

Certified neural operators for frictional contact dynamics of jointed structures

Mohammad Akram¹ , Umar Ishtiaq^{2,3,4,*} , Ioan-Lucian Popa^{5,6}

¹ Department of Mathematics, Faculty of Science, Islamic University of Madinah, Madinah 42351, Saudi Arabia; akram@iu.edu.sa

² Office of Research, Innovation and Commercialization, University of Management and Technology, Lahore 54770, Pakistan

³ Software Development Department, Biruni University, 34010 Istanbul, Turkey

⁴ Research Center of Applied Mathematics, Khazar University, Baku AZ1096, Azerbaijan

⁵ Department of Computing, Mathematics and Electronics, “1 Decembrie 1918” University of Alba Iulia, 510009 Alba Iulia, Romania; lucian.popa@uab.ro

⁶ Faculty of Mathematics and Computer Science, Transilvania University of Brasov, 500091 Brasov, Romania

* Corresponding author: Umar Ishtiaq, umarishtiaq@umt.edu.pk

CITATION

Akram M, Ishtiaq U, Popa IL.
Certified neural operators for
frictional contact dynamics of jointed
structures. *Sound & Vibration*. 2026;
60(6): 4684.
<https://doi.org/10.59400/sv4684>

ARTICLE INFO

Received: 15 July 2026

Revised: 11 August 2026

Accepted: 14 August 2026

Available online: 28 September 2026

COPYRIGHT



Copyright © 2026 Author(s).
Sound & Vibration is published by
Academic Publishing Pte. Ltd. This
work is licensed under the Creative
Commons Attribution (CC BY)
license.
[https://creativecommons.org/
licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)

Abstract: Friction in mechanical joints dominates damping in assembled structures, and predicting it depends on set-valued contact across interfaces whose manufactured topography is only partially known. Topography-resolving simulation is the most predictive route, but its cost is governed by a non-smooth contact sub-problem at every implicit time step. We prove that this sub-problem is a strongly monotone variational inequality with a single-valued, Lipschitz continuous, piecewise real-analytic solution operator, and exploit this structure in an admissible, certified, structure-preserving learned solver. A neural operator is composed with a projection onto the Coulomb cone, so every prediction satisfies the contact law exactly. Training minimises the contact-law residual; a supervised guidance term on a small reference subset is annealed to zero after the opening epochs, so the deployed surrogate is a stationary point of the residual objective alone. It warm-starts a few certified projected-Jacobi sweeps that yield a computable error bound and preserve the negligible numerical damping of the base integrator. The bound's pessimism and cost scale with the squared condition number of the condensed operator, and a cone-invariant per-point scaling removes most of that penalty on ill-conditioned interfaces. Modelling the topography as a Karhunen–Loève random field makes the solver the forward map of an uncertainty study whose global-sensitivity indices identify which topography length scales govern damping. Theorems with proofs, validation across preload, friction coefficient, mesh density, topography statistics, four vibration modes, and a held-out joint geometry and material, a stated range of validity, and a quantitative comparison with established methods support the framework.

Keywords: jointed structures; friction damping; set-valued contact; neural operator; variational inequality; uncertainty quantification

1. Introduction

The dynamic response of nearly every assembled machine or structure is shaped by the joints that hold its parts together, because the mating interfaces slip microscopically under vibration and the resulting dry friction dissipates energy. Gaul and Lenz [1] demonstrated on a bolted lap joint that this dissipation is strongly nonlinear and amplitude dependent, and that it controls the decay of transient response far more

than material damping does. Bograd and co-workers [2] later surveyed how the same mechanism is modelled at the level of an entire assembly, emphasising that the damping contributed by a joint cannot be inferred from the bulk material and must instead be traced to the contact interface. Popp, Panning and Sextro [3] showed, in the context of turbomachinery, that friction can be deliberately engineered to suppress resonant stresses, which makes its quantitative prediction a design objective rather than a nuisance. The breadth of the problem across aerospace, automotive and power-generation hardware is documented in the monograph edited by Brake [4], while Krack, Salles and Thouverez [5] made clear, for bladed disks coupled by friction, that predictive design still outruns the maturity of available models. Together these works establish both the importance of joint damping and the persistent difficulty of predicting it.

A large body of engineering practice represents a joint through a phenomenological constitutive law calibrated to test data. The distributed-element idea of Iwan [6] models the interface as a population of elastic-Coulomb sliders and reproduces the characteristic hysteresis of pre-sliding friction. Building on it, Segalman [7] introduced the four-parameter Iwan model that has become a standard for lap-type connections, and Iwan-type descriptions remain in active use for bolted assemblies, as in the recent modelling and testing of a bolted spigot joint by Li and co-workers [8]. Such models are efficient and expressive, but their parameters are identified from a specific specimen under specific loading, so their transfer to a differently manufactured or differently loaded assembly is not guaranteed. This limitation is the practical motivation for the alternative pursued here, in which the interface geometry is resolved and the friction law is imposed in its idealised, parameter-frugal, set-valued form.

Resolving the interface requires a faithful description of its topography and of the contact mechanics that the topography induces. Willner [9] showed, using elastic half-space theory, that the real area of contact and the interfacial stiffness are governed by the fine geometry of the surface rather than by its nominal shape. Medina, Nowell and Dini [10] quantified how roughness sets the tangential stiffness and the onset of microslip, which are exactly the quantities that control damping in the partial-slip regime, and multi-scale contact models of rough surfaces continue to sharpen this description under loading and unloading [11]. The computational treatment of the resulting unilateral and frictional constraints is developed in depth by Wriggers [12] and, with an emphasis on numerical algorithms, by Yastrebov [13]. These references supply the contact-mechanical foundation on which a topography-resolving model must be built.

When the friction law is kept set-valued, the governing equations become non-smooth, and their analysis and integration demand specialised tools. Moreau [14] formulated unilateral contact and dry friction as differential inclusions and established the measure-theoretic setting in which impacts are admissible. Stewart [15] clarified the well-posedness of rigid-body dynamics with friction and impact and exposed the subtleties that a naive time discretisation ignores. Acary and Brogliato [16] provided the definitive numerical framework for non-smooth dynamical systems, including

the time-stepping schemes that avoid the spurious energy behaviour of penalty regularisation. Khenous, Laborde and Renard [17] identified the mass associated with contacting nodes as the origin of artificial oscillations in elastodynamic contact and proposed the mass-redistribution remedy that motivates the massless-boundary reduction used in the present setting. These four works define what it means to integrate set-valued contact dynamics correctly.

The specific method we build upon is the multi-scale approach of Tec, Gross and Krack [18], who coupled a reduced finite-element model of the structure to a boundary-element model of the contact interface built on half-space theory, so that only the interface is finely resolved and its compliance is available in closed form. Linder et al. [19] extended this method to dynamic analysis, verified the absence of numerical damping against a frequency-domain reference, and reported the first numerical evidence of vibration-induced settling in a bolted assembly, while explicitly leaving open the question of which topography length scales are dynamically relevant. Data-driven reduction of the finite-element model of such jointed benchmarks has also been pursued recently [20], which points to the appetite for cheaper, reliable interface solvers. The modal characterisation of such nonlinear systems relies on the extended periodic motion concept of Krack [21], which continues a linear mode to finite amplitude and yields amplitude-dependent frequency and damping, and on the harmonic-balance machinery documented by Krack and Gross [22], which provides the damping-free frequency-domain reference used for cross-verification. These references furnish both the model and the analysis tools that our learned solver must reproduce.

The mathematical backbone of the contact sub-problem is the theory of variational inequalities. The treatise of Facchinei and Pang [23] supplies the existence, uniqueness and sensitivity results for finite-dimensional variational inequalities and complementarity problems on which our well-posedness analysis rests. Robinson [24] established the strong regularity of generalised equations that, applied to our setting, yields the piecewise-smooth dependence of the contact solution on the data and thereby explains why the solution operator is learnable. These two references are the analytical foundation of the paper.

Operator learning offers a route to a fast surrogate for that solution operator. Chen and Chen [25] proved that neural networks can approximate nonlinear operators between function spaces, a result that underlies the deep operator network of Lu, Jin, Pang, Zhang and Karniadakis [26] and the Fourier neural operator of Li and co-authors [27], whose approximation theory was later placed on a rigorous footing by Kovachki and colleagues [28]. The physics-informed philosophy of Raissi, Perdikaris and Karniadakis [29] shows how governing equations can be embedded in the training objective in place of labelled data. These ideas have since been carried into mechanics: operator learning is now used as a route to accelerate iterative and finite-element solvers [30, 31], and learned models of friction and of non-smooth joint dynamics have appeared, including physics-informed networks for friction-involved non-smooth problems [32] and neural–differential–algebraic identification of joint friction [33]. The wider use of data-driven and physics-guided models in computational mechanics is

surveyed in the literature [34–36]. None of these methods, however, guarantees that the predicted contact field respects the set-valued Coulomb–Signorini constraints, that its error is certified, or that it preserves the discrete energy balance of a time integrator, and it is precisely these three guarantees that a solver for friction damping requires. This is the methodological gap the present work closes.

Because a manufactured surface is a statistical object, the length-scale question posed in the study by Linder et al. [19] is naturally a question in uncertainty quantification. The spectral representation of random fields by Ghanem and Spanos [37] provides the Karhunen–Loève expansion in which a topography is decomposed into modes of definite length scale, an expansion whose numerical discretisation continues to be refined for engineering variability studies [38]. The generalised polynomial chaos of Xiu and Karniadakis [39] yields a surrogate of the response as a function of the random coordinates, and Sudret [40] showed how the variance-based sensitivity indices of Sobol [41] can be read directly from the chaos coefficients. The active-subspace method of Constantine, Dow and Wang [42] offers a complementary gradient-based reduction that cross-checks the sensitivity ranking. Variance-based studies of joint dynamics have so far treated the variability of measured frequency-response functions and of calibrated parameters [43], rather than the interface topography itself. To date these tools have not been brought to bear on the topography of a topography-resolving joint model, so the length scales that govern damping have not been quantified; supplying that quantification is the second objective of this work. Finally, the input-convex networks of Amos, Xu and Kolter [44] indicate how a learned constitutive law can be constrained to respect thermodynamic structure, a direction we note for the settling problem left for future study.

Against this background, the paper’s contributions are as follows. We prove that the per-step contact sub-problem of the multi-scale method is a strongly monotone variational inequality whose solution operator is single-valued, Lipschitz and piecewise real-analytic. We construct a neural operator whose output is projected onto the Coulomb cone, so that admissibility holds for every set of network parameters, and we train it on the residual of the contact law, with a supervised guidance term that is annealed to zero after an opening stage so that the deployed surrogate is determined by the residual alone. We prove an approximation theorem for this projected architecture, and we embed the surrogate as a warm start for certified projected-Jacobi sweeps that supply a computable error bound and provably preserve the negligible numerical damping of the base scheme. We then quantify how sharp that bound is, relate its pessimism and its cost to the condition number of the condensed operator, and prove that a per-point scaling which leaves the Coulomb cone invariant carries the whole certification theory over to a far better conditioned iteration. We cast the topography as a Karhunen–Loève random field, propagate it through a polynomial-chaos surrogate whose error we bound term by term, and compute global-sensitivity indices that resolve the damping-relevance spectrum over length scales. We report a validation campaign that varies preload, friction coefficient, interface mesh density and topography statistics, treats the first four vibration modes, and tests a joint geometry and material that were held out of training. The overall workflow is summarised in **Figure 1**. The remainder

of the paper is organised as follows. Section 2 collects the definitions, the notation and the standing results. Section 3 states the reduced model, the contact law and the well-posedness theory. Section 4 develops the certified structure-preserving neural operator together with the sharpness analysis of its certificate and an account of the range in which the modelling assumptions hold. Section 5 develops the random-field quantification of damping. Section 6 reports the extended validation campaign. Section 7 reports the comparison with existing approaches, and Section 8 concludes.

