[1], Sahana Iyer, PhD[1], Miguel A. Rojas, MS[1]

[a]*Biomechanics and Computation Laboratory, OrthoMotion Research Group, Boston, MA, USA,*

Abstract

Background: Energy-return prosthetic feet store and release elastic work during stance and push-off. Implicit finite element analysis (FEA) is routinely used to design and evaluate such devices, yet decision-grade credibility requires structured evidence. Objective: To assess the credibility of two implicit FEA models of a composite-keel prosthetic foot—one high-fidelity (HF) and one reduced-order (RO)—for two mechanisms: (M1) mid-stance-to-toe-off elastic work return during level walking, and (M2) high-rate push-off rebound during sprint initiation. Methods: We executed a NASA-STD-7009A-aligned assessment using an internal 0–9 scale to rate evidence across eight factors, one table per model. Model HF comprises a nonlinear 3D solid FEA of the composite keel and elastomeric interface with layered orthotropy, viscoelasticity, and frictional contact. Model RO employs equivalent orthotropic shells and Timoshenko beams with a Winkler contact foundation for rapid design exploration. Both models were solved with an implicit Newton scheme and line search, with substeps resolving gait-like and sprint-like load histories. We evaluated parameter match to experiments, mesh/time resolution, and predictive agreement against bench tests and in-shoe measurements from human participants. Results: For M1, Model HF predicted energy return within 1.6 percentage points of bench measurements (mean bench 87% ± 4%, HF 88.6%), with force–deformation RMSE 4.2% of span and hysteresis error 0.32 J. Mesh refinement from 3.5 to 2.5 mm changed peak stored energy by 1.1% and halving the time step altered integrated work by 0.5%. For M2, HF matched energy return within 2.8 percentage points (mean bench 83% ± 5%, HF 85.8%), with RMSE 6.1% of span; time-step sensitivity increased, but remained below 1.2% for integrated work at 0.2 ms. RO reproduced M1 effective stiffness within 7.5% and energy return within 4.9 percentage points; for M2, energy return error was 7.5 percentage points, and stiffness deviation 12%. Evidence for parameter alignment, material law selection, contact modeling, and process control was stronger in HF than RO. Conclusions: Model HF attained higher credibility across both mechanisms, with stronger agreement to data and tighter numerical resolution control. RO delivered acceptable screening accuracy for M1, with limitations for M2 tied to rate-dependent viscoelasticity and simplified contact. We outline targeted actions to elevate credibility centered on high-rate viscoelastic characterization, expanded lot coverage in test articles, and automated convergence auditing. The assessment tables enumerate scores for every factor–mechanism pairing per model, and the discussion provides a single, plain-language finding for each assessed factor away from the sections where the factors are named.

Keywords: finite element analysis, prosthetic foot, credibility

1. Introduction

Prosthetic feet designed to return elastic energy during stance have matured over several decades, moving from laminated leaf-springs toward complex composite keels that integrate elastomeric interfaces and tuned toe geometries. Modeling and simulation are central to this evolution. Among available approaches, implicit nonlinear finite element analysis (FEA) allows engineers to quantify load–deformation behavior, strain energy storage and release, hysteretic losses, and effective push-off stiffness with stability across quasi-static and moderately dynamic conditions. However, as the resulting predictions are increasingly used to make design, verification, or performance claims, model credibility must be demonstrated with traceable evidence that is relevant to the intended use.

NASA-STD-7009A defines a structured approach to assessing the credibility of models and simulations. While created for aerospace applications, its factor-based framework is broadly applicable to medical and assistive device modeling where quantitative decisions must be informed by simulation. Here, we apply a NASA-STD-7009A-aligned assessment to two implicit FEA models used for an energy-return prosthetic foot: a high-fidelity (HF) 3D solid model of the composite keel and elastomeric interface, and a reduced-order (RO) beam–shell model created for design space exploration. We consider two mechanisms and their quantities of interest (QoIs): elastic energy storage from mid-stance and release at toe-off during level walking (M1), and high-rate rebound during sprint initiation (M2).

Our goals were to: (1) define the modeling form and solver settings suitable for each mechanism; (2) establish how experimental measurements map to model inputs and outputs; (3) quantify numerical resolution effects; and (4) compare predictions to bench and in-shoe measurements gathered under conditions matched to the model inputs. We then rate the evidence for eight credibility factors using a private 0–9 scale and report the scores as one table per model. The synthesis includes a succinct, plain-language statement of what we conclude for each assessed factor, placed away from the factor-named sections and grounded by the numeric results reported in the body. Two additional topics—independent technical review and prior usage—are mentioned for context without formal findings in accordance with the plan.

Scoring philosophy (private 0–9 scale) — gradation.

a.
Uncalibrated; no rationale for numeric levels; ad hoc thresholds.

b.
Descriptive ladder exists; qualitative anchors given but not tied to testable evidence.

c.
Quantitative anchors linked to evidence types and acceptance bands; examples provided.

d.
Scale fully operationalized with traceable evidence classes, numeric acceptance limits, and examples spanning low to high maturity.

2. Device Description and Energy-Return Mechanisms

The device under study is a commercially representative energy-return foot composed of a carbon/epoxy composite keel bonded to an elastomeric interface. The keel transitions from a curved forefoot with a tapered toe region to a stiffer mid-foot and a posterior segment that mates to a pylon adapter. The composite lay-up is tailored to supply bending compliance in the sagittal plane while limiting transverse shear and providing torsional stability. The elastomeric interface, positioned between the keel and the socket-side adapter, accommodates local deformations and dissipates energy to shape the feel of the foot. The assembly includes a toe bumper made of a higher durometer elastomer to control end-range stiffness as the toe contacts the ground.

Two mechanisms motivate this credibility assessment. Mechanism M1 centers on energy exchange during level walking. As stance progresses from mid-stance to terminal stance, the keel stores strain energy and then releases it near toe-off. The QoIs for M1 include the input work during loading, released work during unloading, the fraction of input returned (the energy return percentage), the shape of the load–deformation curve including the slope in a mid-range segment, and the hysteresis loss represented by the area enclosed by the loop. Mechanism M2 focuses on push-off when the loading rate is substantially higher. During sprint initiation, the composite and elastomer respond with rate-dependent stiffness and loss factors that can influence propulsion and the transmission of peak forces to the residual limb. The M2 QoIs are similar to M1 but measured at higher sweep rates and with loading histories designed to mimic sprint-like ground reaction force profiles. Together, M1 and M2 span the common design and evaluation envelope for this class of devices.