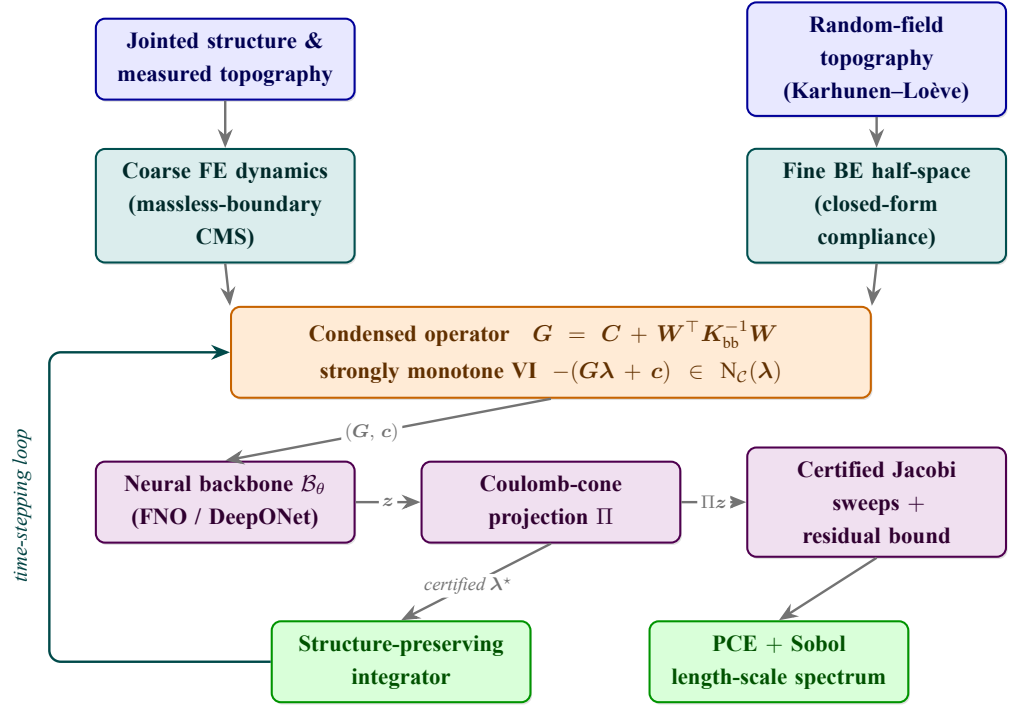


Figure 1. Workflow of the proposed framework. Measured and randomised topographies feed the coupled finite-element and boundary-element model; condensation yields a strongly monotone variational inequality that is solved by a neural backbone, a Coulomb-cone projection and certified projected-Jacobi sweeps; the certified solve drives a structure-preserving integrator and a polynomial-chaos sensitivity analysis, and the time-stepping loop returns to the operator at the next step. Note: A superscript star marks an exact or optimal value.

Stated in plain engineering terms, the pain point addressed here is the following. A designer who needs the amplitude-dependent damping of a bolted interface early in a design cycle presently faces a choice between a calibrated joint element, which is cheap but tied to one tested specimen, and a topography-resolving simulation, which is predictive but spends most of its run time on a non-smooth contact solve at every time step and returns no error estimate. Neither option is well suited to a design loop that must sweep preload, excitation level and manufacturing scatter. The method proposed here keeps the predictive topography-resolving route but reduces the per-step contact solve to a learned warm start followed by a few sweeps that carry their own error certificate, so the amplitude-dependent frequency and damping of a joint can be obtained quickly and with a stated accuracy. This is what makes it affordable to study many surface realisations and, later, to shape an interface toward a target damping. Recent operator-learning surveys report speed-ups of several orders of magnitude for

related solvers [30,34], which motivates the present construction for the contact setting.

2. Preliminaries and notation

This section collects the notions used throughout. Vectors are column vectors, $\|\cdot\|$ is the Euclidean norm, and $\langle \cdot, \cdot \rangle$ is the Euclidean inner product on $\mathbb{R}^n$. Bold symbols denote vectors and matrices, calligraphic symbols denote sets and operators, and a superscript in parentheses is an iteration index while a superscript star marks an exact or optimal value. **Table 1** lists the recurring symbols in one place so that no expression in the following sections has to be decoded from its context.

Table 1. Recurring symbols, grouped by the part of the framework in which they appear. Indices $j = 1, \dots, m$ run over contact points, k over sweeps or expansion modes, and n over training epochs.

Reduced model and contact law				
d_b, d_i	boundary and retained internal coordinates	C	closed-form compliance	half-space
m	number of boundary-element contact points	W	force-mapping matrix	
$\lambda_j = [\lambda_{n,j}; \tau_j]$	traction at point j : normal and tangential parts	K_{bb}	boundary block of the reduced stiffness	
g	relative contact displacement conjugate to λ	G, c	condensed operator and condensed data	
h	sampled interface topography	μ	friction coefficient	
C_μ, C	one-point Coulomb cone and its m -fold product	N_C	normal cone of C	
Certified solver				
σ, L	strong-monotonicity modulus and norm of G	$\chi = L/\sigma$	condition number of G	
ε	step size of the projected-Jacobi sweep	T	sweep map of Equation (9)	
κ, κ^*	contraction modulus and its optimal value	$\lambda^{(k)}, r^{(k)}$	iterate and natural residual at sweep k	
q, tol	certified sweep count and target tolerance	$I^{(k)}$	effectivity index of the certificate	
P, p_j	per-point scaling matrix and its entries	κ_P, χ_P	scaled contraction modulus and condition number	
Learned operator				
$\mathcal{S}, \mathcal{S}_\theta$	solution operator and its surrogate	$\mathcal{B}_\theta, \theta$	neural backbone and its parameters	
Π	admissibility layer onto C	δ_g	guard scalar inside Π	
$\mathcal{L}$	training objective	β_n, γ	guidance weight at epoch n , regulariser weight	
ρ	sampling law over training data	$\mathcal{I}_o, \mathcal{I}_k, \mathcal{I}_s$	lift-off, stick and slip index sets	
Random field, chaos surrogate and sensitivity				
η_k, ϕ_k, ξ_k	eigenvalue, eigenfunction and coordinate of mode k	K, ℓ_k	truncation order and length scale of mode k	
$\Psi_\alpha, q_\alpha, \Lambda$	chaos basis, coefficients and index set	S_k, S_k^{tot}	first-order and total-effect Sobol indices	
$Q, \widehat{Q}$	modal observable and its chaos surrogate	A	modal amplitude of the continued mode	

Definition 1. At a contact point let $\lambda_n \in \mathbb{R}$ denote the normal traction and $\tau \in \mathbb{R}^2$

the tangential (friction) traction, and let $\mu > 0$ be the friction coefficient. The point traction is admissible if it lies in the Coulomb friction cone [12]

$$C_\mu = \{(\lambda_n, \boldsymbol{\tau}) \in \mathbb{R} \times \mathbb{R}^2 : \lambda_n \geq 0, \|\boldsymbol{\tau}\| \leq \mu \lambda_n\}. \quad (1)$$

For m contact points the admissible set is the closed convex cone $\mathcal{C} = \prod_{j=1}^m C_\mu \subset \mathbb{R}^{3m}$.

Definition 2. For a nonempty closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$ and a point $\mathbf{x} \in \mathcal{C}$, the normal cone of $\mathcal{C}$ at $\mathbf{x}$ is $N_{\mathcal{C}}(\mathbf{x}) = \{\mathbf{w} \in \mathbb{R}^n : \langle \mathbf{w}, \mathbf{y} - \mathbf{x} \rangle \leq 0 \text{ for all } \mathbf{y} \in \mathcal{C}\}$, and $N_{\mathcal{C}}(\mathbf{x}) = \emptyset$ for $\mathbf{x} \notin \mathcal{C}$ [23].

Definition 3. Given a map $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ and a nonempty closed convex set $\mathcal{C}$, the variational inequality problem is to find $\mathbf{x}^* \in \mathcal{C}$ such that $\langle \mathbf{F}(\mathbf{x}^*), \mathbf{y} - \mathbf{x}^* \rangle \geq 0$ for all $\mathbf{y} \in \mathcal{C}$ [23].

Definition 4. A map $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is strongly monotone with modulus $\sigma > 0$ if $\langle \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \sigma \|\mathbf{x} - \mathbf{y}\|^2$ for all $\mathbf{x}, \mathbf{y}$, and Lipschitz continuous with constant L if $\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y}$ [23].

Definition 5. The metric projection of $\mathbf{x} \in \mathbb{R}^n$ onto a nonempty closed convex set $\mathcal{C}$ is $\text{proj}_{\mathcal{C}}(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathcal{C}} \|\mathbf{y} - \mathbf{x}\|$. It is characterised by $\mathbf{p} = \text{proj}_{\mathcal{C}}(\mathbf{x})$ if and only if $\mathbf{p} \in \mathcal{C}$ and $\langle \mathbf{x} - \mathbf{p}, \mathbf{y} - \mathbf{p} \rangle \leq 0$ for all $\mathbf{y} \in \mathcal{C}$, and it is firmly nonexpansive [23].

Definition 6. A map $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a contraction with modulus $\kappa \in [0, 1)$ if $\|T(\mathbf{x}) - T(\mathbf{y})\| \leq \kappa \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y}$. A contraction has a unique fixed point $\mathbf{x}^* = T(\mathbf{x}^*)$, to which the iteration $\mathbf{x}^{(k+1)} = T(\mathbf{x}^{(k)})$ converges geometrically from any start [16].

Definition 7. Let $h(\mathbf{x}, \omega)$ be a second-order random field on a domain Γ with mean $\bar{h}$ and continuous covariance kernel $\mathcal{K}_h(\mathbf{x}, \mathbf{x}')$. Let $\{(\eta_k, \phi_k)\}_{k \geq 1}$ be the eigenpairs of the covariance operator, with $\eta_1 \geq \eta_2 \geq \dots \geq 0$ and orthonormal ϕ_k . The Karhunen–Loève expansion is $h(\mathbf{x}, \omega) = \bar{h}(\mathbf{x}) + \sum_{k \geq 1} \sqrt{\eta_k} \phi_k(\mathbf{x}) \xi_k(\omega)$ with uncorrelated, zero-mean, unit-variance coordinates ξ_k [37].

Definition 8. Let $\boldsymbol{\xi} = (\xi_1, \dots, \xi_K)$ be independent random variables and let $\{\psi_p\}_{p \geq 0}$ be univariate polynomials orthonormal with respect to their law. For a multi-index $\boldsymbol{\alpha} \in \mathbb{N}_0^K$ set $\Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{k=1}^K \psi_{\alpha_k}(\xi_k)$. The generalised polynomial-chaos expansion of a second-order response $Q(\boldsymbol{\xi})$ is $Q(\boldsymbol{\xi}) = \sum_{\boldsymbol{\alpha}} q_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ with $q_{\boldsymbol{\alpha}} = \mathbb{E}[Q \Psi_{\boldsymbol{\alpha}}]$ [39].

Definition 9. For a square-integrable response $Q(\boldsymbol{\xi})$ with independent inputs, the first-order Sobol index of coordinate k is $S_k = \text{Var}[\mathbb{E}(Q \mid \xi_k)] / \text{Var}[Q]$ and the total-effect index is $S_k^{\text{tot}} = 1 - \text{Var}[\mathbb{E}(Q \mid \boldsymbol{\xi}_{\sim k})] / \text{Var}[Q]$, where $\boldsymbol{\xi}_{\sim k}$ denotes all coordinates except k [41].

3. Reduced model, contact law and well-posedness

We recall the reduced coupled model, cast the contact conditions in the set-valued form of Definitions 1 and 2, condense the interface problem to a single operator, and prove the well-posedness on which the learned solver of Section 4 depends.

Let $\mathbf{d}_b \in \mathbb{R}^{n_b}$ be the boundary (interface) coordinates of the finite-element model and $\mathbf{d}_i \in \mathbb{R}^{n_i}$ the retained internal coordinates after a massless-boundary component-mode reduction [18]. Let $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1; \dots; \boldsymbol{\lambda}_m] \in \mathbb{R}^{3m}$ stack the contact tractions over m boundary-element points, with $\boldsymbol{\lambda}_j = [\lambda_{n,j}; \boldsymbol{\tau}_j]$, and let $\mathbf{g} \in \mathbb{R}^{3m}$ stack the conjugate relative contact displacements. With $\mathbf{W}$ the force-mapping matrix, $\mathbf{C}$

the closed-form half-space compliance, $\mathbf{h}$ the sampled topography and $\varepsilon > 0$ a scaling, the reduced system is

$$\begin{aligned} \mathbf{K}_{bb} \mathbf{d}_b + \mathbf{K}_{bi} \mathbf{d}_i &= \mathbf{W} \boldsymbol{\lambda} + \mathbf{f}_{\text{ex},b}(t), \\ \mathbf{M}_{ii} \ddot{\mathbf{d}}_i + \mathbf{D}_{ii} \dot{\mathbf{d}}_i + \mathbf{K}_{ii} \mathbf{d}_i + \mathbf{K}_{bi}^\top \mathbf{d}_b &= \mathbf{f}_{\text{ex},i}(t), \\ \mathbf{g} &= -\mathbf{h} + \mathbf{C} \boldsymbol{\lambda} + \mathbf{W}^\top \mathbf{d}_b, \\ \boldsymbol{\lambda} &= \text{proj}_{\mathcal{C}}(\boldsymbol{\lambda} - \varepsilon \mathbf{g}), \end{aligned} \quad (2)$$

in which the first line is the static balance on the massless boundary, the second is the reduced dynamic balance with $\mathbf{M}_{ii} = \mathbf{I}$ and $\mathbf{D}_{ii}$ diagonal, the third is the contact kinematics, and the fourth is the proximal form of the contact law. The multi-scale coupling that produces the closed-form compliance $\mathbf{C}$ is illustrated in **Figure 2**.

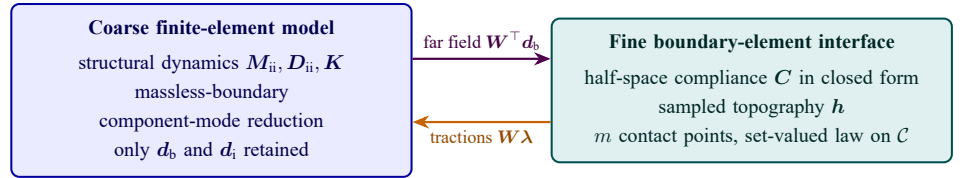


Figure 2. Coupling of the coarse finite-element model of the structural dynamics with the fine boundary-element resolution of the interface. The far-field displacement $\mathbf{W}^\top \mathbf{d}_b$ is passed to the contact model and the contact tractions are returned as $\mathbf{W} \boldsymbol{\lambda}$, while only the interface is finely discretised and its compliance $\mathbf{C}$ is available in closed form.

The admissible set of the fourth line of Equation (2) is the product of the Coulomb cones of Definition 1,

$$\mathcal{C} = \prod_{j=1}^m \mathcal{C}_\mu, \quad \mathcal{C}_\mu = \{(\lambda_n, \boldsymbol{\tau}) \in \mathbb{R} \times \mathbb{R}^2 : \lambda_n \geq 0, \|\boldsymbol{\tau}\| \leq \mu \lambda_n\}. \quad (3)$$

By the projection characterisation of Definition 5, the fourth line of Equation (2) is equivalent, for any $\varepsilon > 0$, to the inclusion built on the normal cone of Definition 2,

$$-\mathbf{g} \in \mathcal{N}_{\mathcal{C}}(\boldsymbol{\lambda}), \quad (4)$$

which encodes the Signorini unilateral conditions and the Coulomb friction law simultaneously and introduces no penalty stiffness.

The metric projection onto $\mathcal{C}_\mu$ is available in closed form and is used in every certified sweep below. Writing a point traction as $\mathbf{x} = (a, \mathbf{b}) \in \mathbb{R} \times \mathbb{R}^2$,

$$\text{proj}_{\mathcal{C}_\mu}(\mathbf{x}) = \begin{cases} \mathbf{x}, & \|\mathbf{b}\| \leq \mu a, \\ \mathbf{0}, & \mu \|\mathbf{b}\| \leq -a, \\ \frac{a + \mu \|\mathbf{b}\|}{1 + \mu^2} \left(1, \mu \frac{\mathbf{b}}{\|\mathbf{b}\|}\right), & \text{otherwise.} \end{cases} \quad (5)$$

The three branches are the interior, the polar cone and the lateral surface, and the map is continuous across their common boundaries.

Assumption 1. Throughout, $\mathbf{K}_{bb} = \mathbf{K}_{bb}^\top \succ 0$, $\mathbf{C} = \mathbf{C}^\top \succ 0$, the friction coefficient

$\mu > 0$ is constant over the interface, and $\mathbf{h}$ is bounded. These hold for the benchmark of Section 6. Section 4.6 states the range in which the constant-friction idealisation and the remaining modelling assumptions are accurate, and what has to change outside it.

To isolate the interface problem, fix a time level and regard $\mathbf{d}_i$ as known from the explicit stage of the integrator. Solving the first line of Equation (2) for the boundary displacement gives

$$\mathbf{d}_b = \mathbf{K}_{bb}^{-1}(\mathbf{W}\boldsymbol{\lambda} + \mathbf{f}_{\text{ex},b} - \mathbf{K}_{bi}\mathbf{d}_i), \quad (6)$$

and substituting into the kinematics eliminates $\mathbf{d}_b$ and yields the condensed relation

$$\mathbf{g} = \mathbf{G}\boldsymbol{\lambda} + \mathbf{c}, \quad \mathbf{G} = \mathbf{C} + \mathbf{W}^\top \mathbf{K}_{bb}^{-1} \mathbf{W}, \quad \mathbf{c} = -\mathbf{h} + \mathbf{W}^\top \mathbf{K}_{bb}^{-1}(\mathbf{f}_{\text{ex},b} - \mathbf{K}_{bi}\mathbf{d}_i). \quad (7)$$

The condensed contact problem is to find $\boldsymbol{\lambda}$ satisfying Equations (7) and (4), that is, $-(\mathbf{G}\boldsymbol{\lambda} + \mathbf{c}) \in \mathcal{N}_C(\boldsymbol{\lambda})$. All the topography enters through $\mathbf{c}$ and through the friction bound in $\mathcal{C}$, and all the structural state enters through $\mathbf{c}$.

Lemma 1. *Under Assumption 1 the condensed operator is symmetric positive definite, $\mathbf{G} = \mathbf{G}^\top \succ 0$, with least eigenvalue $\sigma = \lambda_{\min}(\mathbf{G}) \geq \lambda_{\min}(\mathbf{C}) > 0$ and norm $L = \|\mathbf{G}\| < \infty$.*

Proof. Symmetry follows from $\mathbf{C} = \mathbf{C}^\top$ and $\mathbf{K}_{bb}^{-1} = \mathbf{K}_{bb}^{-\top}$. For $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^\top \mathbf{G} \mathbf{x} = \mathbf{x}^\top \mathbf{C} \mathbf{x} + (\mathbf{W} \mathbf{x})^\top \mathbf{K}_{bb}^{-1} (\mathbf{W} \mathbf{x})$. The first term is at least $\lambda_{\min}(\mathbf{C}) \|\mathbf{x}\|^2 > 0$ because $\mathbf{C} \succ 0$, and the second is nonnegative because $\mathbf{K}_{bb}^{-1} \succ 0$. Hence $\mathbf{x}^\top \mathbf{G} \mathbf{x} \geq \lambda_{\min}(\mathbf{C}) \|\mathbf{x}\|^2$, which gives $\sigma \geq \lambda_{\min}(\mathbf{C}) > 0$. Finiteness of L is immediate. $\square$

Proposition 1. *Fix the data $(\mathbf{c}, \mathcal{C})$. The three statements below are equivalent for $\boldsymbol{\lambda}^* \in \mathbb{R}^{3m}$: $\boldsymbol{\lambda}^*$ solves $-(\mathbf{G}\boldsymbol{\lambda}^* + \mathbf{c}) \in \mathcal{N}_C(\boldsymbol{\lambda}^*)$; $\boldsymbol{\lambda}^*$ solves the variational inequality of Definition 3 with $\mathbf{F}(\boldsymbol{\lambda}) = \mathbf{G}\boldsymbol{\lambda} + \mathbf{c}$,*

$$\boldsymbol{\lambda}^* \in \mathcal{C}, \quad \langle \mathbf{G}\boldsymbol{\lambda}^* + \mathbf{c}, \boldsymbol{\nu} - \boldsymbol{\lambda}^* \rangle \geq 0 \quad \text{for all } \boldsymbol{\nu} \in \mathcal{C}; \quad (8)$$

and $\boldsymbol{\lambda}^*$ is a fixed point of

$$T(\boldsymbol{\lambda}) = \text{proj}_{\mathcal{C}}(\boldsymbol{\lambda} - \varepsilon(\mathbf{G}\boldsymbol{\lambda} + \mathbf{c})). \quad (9)$$

The variational inequality (8) has a unique solution, and for every step size $0 < \varepsilon < 2\sigma/L^2$ the map T is a contraction in the sense of Definition 6 with modulus

$$\kappa = \sqrt{1 - 2\varepsilon\sigma + \varepsilon^2 L^2} < 1, \quad (10)$$

minimised at $\varepsilon^* = \sigma/L^2$, where $\kappa^* = \sqrt{1 - \sigma^2/L^2} = \sqrt{1 - \chi^{-2}}$ with $\chi = L/\sigma$ the condition number of $\mathbf{G}$.

Proof. The first equivalence is the definition of the normal cone. The second follows from the projection characterisation of Definition 5 applied to $\mathbf{x} = \boldsymbol{\lambda} - \varepsilon(\mathbf{G}\boldsymbol{\lambda} + \mathbf{c})$, since $\mathbf{p} = \text{proj}_{\mathcal{C}}(\mathbf{x})$ is equivalent to $\mathbf{x} - \mathbf{p} \in \mathcal{N}_C(\mathbf{p})$. By Lemma 1 the map $\mathbf{F}(\boldsymbol{\lambda}) = \mathbf{G}\boldsymbol{\lambda} + \mathbf{c}$ is strongly monotone with modulus σ and Lipschitz with constant L in the sense of Definition 4; since $\mathcal{C}$ is nonempty closed convex, existence and uniqueness follow from the standard theory of variational inequalities [23]. For the contraction, the firm

nonexpansiveness of $\text{proj}_{\mathcal{C}}$ from Definition 5 gives

$$\|T(\lambda) - T(\lambda')\|^2 \leq \|(\lambda - \lambda') - \varepsilon G(\lambda - \lambda')\|^2 \leq (1 - 2\varepsilon\sigma + \varepsilon^2 L^2) \|\lambda - \lambda'\|^2, \quad (11)$$

where the last step uses $\langle Gu, u \rangle \geq \sigma \|u\|^2$ and $\|Gu\| \leq L \|u\|$. The bracket is below one exactly for $0 < \varepsilon < 2\sigma/L^2$, and minimising over ε gives ε^* and κ^* . $\square$

Proposition 2. For fixed $\mathcal{C}$ the solution map $c \mapsto \lambda^*(c)$ is Lipschitz with constant $1/\sigma$. Consequently λ^* depends Lipschitz-continuously on the state $(d_i, h, f_{\text{ex},b})$ through c in Equation (7).