Bench testing to characterize these mechanisms uses a custom fixture that allows repeatable load paths representative of compressive and bending loads seen in gait and sprint starts. A curved anvil contacts the toe region to capture realistic forefoot support while a six-axis load cell under the anvil and a second load cell at the pylon provide net load measurement. For in-shoe measurements, instrumented insoles record plantar pressure distribution and center-of-pressure progression, while motion capture provides joint kinematics, and a treadmill with integrated force plates supplies global ground reaction forces. These measurements define the load–deformation loop and enable calculation of input and returned work by numerical integration of force–displacement curves aligned to stance phase.

For this paper, we restrict attention to one geometry family, sizes M26–M28, at a nominal user mass of 75–85 kg. The toe bumper and interface elastomer hardnesses are fixed within the manufacturer's standard offering for the size range. Using this set of components mitigates confounding differences in geometry and durometer that would otherwise broaden the range of outputs and complicate the comparisons necessary for credibility assessment.

Relevance of the quantities of interest — gradation.

a.
Model outputs are proxies with unclear ties to decisions; no mapping to user outcomes.

b.
Outputs are plausible but only qualitatively tied to performance; limited rationale for selection.

c.
Outputs are quantitatively defined and traceably mapped to bench or in vivo measures that underlie decisions.

d.
Outputs are formally traced to acceptance criteria linked to clinical or biomechanical endpoints; decision rules are pre-specified.

3. Computational Models: Formulation and Solvers (Implicit FEA)

Model HF: The high-fidelity model resolves 3D geometry for the composite keel, toe bumper, and interface elastomer. The composite is represented with layered solid elements (hexahedral-dominant mesh with tetrahedral fillers where necessary) integrating lamina properties through the thickness. Each lamina follows linear orthotropic elasticity in the fiber-dominated regime with matrix viscoelasticity introduced via a generalized Maxwell model applied to the transverse moduli and shear moduli. Representative properties for the unidirectional lamina at 23 °C are: E1 = 125 GPa, E2 = E3 = 9.0 GPa, G12 = G13 = 5.2 GPa, G23 = 3.1 GPa, nu12 = 0.28, density 1600 kg/m3. Viscoelasticity is captured with a two-term Prony series acting on E2, E3, G12, G13, and G23, with g1 = 0.08 (tau1 = 0.10 s) and g2 = 0.03 (tau2 = 1.0 s), established from DMA coupons cut transverse to the fiber. Ply thickness is 0.18 mm with a quasi-isotropic outer stack around a fiber-biased core to achieve sagittal compliance. The elastomeric interface is modeled using an incompressible Ogden N=2 hyperelastic form with Prony viscoelasticity: mu1 = 0.60 MPa (alpha1 = 1.3), mu2

Figure 1: Schematic of the composite-keel prosthetic foot and the two modeled mechanisms: M1 mid-stance-to-toe-off energy exchange during level walking and M2 high-rate push-off during sprint initiation. The figure depicts the keel and elastomeric interface, load paths, and representative force–deformation loops.

= 0.15 MPa (alpha2 = 5.0), D1 corresponding to an effective bulk modulus of 800 MPa, with viscoelastic g1 = 0.20 (tau1 = 0.02 s). The toe bumper uses a stiffer Ogden set: mu1 = 1.2 MPa (alpha1 = 1.2), mu2 = 0.40 MPa (alpha2 = 4.5), with g1 = 0.10 (tau1 = 0.04 s).

Contact between the forefoot and ground anvil is enforced with finite sliding, penalty contact, and isotropic friction coefficient mu = 0.70 with a Stribeck-like regularization to avoid stick–slip chatter at low velocities (transition velocity 0.01 m/s). The pylon–adapter joint is tied to a rigid reference, with prescribed vertical and fore–aft kinematics applied consistent with bench and in-shoe measured boundary motions. The solver employs full Newton–Raphson with line search and automatic time increment control, a quasi-static step for M1 (nominal substep size 5 ms, minimum 0.2 ms) and a dynamic implicit step for M2 (nominal substep 1 ms, minimum 0.1 ms) including mass and Rayleigh damping (alpha = 0, beta = 2e5 s) to control high-frequency oscillations without artificially damping the push-off.

Model RO: The reduced-order model replaces the 3D keel solid with a mid-surface orthotropic shell calibrated to recover bending and torsional stiffness about the principal axes, joined to a Timoshenko beam element representing the posterior segment. The toe region is discretized with a fine shell mesh to capture contact footprint evolution. The elastomeric interface is represented by a distributed set of nonlinear springs in compression and shear whose force–displacement curves are identified by matching Model HF responses in a set of calibration load cases. Ground interaction is a Winkler foundation with nonlinear springs distributed over the plantar surface and Coulomb friction with an effective coefficient mu = 0.68; micro-slip is introduced by a tangential penalty parameter tuned to replicate energy dissipation observed in Model HF. This structure reduces the number of degrees of freedom by roughly 85% relative to Model HF, enabling rapid parametric sweeps.

Both models were built and solved in a commercial implicit FEA package with user subroutines for viscoelasticity and friction regularization. Geometry was imported from the same native CAD master; mesh generation followed common templates specific to each model. For M1, applied boundary histories reflect a stance duration of approximately 0.62 s at a walking speed of 1.25 m/s; for M2, the push-off event is isolated into a 0.18 s window with a fast-rising force–displacement input. Forces and moments at the pylon are constrained to measured in-shoe kinematic displacements to avoid double-counting interaction with the residual limb–socket system. Model outputs are sampled in substeps, and work is computed by trapezoidal integration of the contact force projected along the toe displacement trajectory.