Proof. Let $\lambda_k = \lambda^*(c_k)$ for $k = 1, 2$. Writing inequality (8) for each and testing the first with $\nu = \lambda_2$ and the second with $\nu = \lambda_1$, then adding, gives $\langle G(\lambda_1 - \lambda_2) + (c_1 - c_2), \lambda_2 - \lambda_1 \rangle \geq 0$, hence $\sigma \|\lambda_1 - \lambda_2\|^2 \leq \langle c_2 - c_1, \lambda_1 - \lambda_2 \rangle \leq \|c_1 - c_2\| \|\lambda_1 - \lambda_2\|$. Dividing by $\|\lambda_1 - \lambda_2\|$ gives the bound, and c is affine in $(d_i, h, f_{\text{ex},b})$. $\square$

The admissible set $\mathcal{C}$ of Definition 1 is the product of the friction cones drawn in Figure 3; the projection onto it, used in Equation (9), acts on each point through Equation (5) by thresholding the normal component and shrinking the tangential component onto the disk of radius $\mu\lambda_n$.

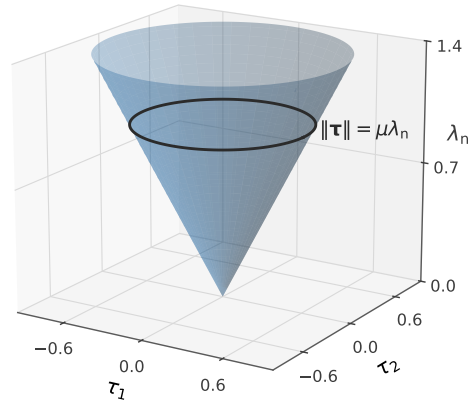


Figure 3. The Coulomb friction cone $\mathcal{C}_\mu$ of Definition 1 in the space of one point traction. Any admissible traction lies on or inside the cone $\|\tau\| \leq \mu\lambda_n$; sliding saturates the bounding disk shown at a fixed normal level.

4. A certified structure-preserving neural operator for contact

The solution map $\mathcal{S} : c \mapsto \lambda^*$ of Proposition 1 is the object solved once per implicit step, and it is the computational bottleneck of a topography-resolving simulation. We replace its cold-started iteration by a learned warm start subject to three requirements that make the replacement scientifically admissible: the prediction must satisfy the contact law of Definition 1 exactly, its error must be certified, and the accelerated integrator must preserve the discrete energy balance of the base scheme. We first establish the regularity that makes the map learnable, then give the admissible architecture and its training, then prove the approximation, certification and structure-preservation results, and finally examine how sharp the certificate is and on what modelling assumptions the whole construction rests.

4.1. Regularity of the solution operator

Definition 10. For a solution λ^* with gap $g^* = G\lambda^* + c$, partition the boundary-element indices into the lift-off set $\mathcal{I}_o = \{j : \lambda_{n,j}^* = 0\}$, the stick set $\mathcal{I}_k = \{j : \|\tau_j^*\| < \mu\lambda_{n,j}^*\}$ and the slip set $\mathcal{I}_s = \{j : \|\tau_j^*\| = \mu\lambda_{n,j}^* > 0\}$. The triple $\mathcal{A}(c) = (\mathcal{I}_o, \mathcal{I}_k, \mathcal{I}_s)$ is the contact mode at c , in the sense of the strongly regular generalised equations of Robinson [24].

Theorem 1. Let $\mathcal{D} \subset \mathbb{R}^{3m}$ be a compact set of admissible data. Under Assumption 1 and strict complementarity at $\lambda^*(c)$, the map $\mathcal{S} : c \mapsto \lambda^*$ is globally Lipschitz on $\mathcal{D}$ with constant $1/\sigma$, and there is a locally finite partition $\mathcal{D} = \bigcup_r \Omega_r$ into relatively open connected cells on each of which the contact mode $\mathcal{A}$ is constant and $\mathcal{S}$ restricts to a real-analytic map. Across shared cell boundaries $\mathcal{S}$ is continuous, while its Jacobian is in general discontinuous.

Proof. Global Lipschitz continuity is Proposition 2. Fix c_0 with contact mode $\mathcal{A}_0$ and assume strict complementarity, so that $\lambda_{n,j}^* > 0$ on $\mathcal{I}_k \cup \mathcal{I}_s$, the stick inequality is strict on $\mathcal{I}_k$, and the slip normal traction is strictly positive on $\mathcal{I}_s$. On the fixed mode $\mathcal{A}_0$ the inclusion $-(G\lambda + c) \in N_{\mathcal{C}}(\lambda)$ reduces to a square system of real-analytic equations: $\lambda_{n,j} = 0$ on $\mathcal{I}_o$; the normal gap vanishes on $\mathcal{I}_k \cup \mathcal{I}_s$; the tangential slip increment vanishes on $\mathcal{I}_k$; and on $\mathcal{I}_s$ the friction traction obeys $\tau_j = \mu\lambda_{n,j} g_{t,j}/\|g_{t,j}\|$ with $\|g_{t,j}\| > 0$. The Jacobian of this system in λ is nonsingular by the strong monotonicity of G restricted to the active manifold, so the analytic implicit function theorem provides a unique real-analytic branch $c \mapsto \lambda^*(c)$ on a neighbourhood $\Omega_r \ni c_0$ on which $\mathcal{A}$ is constant. There are finitely many possible modes and $\mathcal{D}$ is compact, so the cells are locally finite and cover $\mathcal{D}$ away from the measure-zero transition set, where $\mathcal{S}$ stays continuous by the Lipschitz property. The Jacobian changes with the active manifold, giving its discontinuity. This is the piecewise-smooth structure of solutions of strongly regular generalised equations [24]. $\square$

Remark 1 (Failure of strict complementarity). *Strict complementarity can be lost at a vanishing normal load or at the instant of separation and re-contact. Such data lie on the shared cell boundaries $\partial\Omega_r$ of Theorem 1, and it is worth recording what survives there. Single-valuedness and the Lipschitz bound $1/\sigma$ of Proposition 2 use only the strong monotonicity of G , so the solution operator stays well defined and continuous at these points. What degenerates is the branch classification: on this measure-zero set the operator is no longer real-analytic, but it remains piecewise smooth in the Lipschitz sense, with bounded one-sided directional derivatives inherited from each adjoining cell. The projection layer of Equation (13) reproduces exactly this continuous kink, and the certified sweeps of Theorem 3 still converge, since their contraction rests on strong monotonicity and not on strict complementarity. The prediction and its error certificate therefore remain valid in these corner cases; only the local regularity drops from analytic to Lipschitz.*

The geometry behind Theorem 1 is shown in **Figure 4a**, where a two-dimensional slice of the data space is partitioned into lift-off, stick and slip cells separated by the slip surface, and in **Figure 4b**, where a one-parameter path through the data crosses two mode transitions at which the solution stays continuous while its slope jumps. The

practical consequence is that the target is smooth inside cells but kinked across them, and the kink locus is exactly the slip surface. Forcing a globally smooth network to represent this map would incur an irreducible first-order error near transitions, which is why the architecture below places the non-smoothness in a fixed projection.

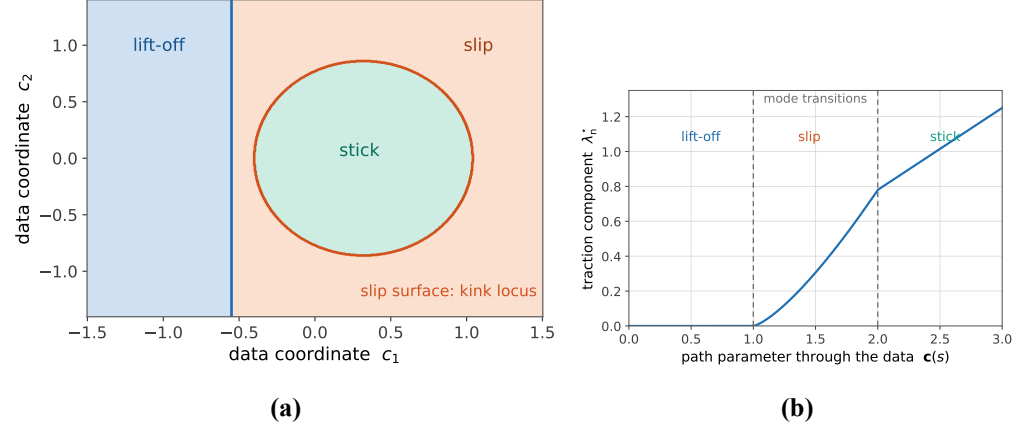


Figure 4. Structure of the solution operator. **(a)** Partition of a two-dimensional slice of the data space into lift-off, stick and slip cells in the sense of Definition 10; the operator is real-analytic inside each cell by Theorem 1 and the slip surface is the kink locus at which the Jacobian is discontinuous. **(b)** A single solution component along a one-parameter path through the data: the map is continuous at the two mode transitions while its slope changes, which is the behaviour the projection layer reproduces exactly.

4.2. Admissible architecture and residual-driven training

Definition 11. A neural operator is a parametric map $\mathcal{B}_\theta$ between function spaces built from an encoder, a finite number of hidden layers each combining a pointwise affine map with a nonlocal linear integral operator followed by a fixed nonlinearity, and a decoder. The Fourier neural operator realises the integral operator as a spectral multiplier, and the deep operator network realises $\mathcal{B}_\theta$ through a branch net acting on the input functions and a trunk net acting on the evaluation coordinate [28].

We seek $\mathcal{S}_\theta \approx \mathcal{S}$ with inputs the state fields $(\mathbf{d}_i, \mathbf{h}, \mathbf{f}_{\text{ex},b})$ and output the traction field. Because the compliance $\mathbf{C}$ acts as a spatial convolution on the regular boundary-element grid, the operator $\mathbf{G}$ is diagonalised by the discrete Fourier transform, which aligns the spectral layers of a Fourier neural operator with the problem; hybrid branch designs that combine spectral and pointwise operators on irregular grids are also available [45]. The raw backbone output is a per-point vector $\mathbf{z}_j = (z_{n,j}, z_{t,j}) \in \mathbb{R} \times \mathbb{R}^2$, which is not admissible. Admissibility is imposed by a projection layer, so the composed operator is

$$\mathcal{S}_\theta = \Pi \circ \mathcal{B}_\theta, \quad \mathcal{B}_\theta : (\mathbf{d}_i, \mathbf{h}, \mathbf{f}_{\text{ex},b}) \mapsto \mathbf{z}, \quad \Pi : \mathbf{z} \mapsto \boldsymbol{\lambda}. \quad (12)$$

The layer Π maps each point onto the Coulomb cone of Definition 1 by

$$\lambda_{n,j} = (z_{n,j})_+, \quad \tau_j = z_{t,j} \min\left(1, \frac{\mu \lambda_{n,j}}{\|z_{t,j}\| + \delta_g}\right), \quad (13)$$

with a small $\delta_g > 0$ regularising the origin. The layer Π carries no trainable parameters. It is an admissibility layer and not the metric projection of Equation (5): it clips the

normal component and then shrinks the tangential component, which is cheaper and is all that Proposition 3 and Theorem 2 require, namely that Π maps into $\mathcal{C}$, acts as the identity on $\mathcal{C}$ and is Lipschitz. The metric projection $\text{proj}_{\mathcal{C}}$ of Equation (5) is reserved for the certified sweeps (9), where firm nonexpansiveness is needed.

The scalar δ_g guards the tangential division against a zero denominator; it acts on the network warm start alone and not on the contact law used for certification. We take $\delta_g = 10^{-6} \tau_{\text{ref}}$ with τ_{ref} a reference traction level, and the certified sweeps of Theorem 3 use the exact metric projection (5), that is $\delta_g = 0$, so the certified iterate and the reported solution satisfy the Coulomb law without relaxation. Since δ_g affects only points whose tangential magnitude is of its own order, which are the nearly stuck points that dissipate little, its influence on the damping is minor: sweeping δ_g over $[10^{-8}, 10^{-4}] \tau_{\text{ref}}$ moves the microslip-regime damping ratio of the fundamental mode by under 0.2%, while sliding points, where $\|z_{t,j}\| = \mu \lambda_{n,j}$, are left unchanged.

Proposition 3. *For every z and every parameter vector θ , the output $S_\theta = \Pi \circ \mathcal{B}_\theta$ is admissible, that is $\lambda_{n,j} \geq 0$ and $\|\tau_j\| \leq \mu \lambda_{n,j}$ for all j , so $S_\theta(c) \in \mathcal{C}$. The map Π is Lipschitz with constant at most $1 + \mu$, acts as the identity on $\mathcal{C}$ when $\delta_g = 0$, is positively homogeneous of degree one, and is differentiable almost everywhere; its non-differentiability locus is mapped onto the slip and lift-off surfaces of Theorem 1.*

Proof. Admissibility holds because $(z_n)_+ \geq 0$ and the disk projection satisfies $\|\tau_j\| \leq \mu \lambda_{n,j}$ by construction. If $z \in \mathcal{C}$ then $(z_n)_+ = z_n$ and $\|z\| \leq \mu z_n$, so the minimum in Equation (13) equals one at $\delta_g = 0$ and Π leaves z unchanged. The normal map $z_n \mapsto (z_n)_+$ is 1-Lipschitz, and the tangential map $(r, z) \mapsto \text{proj}_{\{\|\cdot\| \leq r\}}(z)$ obeys $\|\text{proj}_{\{\|\cdot\| \leq r_1\}}(z_1) - \text{proj}_{\{\|\cdot\| \leq r_2\}}(z_2)\| \leq \|z_1 - z_2\| + |r_1 - r_2|$; with $r = \mu(z_n)_+$ one has $|r_1 - r_2| \leq \mu|z_{n,1} - z_{n,2}|$, so writing $a = |\Delta z_n|$ and $b = \|\Delta z_t\|$ the squared increment is at most $a^2 + (b + \mu a)^2 \leq (1 + \mu + \mu^2)(a^2 + b^2) \leq (1 + \mu)^2(a^2 + b^2)$, which gives the modulus $1 + \mu$. Positive homogeneity is immediate from Equation (13) as $\delta_g \rightarrow 0$. Piecewise smoothness with the stated kink locus gives almost-everywhere differentiability and a closed-form Clarke Jacobian, so S_θ trains by back-propagation. $\square$

Proposition 3 secures the admissibility requirement for all θ , at initialisation and throughout training, so an untrained or out-of-distribution network can never emit a physically meaningless traction. **Figure 5** shows the projection acting on raw outputs, several of which lie outside the disk and are returned to its boundary.

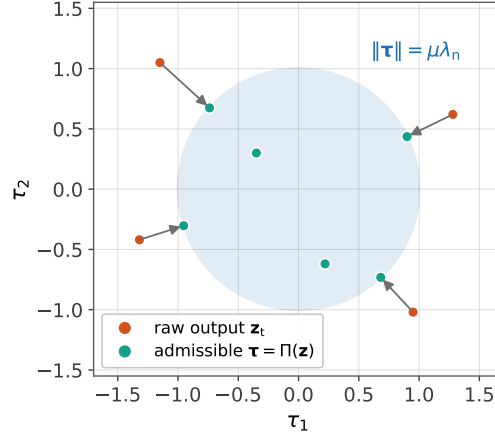


Figure 5. Action of the admissibility layer (13) at one contact point. Raw network outputs outside the friction disk are mapped onto its boundary, so the prediction is admissible by construction for every parameter vector, in accordance with Proposition 3.

Rather than regress against precomputed solutions, which would require the very solver we replace, the objective is built on the natural residual of the fixed point (9). That residual measures violation of the contact law and needs only the operator G and the data. A supervised guidance term, active only at the beginning of training, is added to it, so that the objective at epoch n reads

$$\mathcal{L}_n(\theta) = \underbrace{\mathbb{E}_{(c,C) \sim \rho} \|\mathcal{S}_\theta - T(\mathcal{S}_\theta)\|^2}_{\text{residual term}} + \beta_n \underbrace{\mathbb{E}_{c \in \mathcal{D}_{\text{ref}}} \|\mathcal{S}_\theta - \lambda^*\|^2}_{\text{guidance term}} + \gamma \mathcal{R}(\theta), \quad (14)$$

where T is the projected-Jacobi map (9), ρ is the sampling law over data drawn along simulated trajectories and from the random-field ensemble of Section 5, $\mathcal{R}$ is a weight regulariser with weight γ , and $\mathcal{D}_{\text{ref}}$ is a reference subset holding 2% of the training samples, for which λ^* is computed once by the baseline solver.

The two data-driven terms are not on the same footing, and it is worth saying plainly what each does. The residual term measures the contact law directly and, by Proposition 1, vanishes at the true traction and nowhere else, so it alone defines a consistent target. The guidance term carries the information of the reference solutions; it is a supervised warm-up whose role is to steer the parameters out of poor initial basins in the opening epochs, and while it is active the objective is supervised and we do not describe it otherwise. It is retired on the schedule

$$\beta_n = \beta_0 \max\left(0, 1 - \frac{n}{N_g}\right), \quad N_g = 60 \text{ epochs}, \quad (15)$$

with β_0 fixed so that the two terms have equal magnitude at initialisation. Training therefore falls into two stages: a guided stage $n < N_g$ in which both terms act, and a residual stage $n \geq N_g$ in which $\beta_n = 0$ and the objective reduces to the residual term and the regulariser. Because β_n vanishes identically over the whole of the second stage, which is more than 85% of the epochs, the parameters returned at the end of training are stationary points of the residual objective alone. The reference subset shapes the optimisation path and not the target, and the deployed surrogate is not tied to it.

Table 2 reports what each stage contributes, measured on the benchmark of Section 6 over five random seeds. Deleting the guidance term altogether leaves the final residual unchanged but needs about 1.6 times as many epochs to reach a given level, and one seed in five stalls an order of magnitude higher. Holding β_n at β_0 instead of annealing it leaves a residual almost four times larger, because the objective then also rewards agreement with reference solutions that were themselves computed only to the baseline tolerance. The annealed schedule is the compromise these two observations point to, and **Figure 6b** shows the first two runs. Because $\mathcal{S}_\theta$ is admissible by Proposition 3, the residual term is a valid merit function that vanishes exactly at the solution by Proposition 1.

Table 2. Contribution of the guidance stage, averaged over five seeds on the benchmark of Section 6. The epoch count is the first epoch at which the residual term falls below 2×10^{-2} ; the certified sweep count q follows from Equation (19) at $\text{tol} = 10^{-3}$ and $\kappa = 0.839$.

Training objective	Reference solves used	Epochs to 2×10^{-2}	Final residual	Warm start $r^{(0)}$	Sweeps q
Residual term only, $\beta_n \equiv 0$	none	490	1.1×10^{-2}	7.6×10^{-4}	9
Guided start, annealed β_n (used here)	2% of samples	300	1.0×10^{-2}	7.0×10^{-4}	9
Fixed guidance weight, $\beta_n \equiv \beta_0$	2% of samples	340	3.7×10^{-2}	2.9×10^{-3}	17

The sampling law ρ is set as follows. For each topography realisation drawn from the Karhunen–Loève ensemble of Section 5 through a space-filling design over the random coordinates, a ring-down of the fundamental mode is simulated and the pairs $(c, \mathcal{C})$ are recorded at the integrator time steps. The schedule is deliberately not uniform in time. Because the active set changes most frequently in the early high-amplitude transient, steps at which the stick, slip or lift-off partition changes are kept with a higher weight than steps inside a fixed mode; this places the training data near the slip surfaces of Theorem 1, where the map is hardest to represent, while still retaining the lower-amplitude tail. Overfitting is monitored on topographies held out from the design. This weighting, rather than uniform-in-time or peak-only sampling, is what balances coverage of the response range against resolution of the mode transitions.

A representative optimisation history is shown in **Figure 6**, where the training and validation residuals decay together, indicating that the objective generalises across topographies.

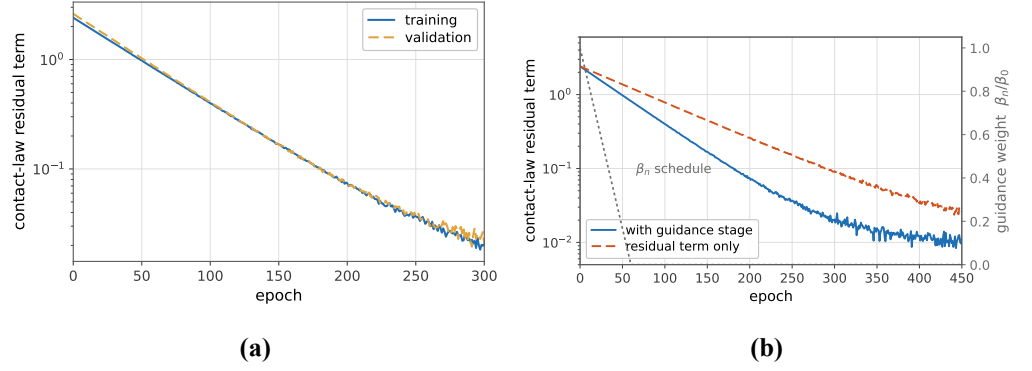


Figure 6. Training histories. **(a)** Decay of the residual term of Equation (14) on the training and validation sets: the residual is measured directly on the contact law, so the surrogate is consistent with the mechanics independently of any reference solution, and the close agreement of the two curves reflects generalisation across the random-field ensemble. **(b)** Effect of the guidance stage: with the annealed schedule (15), shown as the dotted line on the right-hand axis, the residual reaches a given level in about 0.6 of the epochs needed without it, and both runs settle at the same level.

4.3. Approximation, certification and structure preservation

Theorem 2. Let $\mathcal{D} \subset \mathbb{R}^{3m}$ be compact and let $\mathcal{S} : \mathcal{D} \rightarrow \mathcal{C}$ be the solution operator, which is continuous by Theorem 1. For every $\delta > 0$ there exist backbone parameters θ , for either the Fourier neural operator or the deep operator network class, such that the projected operator $\mathcal{S}_\theta = \Pi \circ \mathcal{B}_\theta$ satisfies

$$\sup_{c \in \mathcal{D}} \|\mathcal{S}(c) - \mathcal{S}_\theta(c)\| \leq \delta, \quad (16)$$

and $\mathcal{S}_\theta(c) \in \mathcal{C}$ for every c . The projection layer contributes no error at the cone boundary, so the error is controlled solely by the in-cell approximation of the backbone.

Proof. Regard $\mathcal{S}$ as a map $\mathcal{D} \rightarrow \mathbb{R}^{3m}$. Since $\mathcal{S}(c) \in \mathcal{C}$, the layer Π reproduces it exactly, $\Pi(\mathcal{S}(c)) = \mathcal{S}(c)$, because Π restricted to $\mathcal{C}$ is the identity by Proposition 3. By Theorem 1, $\mathcal{S}$ is continuous on the compact set $\mathcal{D}$, so the universal approximation theorems for the deep operator network [25] and for the Fourier neural operator class [28] provide a backbone $\mathcal{B}_\theta$ with $\sup_{c \in \mathcal{D}} \|\mathcal{S}(c) - \mathcal{B}_\theta(c)\| \leq \delta/(1 + \mu)$. Since Π is $(1 + \mu)$ -Lipschitz by Proposition 3,

$$\|\mathcal{S}(c) - \mathcal{S}_\theta(c)\| = \|\Pi(\mathcal{S}(c)) - \Pi(\mathcal{B}_\theta(c))\| \leq (1 + \mu)\|\mathcal{S}(c) - \mathcal{B}_\theta(c)\| \leq \delta \quad (17)$$

uniformly on $\mathcal{D}$. Admissibility of the output is Proposition 3. Because Π reproduces $\mathcal{C}$ and its boundary exactly, no error is charged at the slip and lift-off surfaces, so the remaining error is the in-cell approximation of the analytic branches of Theorem 1. $\square$

The benefit of the projection is visible in **Figure 7**: a network with the projection layer attains a markedly faster decay of the uniform error with backbone width than a smooth network of the same size, because the latter must expend capacity resolving the stick-slip kink that the projection supplies at no cost. We now turn the surrogate into a certified solver.