Both models underwent sensitivity evaluations to identify influential parameters, including the transverse viscoelastic terms of the composite, the Ogden parameters for the interface, and the friction coefficient for the toe anvil interaction. The top sensitivities for energy return were, in order: interface viscoelastic g1 (M2 dominant), toe friction coefficient (M1 and M2), and composite transverse shear viscoelasticity. These findings were used to prioritize experimental characterization and guide uncertainty quantification bounds reported later.

Model inputs — gradation.

a.
Key quantities are undocumented or tuned to match outputs; no traceable origin.

b.
Inputs partially documented; some are literature values without justification.

c.
Major inputs measured or derived from device-specific tests with stated tolerances; limited tuning.

d.
All influential inputs measured on test articles or substantiated by targeted characterization; priors and uncertainties documented; no post hoc tuning.

4. Model Inputs and Equivalency of Input Parameters

Mapping between experimental configuration and model parameters is central to avoiding hidden discrepancies. The lamina properties used in the composite keel came from coupon tests manufactured with the same pre-preg lot and cure cycle as the tested keels. Ply thickness measured from micrographs across five cross-sections averaged 0.180 ± 0.006 mm, aligning with the 0.18 mm per ply used in the model. The elastomeric interface coupons were die-cut from the process-trim region adjacent to the test article and tested in uniaxial compression and simple shear over strain rates from 0.05 to 20 s1 to cover M1 and M2 regimes. Ogden parameters were regressed simultaneously to shear and compression data with a 2-norm target and Tikhonov regularization to constrain non-identifiability; the selected values reproduced stress–strain curves within ±6% over the tested range. The Prony term g1 and time constant tau1 were extracted from stress relaxation tests using a Prony series fit (R2 = 0.992 for compression, 0.987 for shear), which informed the implementation in both models.

The friction coefficient between the toe region and anodized aluminum anvil was measured in a custom reciprocating rig under normal loads of 300–900 N and sliding speeds from 0.005 to 0.2 m/s, with mean mu = 0.68 ± 0.04. The HF model used mu = 0.70 to fall within the measured range; the RO model used mu = 0.68 due to its Winkler foundation parameterization. The mechanical attachment of the foot to the adapter in the bench rig was torqued to 12 N·m, with a corresponding clamping pressure of approximately 11 MPa. In the model, the adapter was tied to a rigid reference; the effect of clamp compliance was evaluated by introducing a torsional spring in sensitivity runs and found to contribute less than 0.2 percentage points to energy return across both mechanisms.

Boundary conditions were aligned to in-shoe measured kinematics and the bench rig geometry. The vertical displacement of the pylon reference was prescribed to match the stance-phase displacement from motion capture averaged across participants for M1 and across sprint initiation trials for M2, appropriately scaled to the M26–M28 sizes by anthropometric normalization to foot length. The fore–aft displacement was similarly prescribed, while toe contact with the anvil emerged from the contact solution. Post-processing included rigid-body removal to ensure that work calculations for energy storage and release occurred in the deforming substructure, not in rigid motion.

To maintain consistency across M1 and M2 conditions, we selected input sets that are portable between bench and in-shoe contexts. For example, the angle between the pylon and ground frame at mid-stance was within 1.2 degrees between bench replay and in-shoe stance, a level that kept torsional moments comparable. We verified that the resulting contact footprint in the HF model matched the pressure insole patterns to within 8% of area after morphological dilation to account for insole sensor spacing.

Equivalency of input parameters — gradation.

a.
No demonstration that the model's inputs correspond to the test or use case conditions.

b.
Qualitative mapping provided; some inputs likely mismatched or unstated tolerances.

c.
Quantitative crosswalk establishes alignment within stated tolerances for major inputs; minor adjustments documented.

d.
Formal equivalence demonstrated with pre-defined tolerances, direct measurement crosswalk, and residual differences bounded with sensitivity analysis.

5. Test Samples and Data Pedigree

Physical test articles included six feet drawn from two production lots across sizes M26–M28. For each size, two articles were selected: one from an earlier position in the cure cycle and one from a later position to capture intra-lot variation. Serial numbers and cure batch identifiers were recorded. All articles underwent conditioning equivalent to 2000 cycles at 1.0 body weight on the bench fixture prior to data collection to stabilize initial seating and micro-settlement. The toe bumper and interface elastomer were the standard hardness for the size range, with hardness verified using Shore A durometers (interface: 57 ± 1 Shore A; toe bumper: 72 ± 1 Shore A).

Bench tests for M1 used a stance-mimicking displacement input with a 0.62 s cycle time; four repeats per article were recorded, with loading up to a peak toe displacement corresponding to 1.2 body weight reaction forces. For M2, a fast ramp-and-hold sequence reached a comparable peak displacement in 0.09 s with a total cycle of 0.18 s; four repeats per article were recorded. The six-axis load cell at the anvil (capacity 5 kN, calibration uncertainty ±0.3%) and the pylon load cell (capacity 2 kN, calibration uncertainty ±0.3%) were calibrated within six months of the tests by an ISO/IEC 17025-accredited laboratory; the calibration certificates were archived with the test data. Environmental conditions were recorded (22–24 °C, 35–45% RH). Data acquisition used 2 kHz sampling for force signals and 500 Hz for displacement encoders.

In-shoe data came from nine unilateral transtibial prosthesis users (7 male, 2 female; 31–55 years; mass 70–88 kg; K3–K4 mobility) under IRB approval. For M1, participants walked at 1.25 m/s ± 0.05 m/s; for M2, five of the nine performed a sprint initiation on an instrumented treadmill from a static start to the first propulsion step. Instrumented insoles recorded pressure at 100 Hz; motion capture at 200 Hz; treadmill ground reaction forces at 1000 Hz. The prosthetic components used in the trials were the same model family as the bench articles, with size matched to user foot length. The specific feet instrumented for in-shoe use were not the same serial-numbered units as on the bench to avoid overuse; however, all came from the same two production lots as the bench articles.

Data management followed a documented procedure. Raw and processed data were stored in a controlled repository with read-only access after the initial 72-hour window for QA checks. File naming included the serial number, participant ID, and test condition. Analysis scripts were versioned alongside the data, with hash-verifiable inputs and outputs. Reproducibility audits were conducted by re-processing 10% of the trials from raw signals to derived QoIs. These details support traceability of the measurement sources, sensor calibrations, processing steps, and links to the modeled components.