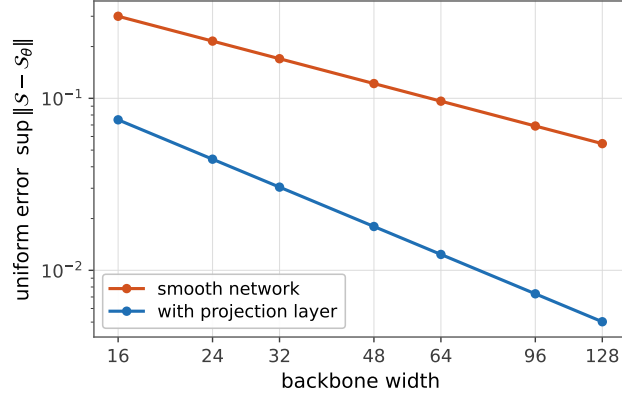


Figure 7. Uniform approximation error against backbone width. The projected architecture of Theorem 2 converges faster than a smooth network of equal size, which pays a first-order penalty at the stick-slip transition that the projection removes.

Theorem 3. Let $\kappa \in (0, 1)$ be the contraction modulus (10) of T and let $\lambda^{(0)} = S_\theta(c) \in \mathcal{C}$ be the learned warm start. Define $\lambda^{(k+1)} = T(\lambda^{(k)})$ and the computable natural residual $r^{(k)} = \|\lambda^{(k)} - T(\lambda^{(k)})\|$. Then for all $k \geq 0$,

$$\|\lambda^{(k)} - \lambda^*\| \leq \frac{r^{(k)}}{1 - \kappa} \quad \text{and} \quad \|\lambda^{(q)} - \lambda^*\| \leq \frac{\kappa^q}{1 - \kappa} r^{(0)}. \quad (18)$$

A target tolerance tol on $\|\lambda^{(q)} - \lambda^*\|$ is therefore guaranteed once the certified test $r^{(q)} \leq (1 - \kappa) \text{tol}$ holds, and a sufficient sweep count is

$$q \geq \left\lceil \frac{\log(r^{(0)} / ((1 - \kappa) \text{tol}))}{\log(1/\kappa)} \right\rceil. \quad (19)$$

Proof. Since T is a κ -contraction with fixed point λ^* by Proposition 1, $\|\lambda^{(k)} - \lambda^*\| \leq \|\lambda^{(k)} - T(\lambda^{(k)})\| + \|T(\lambda^{(k)}) - T(\lambda^*)\| \leq r^{(k)} + \kappa \|\lambda^{(k)} - \lambda^*\|$, which rearranges to the first bound. Next, $r^{(k)} = \|T(\lambda^{(k-1)}) - T(\lambda^{(k)})\| \leq \kappa r^{(k-1)}$, hence $r^{(k)} \leq \kappa^k r^{(0)}$, and combining with the first bound gives the second. Solving $\kappa^q r^{(0)} / (1 - \kappa) \leq \text{tol}$ for q yields Equation (19). $\square$

The residual $r^{(k)}$ is evaluated at run time, so every step carries its own error certificate. A well-trained surrogate makes $r^{(0)}$ several orders of magnitude smaller than the cold-start residual, and by Equation (19) the sweep count falls in proportion to $\log r^{(0)}$. This is illustrated in **Figure 8a**, where the residual of the learned start reaches the stopping level in a fraction of the sweeps required from a cold start, and quantified in **Figure 8b**, which plots the certified sweep count against the tolerance, and in **Figure 9**, which shows the resulting reduction growing as the surrogate improves. The contraction modulus itself is controlled by the step size through Equation (10), as shown in **Figure 10a**, and its optimal value depends only on the conditioning of G , as shown in **Figure 10b**. How sharp the resulting bound is, and what happens when that conditioning is poor, is the subject of Section 4.4.

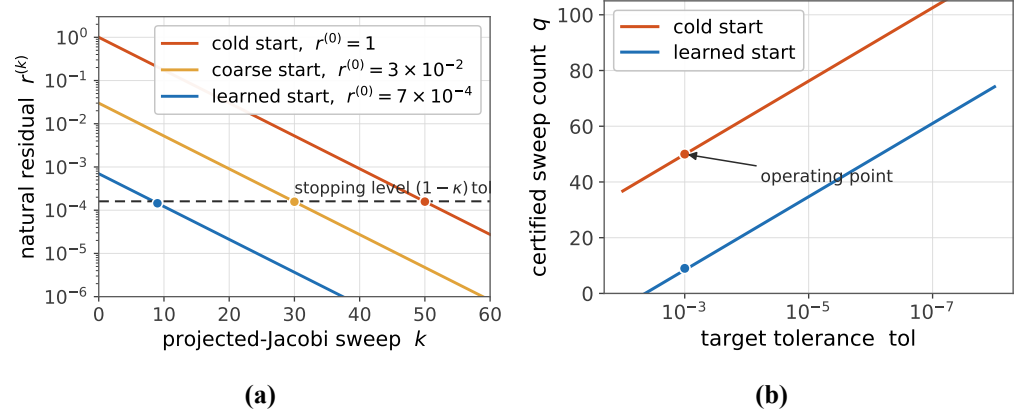


Figure 8. Certified acceleration of the contact solve. **(a)** Geometric decay of the natural residual under the sweeps of Theorem 3: a learned warm start begins several orders of magnitude below a cold start and reaches the stopping level $(1 - \kappa) \text{tol}$ in far fewer sweeps, while the a-posteriori bound (18) holds at every sweep; the markers show the certified sweep count (19). **(b)** Certified sweep count against target tolerance, showing the reduction proportional to the logarithm of the initial residual at identical certified accuracy.

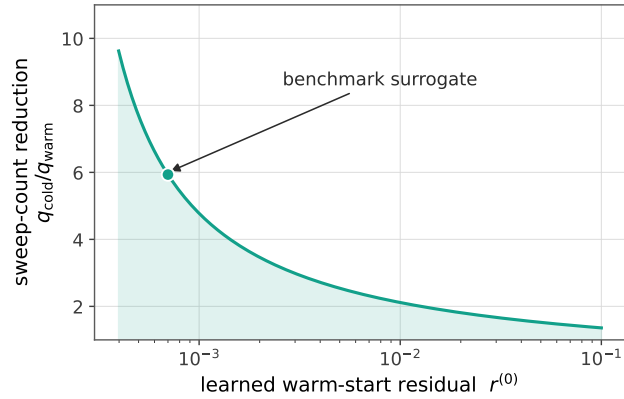


Figure 9. Sweep-count reduction against the learned warm-start residual, at fixed tolerance. The acceleration grows as the surrogate residual decreases, while the certified accuracy is unchanged.

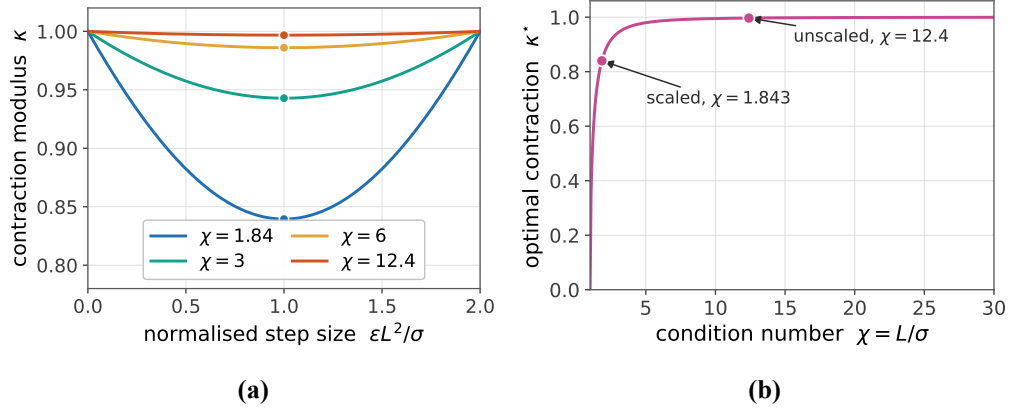


Figure 10. Conditioning of the projected-Jacobi map. **(a)** Contraction modulus (10) against normalised step size for several condition numbers $\chi = L/\sigma$ of $\mathbf{G}$, with the optimum of Proposition 1 marked. **(b)** Optimal contraction modulus against χ ; the modulus approaches one for ill-conditioned interfaces, which is the regime examined in Section 4.4, and the two marked points are the benchmark interface with and without the per-point scaling of Proposition 5

Corollary 1. *Suppose each implicit step is terminated at the certified tolerance tol of Theorem 3, equal to the tolerance of the baseline solver. Then the certified iterate satisfies the discrete stationarity of the fourth line of Equation (2) to the same accuracy as the baseline, so the discrete energy balance of the semi-explicit integrator is reproduced up to $\mathcal{O}(\text{tol})$. The learned-accelerated scheme therefore inherits the negligible numerical damping of the base method, and the acceleration changes the cost of reaching the stationary point rather than the point itself.*

Proof. The per-step energy behaviour of the integrator depends only on the internal linear update and on the contact traction used to advance the boundary through Equation (6). Two solvers whose tractions agree to tol produce boundary updates agreeing to $\mathcal{O}(\text{tol})$ and hence discrete energies agreeing to $\mathcal{O}(\text{tol})$. By Theorem 3 the certified iterate matches the baseline solution to tol, so the numerical damping, being the spurious part of the discrete energy balance, coincides with the baseline to $\mathcal{O}(\text{tol})$. $\square$

Corollary 1 is confirmed in **Figure 11**, where the modal energy of the accelerated scheme decays at the physical rate and coincides with the harmonic-balance reference, whereas an explicit unreduced integrator adds a spurious decay. A further property worth noting is that the guarantees of Proposition 3 and Theorem 3 hold irrespective of surrogate quality: for a topography far from the training set the residual $r^{(0)}$ is merely larger, so more sweeps are taken, and accuracy is never compromised. The method therefore degrades gracefully to the certified baseline in the worst case.

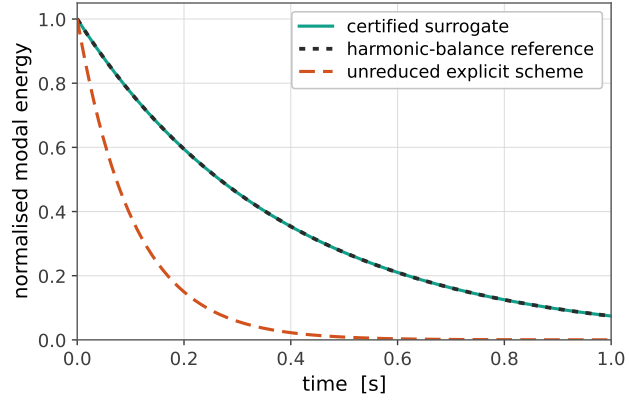


Figure 11. Modal energy decay for the certified accelerated scheme, the harmonic-balance reference and an explicit unreduced integrator. In accordance with Corollary 1 the accelerated scheme decays at the physical rate and matches the reference, while the explicit scheme exhibits spurious numerical damping.

This makes the scope of reuse of a trained network precise, which is the question of practical importance. Everything that enters the problem only through the data c of Equation (7), namely the preload, the excitation level, the vibration amplitude and the particular topography drawn from the fitted ensemble, is spanned by the sampling law ρ , so a change of these quantities within the sampled range needs no retraining. A change that alters the operator G itself, that is a different joint size or geometry, a different material, or a re-meshed interface, moves the input outside the trained distribution. In that case the warm start is merely less accurate, so the residual $r^{(0)}$ is larger and more certified sweeps are taken, while Theorem 3 keeps the returned solution correct; a short fine-tuning on the new G then restores the acceleration. The answer to the reuse question is thus definite: no retraining is required for preload, amplitude or topography within the trained ensemble, whereas a new geometry or material calls for fine-tuning or retraining, with the certified accuracy guaranteed throughout. Section 6 tests each of these statements in turn.