Data pedigree — gradation.

a.
Unknown or undocumented origin; no calibration or chain of custody.

b.

Partial documentation; some sources or calibrations missing.

c.
Complete origin, calibration, and processing history for principal datasets; partial chain-of-custody.

d.
Full traceability from source to analysis with accredited calibrations, preserved chain-of-custody, and reproducibility checks.

6. Relevance of the Quantities of Interest

The decision to accept or refine the design relies on how well the model reports the work exchange during stance. For level walking, the energy returned at toe-off relative to the energy stored during loading is one of the primary determinants of perceived responsiveness and has been associated with metabolic cost differences across prosthetic feet. Therefore, the percentage of input work returned, the effective mid-range stiffness of the forefoot, and the hysteresis loss are the central QoIs for M1. These are computed consistently across bench and in-shoe contexts by integrating the anvil reaction force over the toe displacement during the appropriate phases of the cycle, with the loop area corresponding to losses.

For sprint initiation, the rate of loading increases and the elastomer contributes proportionally more to loss. The returned work fraction at high rate influences peak push-off forces and the impulse delivered by the foot. Thus, for M2, we retain the same QoIs but focus on the sensitivity to rate, resolving work within the fast ramp and immediate unloading. Because push-off in sprint initiation is dominated by forefoot mechanics with a shorter time budget for error correction, small differences in damping or contact micro-slip can produce noticeable changes in performance. The models must therefore resolve time scales on the order of 0.1 s or less, and their stiffness and damping formulations cannot obscure the working range of these QoIs.

These choices for QoIs connect naturally to bench and in-shoe measurement signals: they are observable, calculable with standard methods, and relevant to design and clinical decisions. Specifically, the acceptance criteria employed during this study were prespecified as: for M1, energy return within ±3 percentage points and effective stiffness within ±10% relative to bench measurements; for M2, energy return within ±5 percentage points and stiffness within ±15%. The looser target for M2 reflects expected rate-dependent uncertainties that are more difficult to control.

Relevance of the quantities of interest — gradation.

a.
Outputs loosely connected to the decision; not directly measurable.

b.
Outputs measurable but not aligned to decision thresholds.

c.
Outputs measurable and linked to decision thresholds that are informally defined.

d.
Outputs directly tied to pre-specified acceptance criteria reflecting biomechanical endpoints.

7. Discretization Error: Mesh Convergence and Time-Step Study

We evaluated spatial and temporal resolution effects by systematic refinement. For Model HF, three mesh densities were generated using the same meshing algorithm and growth controls: coarse (5.0 mm target edge length in the forefoot, 6–8 elements through thickness of the keel laminate), medium (3.5 mm, 10–12 elements through thickness), and fine (2.5 mm, 14–16 elements through thickness). The elastomeric interface followed the same in-plane sizes to maintain contact footprint consistency and had 6–10 elements through thickness due to higher deformation. For Model RO, shell element sizes of 8.0, 5.0, and 3.5 mm were used in the forefoot, with beam element subdivision adjusted to match curvature; Winkler foundation point spacing matched shell edge lengths.

For M1, Model HF showed monotonic convergence in peak stored energy: 20.9 J (coarse), 21.2 J (medium), 21.4 J (fine). Using Richardson extrapolation with an assumed second-order convergence rate for energy, we estimated an asymptotic value of 21.6 J, implying a relative error of 1.1% at the fine mesh. The energy return percentage at the fine mesh was 88.6%, compared to 88.3% (medium) and 87.8% (coarse). Halving the nominal time increment from 5 ms to 2.5 ms and then to 1.25 ms at the fine mesh changed integrated work by 0.5% and the energy return percentage by 0.3 percentage points. For M2, at the fine mesh with a nominal time increment of 1.0 ms, halving to 0.5 ms and 0.25 ms altered integrated work by 0.8% and 1.1%, respectively; peak force and unloading slope were more sensitive, shifting by up to 1.9% at 0.25 ms compared to 1.0 ms. We found no evidence of numerical damping beyond the specified Rayleigh term using this time resolution.

For Model RO, mesh refinement influenced stiffness prediction more than work integrals. In M1, the effective stiffness over a mid-range segment increased from 3.4 kN/m (coarse) to 3.6 kN/m (fine), while energy return moved from 84.2% to 85.8%. In M2, effective stiffness changed from 4.1 to 4.6 kN/m across the refinement, with energy return moving from 76.3% to 78.5%. Temporal refinement in RO had limited effect because the foundation springs dominated contact behavior; substep reduction from 5 to 1.25 ms changed energy return by less than 0.6 percentage points in M1 and 1.0 percentage point in M2.

Based on these studies, we used the fine mesh for all reported HF results and the fine shell-beam discretization for RO. We retained the 2.5 mm in-plane element size in HF as a practical limit balancing the cost and the sub-percent error in the QoIs, and we used minimum time steps of 0.2 ms for M2 to bound temporal discretization error below 1.2% in integrated work. The reported differences between model predictions and experiments are therefore dominated by modeling assumptions and input uncertainties rather than numerical resolution artifacts.

Discretization error — gradation.

a.
Single mesh/time step; no evidence of resolution adequacy.

b.
Two-level refinement; qualitative judgment of sufficiency.

c.
Multi-level refinement with quantitative changes reported for QoIs; extrapolation for at least one QoI.

d.
Comprehensive space–time refinement with extrapolation and error bounding for all QoIs of interest.

8. Output Comparison: Bench Tests and In-Shoe Measurements

Model HF versus bench tests (M1): Across six articles and four repeats per article, the bench-measured mean energy return was 87% with a standard deviation of 4%. For the fine mesh/time-step settings, Model HF predicted a mean of 88.6% across the corresponding simulated configurations (using measured friction coefficients within their experimental range). The absolute difference was 1.6 percentage points on average. The force–deformation curves over the mid-range showed an RMSE of 4.2% of the force span; the loop area corresponding to hysteresis differed by 0.32 J on average relative to a mean loop area of 2.8 J. Effective stiffness measured between 30% and 70% of peak load differed by 6.1% on average. Agreement to in-shoe derived work estimates was within 2.8 percentage points for energy return when mapping to individual participant trials scaled by anthropometry.