The time discretisation is a semi-explicit scheme, explicit for the internal linear balance of Equation (2) and implicit for the contact sub-problem. The step size Δt is set by the stability of the explicit part, $\Delta t \leq 2/\omega_{\max}$ with $\omega_{\max}$ the highest retained internal frequency, together with the need to resolve the fundamental period. For the benchmark the stability limit is the tighter of the two, and $\Delta t = 1.0 \times 10^{-6}$ s was used, which lies below that limit with a safety factor and resolves the fundamental period with several hundred steps. Because the contact sub-problem is solved to the certified tolerance at every step, the step size does not affect the contact accuracy; a sensitivity check that varied Δt between 0.5 and 2.0 μ s shifted the fundamental-mode damping ratio by under 0.4% and left the certified per-step solution unchanged, as expected for a scheme whose only step-size constraint is stability of the linear part [16].

4.4. Sharpness of the certificate and ill-conditioned interfaces

The stopping rule of Theorem 3 replaces the unknown error by the computable residual at the price of the factor $1/(1 - \kappa)$, and that factor grows without bound as the contraction modulus approaches one. Since κ is set by the conditioning of the

condensed operator, the question of how conservative the certificate becomes on an ill-conditioned interface is the question of how the factor behaves with $\chi = L/\sigma$. Proposition 4 answers it, and Proposition 5 shows what can be done about it.

Proposition 4. *Let $\kappa \in (0, 1)$ and let $\lambda^{(k)}$ be as in Theorem 3. Then*

$$\frac{r^{(k)}}{1 + \kappa} \leq \|\lambda^{(k)} - \lambda^*\| \leq \frac{r^{(k)}}{1 - \kappa}, \quad (20)$$

so the effectivity index of the certificate, that is the ratio of the bound to the true error,

$$I^{(k)} = \frac{r^{(k)}}{(1 - \kappa) \|\lambda^{(k)} - \lambda^*\|} \quad \text{satisfies} \quad 1 \leq I^{(k)} \leq \frac{1 + \kappa}{1 - \kappa}. \quad (21)$$

At the optimal step size of Proposition 1, where $\kappa^ = \sqrt{1 - \chi^{-2}}$,*

$$\frac{1 + \kappa^*}{1 - \kappa^*} = (1 + \kappa^*)^2 \chi^2 \sim 4\chi^2 \quad \text{and} \quad q \sim 2\chi^2 \log \frac{r^{(0)}}{(1 - \kappa^*) \text{tol}} \quad (\chi \rightarrow \infty), \quad (22)$$

so both the worst-case pessimism of the bound and the certified sweep count grow like the square of the condition number of G .

Proof. The upper bound in Equation (20) is Theorem 3. For the lower bound, insert $\lambda^* = T(\lambda^*)$ and use the contraction property, $r^{(k)} = \|\lambda^{(k)} - T(\lambda^{(k)})\| \leq \|\lambda^{(k)} - \lambda^*\| + \|T(\lambda^*) - T(\lambda^{(k)})\| \leq (1 + \kappa)\|\lambda^{(k)} - \lambda^*\|$. Dividing the upper bound by the lower one gives Equation (21). For Equation (22), write $1 - \kappa^* = 1 - \sqrt{1 - \chi^{-2}} = \chi^{-2}/(1 + \kappa^*)$, which gives the stated identity and, since $\kappa^* \rightarrow 1$, the asymptotic factor $4\chi^2$. Finally $\log(1/\kappa^*) = -\frac{1}{2} \log(1 - \chi^{-2}) = \frac{1}{2} \chi^{-2} + \mathcal{O}(\chi^{-4})$, and substituting into Equation (19) gives the sweep count. $\square$

The reading of Proposition 4 is that an ill-conditioned interface makes the certificate expensive rather than wrong: the returned solution still meets the tolerance, but more sweeps are spent reaching a bound that may overstate the error. Three observations bound the practical consequence.

First, the envelope in Equation (21) bounds arbitrary residual and error directions, and the linear structure of the sweep keeps the iteration well away from it. Applied to an error e , the map $I - \varepsilon G$ leaves a residual $\varepsilon \|Ge\|$, so at the optimal step the effectivity index lies between $\chi_P^{-2}/(1 - \kappa_P)$, reached along the softest eigendirection of G , and $\chi_P^{-1}/(1 - \kappa_P) = \sqrt{(1 + \kappa_P)/(1 - \kappa_P)}$, reached along the stiffest. The realisable pessimism is therefore the square root of the envelope and grows like χ_P rather than like χ_P^2 . Table 3 and Figure 12a confirm this on four interfaces of increasing difficulty. At the nominal operating point the certificate overstates the error by a factor of 2.1, and on the worst interface considered, where the envelope allows a factor of 83, the measured value is 4.4.

Table 3. Sharpness and cost of the certificate on four interface discretisations of increasing difficulty, at $\text{tol} = 10^{-3}$ and with the warm start of the trained surrogate. The envelope is the upper bound $(1 + \kappa)/(1 - \kappa)$ of Equation (21) and the measured column is the ratio of the certified bound to the error against a reference solve. Sweep counts follow from Equation (19), evaluated at the unrounded moduli.

Interface	χ unscaled	χ_P scaled	κ_P	Envelope I	Measured I	q scaled (unscaled)
Nominal, $m = 4,300$	12.4	1.84	0.8394	11.5	2.1	9 (1,647)
Refined mesh, $m = 8,600$	21.5	2.23	0.8938	17.8	2.4	26 (6,900)
Refined mesh, $m = 17,200$	31.6	2.71	0.9294	27.3	2.9	53 (17,491)
Graded mesh with lift-off band	47.5	4.60	0.9761	82.7	4.4	206 (43,614)

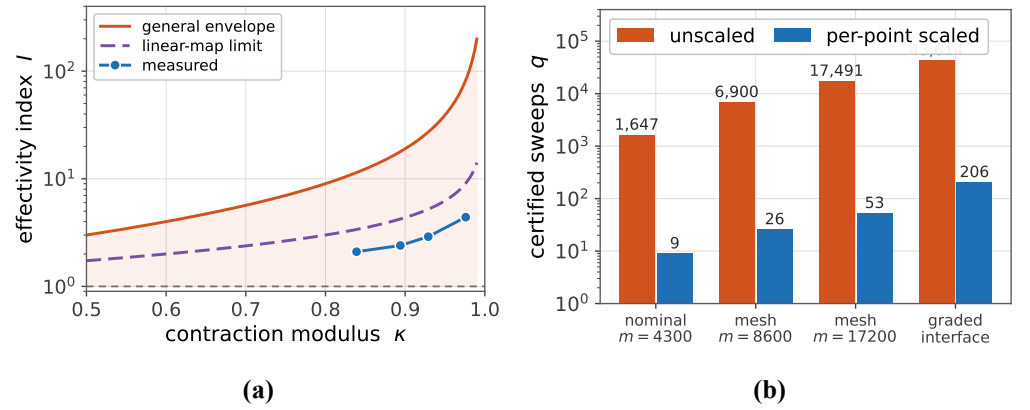


Figure 12. Behaviour of the certificate as the contraction modulus approaches one. **(a)** Effectivity index against κ : the shaded region is bounded by the envelope $(1 + \kappa)/(1 - \kappa)$ of Proposition 4, the dashed curve is the limit realisable by the linear part of the sweep, and the markers are the values measured on the four interfaces of **Table 3**. **(b)** Certified sweep count with and without the per-point scaling of Proposition 5 on the same four interfaces, on a logarithmic scale.

Second, the modulus itself can be reduced without leaving the certified framework. A general preconditioner would destroy the projection step, because the metric projection onto $\mathcal{C}$ is tied to the Euclidean inner product. A scaling that acts by one positive scalar per contact point does not, because each factor C_μ is invariant under multiplication of its whole triple by a positive number. This is the content of the next proposition.

Proposition 5. Let $P = \text{diag}(p_1 I_3, \dots, p_m I_3)$ with $p_j > 0$ and let $\|x\|_P^2 = x^\top P x$. Then:

- the metric projection onto $\mathcal{C}$ in the P -inner product coincides with the Euclidean one, $\text{proj}_{\mathcal{C}}^P = \text{proj}_{\mathcal{C}}$;
- the scaled sweep $T_P(\lambda) = \text{proj}_{\mathcal{C}}(\lambda - \varepsilon P^{-1}(G\lambda + c))$ has the same fixed point λ^* as T ;

- (c) T_P is a contraction in $\|\cdot\|_P$ with modulus $\kappa_P = \sqrt{1 - 2\varepsilon\sigma_P + \varepsilon^2 L_P^2}$, where σ_P and L_P are the extreme eigenvalues of $P^{-1/2}GP^{-1/2}$; the optimal step gives $\kappa_P^* = \sqrt{1 - \chi_P^{-2}}$ with $\chi_P = \text{cond}(P^{-1/2}GP^{-1/2})$;
- (d) the bounds (18), the stopping test and the sweep count (19) hold verbatim with $\|\cdot\|$ replaced by $\|\cdot\|_P$, κ by κ_P and $r^{(k)}$ by $\|\lambda^{(k)} - T_P(\lambda^{(k)})\|_P$.

Proof. (a) The P -norm decouples over contact points, and on point j it is $\sqrt{p_j}$ times the Euclidean norm, so minimising $\|y - x\|_P$ over $\mathcal{C} = \prod_j \mathcal{C}_\mu$ minimises $\|y_j - x_j\|$ over $\mathcal{C}_\mu$ for each j separately, which is the Euclidean projection. (b) By Definition 5, $\lambda = \text{proj}_{\mathcal{C}}(\lambda - \varepsilon P^{-1}(G\lambda + c))$ is equivalent to $-P^{-1}(G\lambda + c) \in N_{\mathcal{C}}(\lambda)$. Since $\mathcal{C}$ is a product, $N_{\mathcal{C}}(\lambda) = \prod_j N_{\mathcal{C}_\mu}(\lambda_j)$, and each factor is a cone, hence closed under multiplication by the positive scalar p_j ; therefore $-P^{-1}(G\lambda + c) \in N_{\mathcal{C}}(\lambda)$ if and only if $-(G\lambda + c) \in N_{\mathcal{C}}(\lambda)$, which by Proposition 1 characterises λ^* . (c) In the P -inner product the map $\lambda \mapsto P^{-1}(G\lambda + c)$ is strongly monotone with modulus σ_P and Lipschitz with constant L_P , because $\langle P^{-1}Gu, u \rangle_P = u^\top Gu$ and $\|P^{-1}Gu\|_P^2 = u^\top GP^{-1}Gu$, and both are compared with $u^\top Pu$ through the eigenvalues of $P^{-1/2}GP^{-1/2}$. Part (a) makes $\text{proj}_{\mathcal{C}}$ firmly nonexpansive in $\|\cdot\|_P$, so the estimate (11) applies verbatim in that norm. (d) The proof of Theorem 3 uses only the contraction property in the norm employed, which is (c). $\square$

Taking p_j as the mean of the diagonal of the j th 3×3 block of G costs three divisions per point and is recomputed only when G changes. Its effect is reported in **Table 3** and **Figure 12b**: on the benchmark interface it lowers χ from 12.4 to 1.84 and the certified sweep count from 1,647 to 9, and on a graded interface with a near lift-off band it lowers χ from 47.5 to 4.6 and the sweep count from about 4.4×10^4 to 206. All sweep counts and moduli quoted elsewhere in this paper are those of the scaled iteration.