Model HF versus bench tests (M2): For sprint-like loading, bench-measured mean energy return was 83% with a standard deviation of 5%. HF predicted a mean of 85.8%, for a 2.8 percentage-point difference. The force–deformation RMSE increased to 6.1% of span, reflecting higher sensitivity to rate-dependent parameters and contact micro-slip. The loop area error averaged 0.58 J relative to a mean of 5.8 J. Effective stiffness over the mid-range differed by 10.2% on average. In-shoe derived estimates for participants who performed sprint initiation agreed with bench and model within 4.2 percentage points, though variability was higher trial-to-trial due to initiation strategies.

Model RO versus bench tests (M1): RO predicted energy return averaging 82.1% at coarse, 84.9% at medium, and 85.8% at fine discretization; we report the fine prediction in the comparison. Against the bench mean of 87%, the difference was 4.9 percentage points. Mid-range stiffness differed by 7.5%. The RO loop area deviated by 0.71 J relative to the 2.8 J mean. The force–deformation RMSE relative to bench was 7.8% of span. The mismatch primarily arose from the simplified contact and the omission of 3D stress redistribution in the toe region.

Model RO versus bench tests (M2): RO predicted an energy return of 75.5% at its fine discretization. Against the bench mean of 83%, the difference was 7.5 percentage points. The RMSE in the force–deformation loop increased to 11.2% of span, with mid-range stiffness deviating by 12%. The loop area was underpredicted by 1.3 J relative to the 5.8 J mean. The greater shortfall in RO under sprint initiation reflects its simplified rate dependence in the interface and the lumped nature of the Winkler foundation under fast loading where micro-slip transitions are not explicitly modeled.

Across both models, the most influential uncertainties in the comparisons were the friction coefficient (±0.04 measured SD yielded ±0.6–0.9 percentage-point spread in predicted energy return) and the interface viscoelastic g1 (±0.03 plausible variation yielded ±0.7–1.4 percentage-point spread across M1–M2). These uncertainty bands do not overturn the observed patterns of agreement for either model.
Output comparison — gradation.

a.
Qualitative or anecdotal comparison; no quantitative metrics.

b.
Limited quantitative comparison on a subset of QoIs or conditions.

c.
Quantitative comparison on all QoIs under representative conditions with error metrics.

d.
Broad quantitative comparison across mechanisms and conditions with pre-specified acceptance bands and uncertainty analysis.

$$\text{CI\_\{score\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \tag{1}$$

$$\text{\textbackslash varepsilon\_\{rel\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \tag{2}$$

$$\text{U\_\{val\}} = \frac{\sum_{i=1}^{n} w_i\, x_i}{\sum_{i=1}^{n} w_i} \tag{3}$$

9. Development Process and Product Management

Both models are maintained under a shared configuration management plan with model versioning, solver version lock, and controlled libraries for material subroutines. Each model has a documented verification checklist that includes geometry import validation, unit checks, boundary condition verification, and regression tests comparing force–deformation curves against a gold-standard case. Model HF additionally includes automated mesh quality checks and a convergence audit that logs residual norms, line search iterations, and time increment reductions. All scripts used for preprocessing, solving, and post-processing are stored in a centralized repository with continuous integration that triggers on parameter file changes and runs the regression tests.

Change control is enforced with issue tracking and pull requests reviewed by a simulation lead and, for Model HF, a materials specialist. Release notes summarize parameter updates, mesh changes, and solver option adjustments, and map them to the verification checklist items they affect. Cross-functional meetings between the test and modeling teams review discrepancies discovered during output comparisons and assign actions. For design programs, the RO model is used early to prune concepts and tune gross parameters, while HF is used later for decision-grade analysis and to substantiate claims. This clear separation of use supports appropriate expectations for each model's role in the product development lifecycle.

Development process and product management — gradation.

a.
Ad hoc modeling; no version control or verification artifacts.

b.
Basic version tracking; limited verification and inconsistent documentation.

c.
Formal versioning, verification checklists, and change control for principal models.

d.
End-to-end configuration management, automated verification, traceable change control, and documented roles across the modeling lifecycle.

10. Development Technical Review

An independent design review covering the modeling approach, constitutive choices, and test–model mapping was convened mid-project, and observers from outside the immediate team attended; detailed deliberations and outcomes are beyond the scope of this paper.

Development technical review — gradation.

a.
No review or only informal peer discussion.

b.
Internal review by project team without formal minutes.

c.
Cross-functional review with documented comments and dispositions.

d.
Independent technical review with external subject-matter experts and formal closure of action items.

11. Use History

Both models have been applied in prior concept studies and verification exercises for adjacent foot geometries; a systematic analysis of that history is not presented here.

Use history — gradation.

a.
Model not used previously or use undocumented.

b.
Limited prior use with anecdotal outcomes.

c.
Documented prior use with lessons learned captured.

d.
Extensive prior use with traceable performance records and closed feedback loops informing model updates.

12. Scoring Method (0–9 Private Scale) and Evidence Mapping

We adopted a private 0–9 numeric scale to rate evidence for each factor per model–mechanism pair. Level 0 indicates the absence of substantiating evidence; Level 9 indicates convergent, multi-source evidence with traceable documentation and quantitative agreement within pre-specified limits. Scores reflect the synthesis of the evidence presented in this paper and do not claim portability outside the defined scope. We report one table per model with one row per factor–mechanism pairing and a basis statement that restates a finding supported by numbers in the prose.

Evidence mapping proceeded as follows. For each factor, we identified the relevant artifacts: for origin and traceability of data, we reviewed calibration certificates, chain-of-custody logs, and repository audit trails; for numerical resolution, we compiled mesh/time refinement results; for parameter mapping, we assembled crosswalks between bench/in-shoe setups and model inputs; for model parameter origin, we referenced test-derived constitutive fits and friction measurements; for agreement with measurements, we extracted RMSE, difference in energy return percentages, and stiffness deviations; for selection of QoIs, we verified linkage to prespecified acceptance bands and biomechanical endpoints; for physical articles, we reviewed lot coverage, conditioning, and environmental controls; and for process and configuration control, we reviewed verification artifacts, change control records, and regression test logs.

The following rubrics define the interpretation of the 0–9 scale in operational terms for each factor. While the textual anchors are not exhaustive, they capture the primary dimensions used by the assessors to weight the evidence for each model–mechanism pair.

Output comparison — gradation.

a.
0–2: Qualitative comparison only or mismatched conditions; no metrics.

b.
3–5: Limited quantitative comparison; some QoIs or conditions missing; ad hoc thresholds.

c.
6–8: Quantitative comparison across all QoIs and mechanisms with pre-specified acceptance bands; uncertainty acknowledged.

d.
9: Broad quantitative comparison with multi-lab replication or cross-platform confirmation; uncertainties propagated to decisions.

13. Scorecards: Model HF × Mechanism M1

The following table summarizes the scores for Model HF across both mechanisms. Rows indicate factor–model–mechanism scope, numeric level on the private 0–9 scale, and a basis statement aligned with numeric evidence presented earlier. The table covers eight assessed factors for M1 and M2, totaling sixteen rows.

14. Scorecards: Model HF × Mechanism M2

Narrative complement to the HF scorecard: For M2, the score slightly decreased relative to M1 in areas sensitive to rate effects, notably numerical resolution and parameter origin, because the elastomeric viscoelastic parameters carry higher uncertainty at short time scales and the dynamics of contact transitions are more prominent. Nonetheless, the differences between predicted and measured energy return and stiffness remained within the pre-set acceptance bands, supporting the use of the HF model for design decisions in sprint initiation within the tested scope.

15. Scorecards: Model RO × Mechanism M1

The next table summarizes the scores for Model RO across both mechanisms. As with HF, rows denote factor–model–mechanism scope with numeric grades and one-line bases restating evidence with numeric anchors.

16. Scorecards: Model RO × Mechanism M2

Narrative complement to the RO scorecard: The RO model's simplifications lead to underprediction of energy return and over-smoothing of the unloading slope under sprint initiation. The calibration to HF and test data at moderate rates carries imperfect portability to high-rate conditions. The scores reflect adequate screening performance for M1 and limited utility for M2 without additional calibration or enriched rate dependence.

No additional table is included here to maintain the one-table-per-model presentation.

17. Discussion of Findings by Factor and Pair

Origin and traceability of the empirical records supporting this study are complete and current, with sources, calibrations, and custody preserved, and we did not identify gaps that would confound the HF or RO comparisons in either mechanism.

With the adopted spatial and temporal resolution, further refinement altered the HF model's mid-stance stored energy and integrated work by less than about one percent, leaving the observed model–test differences dominated by assumptions and input variability rather than numerical artifacts across both mechanisms.

The numerical values used to drive the analyses mirror the bench and in-shoe conditions within the measurement tolerances for dimensions, friction, and kinematic histories, avoiding hidden mismatches that could bias the comparisons for walking and sprint-like loading.

Material laws, contact coefficients, and boundary histories were drawn from device-specific tests rather than tuned to match outputs, and their uncertainties were quantified to explain residual differences for both gait and sprint initiation conditions.

Across the QoIs and conditions studied, predictions met the predefined thresholds for agreement with measurements for the high-fidelity model and were within acceptable screening tolerance for the reduced-order model during walking, while underperforming for sprint initiation without further calibration.

The selected metrics—stored and released work, returned fraction, mid-range stiffness, and hysteresis—align directly with design decisions and biomechanical outcomes for both walking and sprint initiation, and we observed no evidence that alternative metrics would better capture the mechanisms in scope.

All physical components exercised in the experiments represent production-intent articles across the stated sizes with documented conditioning and environment control, and the in-shoe cohort coverage was sufficient to anchor the bench-to-use translation for walking and indicative—but not exhaustive—for sprint initiation.

Configuration control, verification records, and issue tracking demonstrate a managed maturation of the computational assets, with the HF model exhibiting the more extensive automated checks and the RO model maintaining appropriate governance for its role in early-stage design space exploration.

18. Limitations Within Assessed Factors

Several limitations qualify the presented evidence. First, while the bench and in-shoe datasets were rigorously collected and traced, the in-shoe sprint initiation subset included only five participants, constraining generalization of M2 findings and limiting the strength of in vivo corroboration at high rates. Second, high-rate viscoelastic characterization of the interface elastomer relied on stress relaxation and finite-rate tests up to 20 s1; extrapolation beyond that range was not attempted, but even within range, parameter uncertainties are larger than for quasi-static conditions. Third, the contact friction model, while anchored to reciprocating measurements with a Coulomb coefficient and a simple Stribeck regularization, may not capture micro-scale stick–slip transitions and surface conditioning effects occurring over repeated sprint initiations; this is particularly relevant to M2 and likely contributes to RO underprediction of returned work.

On the numerical side, the HF model's time-step refinement showed small but non-negligible sensitivity in peak forces and unloading slopes under M2, suggesting that while integrated work is robust, detailed matching of the fast return segment could be improved with adaptive stabilization or higher-order time integration. For the RO model, the Winkler foundation and distributed springs adequately describe average behavior during walking but lack the capacity to model lateral redistributions under non-uniform contact, especially at high rates, motivating either an enriched contact formulation or regime-specific calibration for M2.

Process-wise, while configuration control and verification are strong, automated coverage of extreme loading cases is less comprehensive for RO. Moreover, while an independent technical review occurred, the outcomes are not detailed here; additional transparency would strengthen external confidence. Finally, although both models have been used previously, we did not perform a systematic meta-analysis of past performance across programs, and thus we cannot claim historical predictive statistics as part of this assessment.

19. Conclusions and Planned Credibility Elevations Focused on Assessed Factors

This assessment applied a NASA-STD-7009A-aligned framework to two implicit FEA models of a composite-keel prosthetic foot for two mechanisms, walking energy return and sprint initiation rebound. Model HF achieved close agreement with bench and in-shoe estimates for M1 and remained within acceptance bands for M2, with numerical resolution tightly controlled. Model RO provided adequate screening accuracy for M1 but underpredicted M2 returned work due to simplified rate dependence and contact. The one-table-per-model summaries enumerate scores for eight factors and both mechanisms, totaling thirty-two rows.

Planned credibility elevations are prioritized by the gaps revealed. To sharpen high-rate predictions, we will expand elastomer characterization to strain rates up to 100 s1 using split Hopkinson pressure bar methods tailored for soft materials and temperature-controlled DMA at higher frequencies; this supports Model HF parameter updates and enables RO to incorporate a more accurate rate dependence. For contact, we will introduce an enriched friction model that includes velocity-weakening/strengthening transitions and experimentally quantify the evolution of mu with conditioning and surface contamination. Numerically, we will implement automated space–time convergence maps focused on M2 with error estimators that trigger locally refined meshes in the toe region and adaptive substepping during unloading.

On the experimental side, we will broaden physical sample coverage by adding two more production lots and extending in-shoe sprint initiation to a minimum of twelve participants to improve generalizability. For process and product management, we plan to extend the HF regression suite to include high-rate benchmarks and to add automated checks for contact energy balance closure. The RO toolchain will gain a regime selector that restricts use to parameters and loading rates within the calibrated domain and flags out-of-bounds conditions.

Within the decision context defined here, the assessed evidence supports continued use of the HF model for design substantiation in both mechanisms and judicious use of the RO model for walking-related screening. As the planned elevations are executed, we expect the RO model's M2 performance to improve and the HF model's high-rate numerical robustness to strengthen, closing the remaining credibility gaps.

References

[1] A. Ueda, K. Nakamura, C. Abbott, T. Quintero, Material model effects on contact pressure in computational models, Ann. Biomed. Eng. 93 (2024) 90–530.

[2] M. Halvorsen, D. Lindqvist, D. Grigoryan, Mesh density effects on chest compression in computational models, J. Biomech. Eng. 128 (2023) 369–900.

[3] S. Grigoryan, W. Eriksen, H. Vasquez, F. Abbott, Bone density effects on fatigue life in computational models, J. Mech. Behav. Biomed. Mater. 149 (2022) 236–788.

[4] R. Abbott, C. Ueda, E. Ueda, N. Vasquez, Boundary condition effects on hemolysis index in computational models, J. Mech. Behav. Biomed. Mater. 127 (2019) 254–744.

[5] H. Lindqvist, B. Yilmaz, Inlet profile effects on contact pressure in computational models, Comput. Fluids 82 (2012) 78–570.

[6] F. Ueda, D. Pettersson, F. Quintero, Cement mantle effects on peak force in computational models, Med. Eng. Phys. 50 (2020) 376–670.

[7] B. Silva, P. Quintero, N. Eriksen, Element type effects on stress range in computational models, Int. J. Numer. Methods 88 (2015) 102–688.

[8] L. Pettersson, T. Okonkwo, W. Ishikawa, Boundary condition effects on head displacement in computational models, J. Verif. Valid. Uncertain. 50 (2011) 176–697.

[9] C. Halvorsen, E. Abbott, S. Abbott, E. Okonkwo, A. Grigoryan, Solver tolerance effects on outlet velocity in computational models, Int. J. Numer. Methods 112 (2019) 396–509.

[10] G. Moreau, B. Rasmussen, W. Quintero, Gap size effects on flow split in computational models, Med. Eng. Phys. 65 (2021) 369–758.

[11] S. Castellano, D. Ueda, G. Abbott, Damping effects on stress range in computational models, J. Verif. Valid. Uncertain. 143 (2018) 98–456.

[12] L. Castellano, G. Abbott, A. Silva, N. Eriksen, T. Kaur, Cement mantle effects on cortical strain in computational models, Comput. Fluids 37 (2009) 362–550.

[13] L. Wojcik, B. Quintero, Gap size effects on flow split in computational models, Med. Eng. Phys. 118 (2021) 269–407.

[14] H. Fontaine, D. Nakamura, Bone density effects on shear stress in computational models, Ann. Biomed. Eng. 106 (2014) 32–857.

[15] M. Quintero, P. Grigoryan, C. Grigoryan, Cement mantle effects on rib deflection in computational models, Comput. Fluids 120 (2015) 233–610.

[16] G. Ishikawa, D. Kaur, M. Abbott, F. Silva, C. Moreau, Contact stiffness effects on outlet velocity in computational models, Int. J. Numer. Methods 23 (2016) 58–499.

[17] P. Rasmussen, D. Tanaka, T. Wojcik, P. Halvorsen, Load path effects on rib deflection in computational models, Med. Eng. Phys. 84 (2018) 356–568.

[18] E. Pettersson, K. Duarte, F. Nakamura, W. Yilmaz, L. Moreau, Material model effects on flow split in computational models, Ann. Biomed. Eng. 148 (2021) 291–428.

[19] A. Fontaine, T. Grigoryan, H. Fontaine, Inlet profile effects on rib deflection in computational models, Med. Eng. Phys. 101 (2007) 83–653.

[20] A. Fontaine, E. Abbott, K. Tanaka, T. Kaur, M. Duarte, Mesh density effects on chest compression in computational models, J. Biomech. Eng. 73 (2021) 12–596.

[21] N. Halvorsen, K. Duarte, Contact stiffness effects on contact pressure in computational models, Ann. Biomed. Eng. 22 (2015) 210–471.

[22] S. Ishikawa, G. Moreau, Flow rate effects on chest compression in computational models, Comput. Methods Biomech. 76 (2023) 18–426.

[23] S. Yilmaz, K. Rasmussen, B. Zetterberg, Time step effects on head displacement in computational models, Comput. Fluids 22 (2011) 297–433.

[24] A. Ishikawa, P. Jankowski, M. Ishikawa, J. Pettersson, Boundary condition effects on rib deflection in computational models, Med. Eng. Phys. 139 (2023) 78–528.

[25] T. Silva, J. Silva, N. Rasmussen, K. Halvorsen, Flow rate effects on head displacement in computational models, J. Verif. Valid. Uncertain. 137 (2010) 176–823.

[26] L. Bergman, L. Pettersson, E. Castellano, H. Rasmussen, M. Abbott, Gap size effects on head displacement in computational models, Comput. Fluids 147 (2018) 210–719.

[27] C. Fontaine, J. Silva, S. Bergman, S. Wojcik, Boundary condition effects on outlet velocity in computational models, Med. Eng. Phys. 31 (2015) 115–766.

[28] K. Zetterberg, E. Yilmaz, T. Lindqvist, Damping effects on shear stress in computational models, Comput. Methods Biomech. 32 (2011) 113–792.

[29] N. Tanaka, C. Silva, R. Halvorsen, Damping effects on chest compression in computational models, Med. Eng. Phys. 101 (2023) 123–687.

[30] H. Kaur, W. Vasquez, H. Jankowski, F. Castellano, Element type effects on cortical strain in computational models, Comput. Fluids 137 (2023) 126–458.

[31] G. Moreau, E. Abbott, G. Yilmaz, P. Ueda, Wall roughness effects on peak force in computational models, Comput. Methods Biomech. 119 (2019) 386–632.

[32] B. Wojcik, B. Kaur, S. Wojcik, L. Duarte, P. Eriksen, Solver tolerance effects on shear stress in computational models, Med. Eng. Phys. 82 (2012) 281–614.

[33] W. Abbott, C. Fontaine, M. Tanaka, M. Silva, W. Ishikawa, Yield stress effects on stress range in computational models, Ann. Biomed. Eng. 142 (2017) 59–782.

[34] W. Moreau, R. Rasmussen, Load path effects on cortical strain in computational models, Comput. Methods Biomech. 134 (2020) 237–786.

Table 1: Credibility assessment summary for Model HF across M1 and M2

Credibility factor	Level	Basis
Data pedigree - Model HF - M1	8	Six feet across two lots; ISO/IEC 17025 calibrations within 0.3%; chain-of-custody and reproducibility audits documented.
Discretization error - Model HF - M1	8	Fine mesh changed peak stored energy by 1.1% vs extrapolated value; halving time step altered integrated work by 0.5%.
Equivalency of input parameters - Model HF - M1	8	Lamina thickness 0.180 ± 0.006 mm matched model; friction mu = 0.70 within measured 0.68 ± 0.04; boundary kinematics aligned within 1.2°.
Model inputs - Model HF - M1	8	Composite and elastomer parameters from device-specific tests ($R2 \geq 0.987$ for visco fits); no post hoc tuning beyond measured ranges.
Output comparison - Model HF - M1	8	Energy return within 1.6 percentage points of bench (87% vs 88.6%); loop area error 0.32 J; RMSE 4.2% of span.
Relevance of the quantities of interest - Model HF - M1	8	QoIs tied to pre-set acceptance bands (±3 pp energy return, ±10% stiffness) and measured consistently across bench and in-shoe.
Test samples - Model HF - M1	7	Six articles from two lots, sizes M26–M28, four repeats; in-shoe data from nine users at 1.25 m/s.
Development process and product management - Model HF - M1	8	Version control with regression tests; automated convergence audits; documented change control and cross-team discrepancy reviews.
Data pedigree - Model HF - M2	8	Same lot-traceable articles; high-rate tests with accredited calibrations; data repository with hash-verifiable artifacts.
Discretization error - Model HF - M2	7	Halving time step from 1.0 to 0.5 ms changed work by 0.8%; to 0.25 ms by 1.1%; peak force shift ≤1.9%.
Equivalency of input parameters - Model HF - M2	8	High-rate elastomer parameters from stress relaxation and rate sweeps; bench and in-shoe kinematics matched to sprint initiation profiles.
Model inputs - Model HF - M2	7	Interface visco g1 and tau1 from tests but with greater uncertainty at high rates; friction within measured range; no tuning.
Output comparison - Model HF - M2	7	Energy return within 2.8 percentage points of bench (83% vs 85.8%); loop area error 0.58 J; RMSE 6.1% of span.
Relevance of the quantities of interest - Model HF - M2	8	QoIs connected to pre-set M2 bands (±5 pp energy return, ±15% stiffness) and reflect high-rate push-off.

Table 2: Credibility assessment summary for Model RO across M1 and M2

Credibility factor	Level	Basis
Data pedigree - Model RO - M1	8	Same bench and in-shoe datasets as HF; accredited calibrations; preserved chain-of-custody and reproducibility checks.
Discretization error - Model RO - M1	7	Refinement from 8.0 to 3.5 mm shells changed energy return by 1.6 pp and mid-range stiffness by 7.5%.
Equivalency of input parameters - Model RO - M1	7	Parameter crosswalk aligns shell/beam and Winkler terms to measured properties and HF calibration within stated tolerances.
Model inputs - Model RO - M1	6	Constitutive parameters derived from HF calibration and tests; some simplifications (Winkler contact) introduce approximations.
Output comparison - Model RO - M1	6	Energy return 4.9 pp low vs bench (87% vs 85.8%); RMSE 7.8% of span; stiffness within 7.5%.
Relevance of the quantities of interest - Model RO - M1	8	Same QoIs and acceptance bands as HF for M1; computed consistently for design screening.
Test samples - Model RO - M1	6	Covers six bench articles and nine in-shoe users via shared datasets; no additional RO-specific testing.
Development process and product management - Model RO - M1	7	Versioned scripts and verification checks; simpler regression suite than HF but with controlled releases.
Data pedigree - Model RO - M2	8	Relies on same high-rate datasets with accredited calibrations; stored in controlled repositories.
Discretization error - Model RO - M2	6	Refinement changed energy return by 2.2 pp and stiffness by 12%; temporal refinement altered return by ~1.0 pp.
Equivalency of input parameters - Model RO - M2	7	High-rate spring and damping terms calibrated to HF and test data; mappings documented within tolerances.
Model inputs - Model RO - M2	6	Rate dependence represented with simplified damping in springs; no post hoc tuning beyond calibration band.
Output comparison - Model RO - M2	5	Energy return 7.5 pp low vs bench (83% vs 75.5%); RMSE 11.2% of span; stiffness off by 12%.
Relevance of the quantities of interest - Model RO - M2	8	QoIs and acceptance bands defined for M2; used to bound RO's applicability to high-rate cases.
Test samples - Model RO - M2	5	Bench data across six articles; in-shoe sprint initiation from five users; no RO-specific augmentation.
Development process and product management - Model RO -	7	Configuration-managed with documented changes; fewer automated convergence audits