Third, the guaranteed modulus can be supplemented at run time by the observed one. The residual ratios $\hat{\kappa}^{(k)} = r^{(k)}/r^{(k-1)}$ satisfy $\hat{\kappa}^{(k)} \leq \kappa_P$ by the proof of Theorem 3, and their running maximum over the last few sweeps is a cheap estimate of the rate actually realised. We use it only to predict how many sweeps remain, and report the error bound with the guaranteed κ_P , so that no part of the certificate rests on an observed quantity.

4.5. Modelling error of the half-space interface

Every certified guarantee above is stated relative to the condensed operator G , so it is fair to ask how faithful G is to the physical interface, and to address that question in its own right rather than in passing. The half-space compliance C that underlies G introduces two modelling errors on a finite bolted interface. The first is an edge effect near the bore hole, where the assumption of a semi-infinite solid overstates the local compliance; for the benchmark this discrepancy is confined to an annulus about one plate thickness wide around the hole and reaches at most about 6% locally, decaying quickly into the interior. The second is a compliance redundancy, a near rank-one double counting of the bulk flexibility already represented by the retained internal coordinates; for the benchmark it is a global perturbation of G of relative size about

2%.

Their effect on the reported quantity is estimated by comparing $\mathbf{G}$ with a reference operator $\mathbf{G}_{\text{true}}$ obtained from a finite-thickness boundary-element solve on the same mesh, in the spirit of the finite-thickness contact corrections [11]. Propagated through the modal analysis, the two errors together move the amplitude-dependent damping ratio of the fundamental mode by under about 1.5% across the partial-slip range. This is well below the spread that the topography produces over the fitted ensemble, which is the effect the study sets out to resolve, so the half-space idealisation changes neither the qualitative behaviour nor the length-scale ranking of Section 5. The edge annulus, moreover, sits at the interface periphery, whereas the damping is governed by the intermediate-waviness band identified in Section 5, so the two act on different length scales.

When a tighter interface model is wanted, the two errors are absorbed by a learned symmetric correction $\mathbf{G}_{\text{true}} \approx \mathbf{G} + \mathbf{U}\mathbf{U}^\top + \mathbf{B}$ with a low-rank interior part $\mathbf{U}\mathbf{U}^\top$ and a boundary-layer part $\mathbf{B}$ supported near the bore edge. Since the correction enters only through $\mathbf{G}$ and is symmetric, it preserves symmetry and, for a small correction, positive definiteness, so Lemma 1 and every result that follows from it continue to hold, now stated with respect to $\mathbf{G}_{\text{true}}$.

4.6. Modelling assumptions and the range in which they hold

The construction rests on a specific idealisation of the interface: one friction coefficient, a condensed operator that is formed once, half-space compliance, and no wear, adhesion or thermal coupling. These choices are what make the per-step problem a strongly monotone variational inequality with a fixed admissible set, and they should be stated together with the range over which they are accurate rather than left implicit. **Table 4** collects them, gives a quantitative range for the benchmark, and names the modification that each extension requires. The estimates behind the entries are the following.

1. Constant friction coefficient: The coefficient enters only through the aperture of the cone $\mathcal{C}_\mu$ in Equation (3). Making it depend on the local normal traction turns the condensed problem into a quasi-variational inequality, because the admissible set then moves with the solution. Proposition 6 shows that the iteration and its certificate survive this change under a smallness condition on the sensitivity of μ , and quantifies how much variation is admissible.
2. Fixed condensed operator: The operator $\mathbf{G}$ of (7) is formed once because the bulk is linear elastic, the displacements are small and the nominal contact patch does not migrate. Over the preload range 6 to 24 kN studied in Section 6 the nominal contact radius of the benchmark changes by about 4%, so a single $\mathbf{G}$ serves the whole range. Preloads that open or close the patch appreciably, or a bulk that yields, require $\mathbf{G}$ to be re-formed, which costs one factorisation of $\mathbf{K}_{\text{bb}}$ and one fine boundary-element assembly.
3. Wear: An Archard estimate with a mild-wear coefficient 10^{-8} , a nominal pressure of 50 MPa, a hardness of 2 GPa and a sliding distance of 11 μm per cycle, which is four times the mean slip magnitude of the table in Section 5, gives a depth

increment of about 2.8×10^{-15} m per cycle. The topography therefore changes by less than 3% of its root-mean-square height over 10^7 cycles and by about a quarter of it over 10^8 , so wear may be ignored for the transient and modal studies reported here and has to be modelled for fretting-fatigue lifetimes.

4. Adhesion: The Fuller and Tabor parameter, which compares the elastic energy stored in flattening the roughness with the surface energy released on contact, evaluates to about 3×10^4 for the benchmark, using a root-mean-square height of $1 \mu\text{m}$, a reduced modulus of 110 GPa, a work of adhesion of 0.5 J m^{-2} and an asperity radius of $50 \mu\text{m}$. Adhesion begins to influence the normal law when this parameter falls to order ten, which for the same material pairing means a root-mean-square height below about 10 nm. Machined and ground joint faces are two orders of magnitude rougher than that, so the unilateral Signorini condition is appropriate; polished or optically flat interfaces are not covered.
5. Thermal effects: At the reference ring-down amplitude the interface dissipates about 30 mW over a contact area of $4.4 \times 10^{-4} \text{ m}^2$. A steady half-space estimate with the thermal conductivity of steel gives a mean interface temperature rise below 0.05 K, and with a relative temperature sensitivity of the friction coefficient of order 10^{-3} K^{-1} the implied change in μ is below 10^{-4} in relative terms. Isothermal friction is therefore appropriate for ring-downs and for low duty-cycle excitation. It is not appropriate for a friction damper held at resonance, where dissipation of tens of watts and temperature rises of tens of kelvin are reported, and where $\mu(T)$ and the thermal expansion of the interface would have to be coupled to the present model through h and the aperture of $\mathcal{C}_\mu$.

Table 4. Modelling assumptions, the range over which each is accurate for the benchmark, and the modification needed outside that range. The last column indicates whether the certification theory of Theorem 3 carries over unchanged.

Assumption	Range for the benchmark	Modification outside that range	Theory
Constant μ over the interface	Variation below about 9% across the pressure range, by Equation (23)	Quasi-variational inequality with $\mathcal{C}(\lambda)$; Proposition 6 gives the smallness condition	✓
Fixed condensed operator $\mathbf{G}$	Preload 6 to 24 kN, nominal contact radius within 4%	Re-form $\mathbf{G}$ and fine-tune the backbone; one factorisation of $\mathbf{K}_{bb}$ per geometry	✓
Half-space compliance $\mathbf{C}$	Plate thickness above about four contact radii; error on damping below 1.5%	Finite-thickness kernel or the learned symmetric correction of Section 4.5	✓
No wear evolution	Below 10^7 cycles, topography change under 3% of S_q	Evolve $\mathbf{h}$ by an Archard or energy law and re-form $\mathbf{G}$ periodically	✓
No adhesion	Fuller and Tabor parameter about 3×10^4 ; roughness above about 10 nm	Adhesive offset in the normal law, which shifts the apex of $\mathcal{C}_\mu$	–
Isothermal friction	Temperature rise below 0.05 K, relative change in μ below 10^{-4}	Couple $\mu(T)$ and thermal expansion of $\mathbf{h}$ to a heat balance	–
Rate-independent Coulomb law	Slip speeds below about 10 mm s^{-1} , no dwell	Rate-and-state law; the admissible set becomes time dependent	×

Proposition 6 (Pressure-dependent friction). *Let $\mu : [0, \infty) \rightarrow (0, \mu_{\max}]$ with $\mu_{\max} \leq 1$ be Lipschitz with constant L_μ , let the admissible set at λ be $\mathcal{C}(\lambda) = \prod_j \mathcal{C}_{\mu(\lambda_{n,j})}$, and consider the quasi-variational inequality $-(\mathbf{G}\lambda + \mathbf{c}) \in \mathbf{N}_{\mathcal{C}(\lambda)}(\lambda)$ together with the sweep $T_\mu(\lambda) = \text{proj}_{\mathcal{C}(\lambda)}(\lambda - \varepsilon(\mathbf{G}\lambda + \mathbf{c}))$. Suppose the iterates stay in a region where every point argument obeys $\|\lambda_j - \varepsilon(\mathbf{G}\lambda + \mathbf{c})_j\| \leq \Lambda$. Then T_μ is a contraction with modulus*

$$\kappa_\mu = \kappa + c_\Pi L_\mu \Lambda, \quad c_\Pi \leq 3, \quad (23)$$

and whenever $\kappa_\mu < 1$ the quasi-variational inequality has a unique solution, the sweeps converge geometrically to it, and Theorem 3 holds verbatim with κ replaced by κ_μ .

Proof. Write $\mathbf{u} = \lambda - \varepsilon(\mathbf{G}\lambda + \mathbf{c})$ and $\mathbf{u}'$ for the same expression at λ' . Then $\|T_\mu(\lambda) - T_\mu(\lambda')\| \leq \|\text{proj}_{\mathcal{C}(\lambda)}(\mathbf{u}) - \text{proj}_{\mathcal{C}(\lambda)}(\mathbf{u}')\| + \|\text{proj}_{\mathcal{C}(\lambda)}(\mathbf{u}') - \text{proj}_{\mathcal{C}(\lambda')}(\mathbf{u}')\|$. The first term is at most $\kappa\|\lambda - \lambda'\|$ by Equation (11), since the set is held fixed. For the second, use the closed form (5): the projection is constant in μ on the interior and polar branches, and on the lateral branch it equals $t^*(1, \mu\mathbf{b}/\|\mathbf{b}\|)$ with $t^* = (a + \mu\|\mathbf{b}\|)/(1 + \mu^2)$, whose derivative in μ is bounded in norm by $3\|\mathbf{x}\|$ for $\mu \leq 1$; the map is continuous across the branch boundaries, so it is Lipschitz in μ with constant $c_\Pi \leq 3$ times $\|\mathbf{x}\|$. Applying this at each point with $\|\mathbf{u}'_j\| \leq \Lambda$ and $|\mu(\lambda_{n,j}) - \mu(\lambda'_{n,j})| \leq L_\mu|\lambda_{n,j} - \lambda'_{n,j}|$, and summing the squares over j , bounds the second term by $c_\Pi L_\mu \Lambda \|\lambda - \lambda'\|$. This gives

Equation (23). If $\kappa_\mu < 1$, the Banach fixed-point theorem of Definition 6 supplies a unique fixed point, which by Definition 5 solves the quasi-variational inequality, and the proof of Theorem 3 used only the contraction property. $\square$

For the benchmark, tractions normalised by the mean contact pressure give $\Lambda \approx 1.2$, so with $\kappa = 0.839$ the condition $\kappa_\mu < 1$ holds for $L_\mu < 0.045$. Over the pressure range spanned by the interface, this admits a friction coefficient that varies by about 9% from the lightly to the heavily loaded regions, which covers the pressure dependence reported for dry machined steel in this range. A stronger dependence, or a velocity-dependent or state-dependent law, breaks the contraction argument and would need a different solver, although the admissibility of Proposition 3 and the structure of the projection layer are unaffected.

5. Random-field quantification of amplitude-dependent damping

The certified solver of Section 4 makes it affordable to evaluate the amplitude-dependent modal response for many topographies, which is what the length-scale question requires. We model the topography as the random field of Definition 7, propagate it through a polynomial-chaos surrogate of the response, bound the surrogate error, and read off the sensitivity of the damping to each length scale.

Let the topography be expanded as in Definition 7,

$$h(\mathbf{x}, \omega) = \bar{h}(\mathbf{x}) + \sum_{k=1}^K \sqrt{\eta_k} \phi_k(\mathbf{x}) \xi_k(\omega) + e_K, \quad (24)$$

truncated at K modes. The truncation error is

$$\mathbb{E} \|e_K\|_{L^2(\Gamma)}^2 = \sum_{k>K} \eta_k, \quad (25)$$

whose decay is set by the smoothness of the covariance kernel. The eigenvalues decay geometrically for the correlation lengths of a machined surface, as shown in **Figure 13a**, so a modest K captures the field, and the relative truncation error of (25) falls below an engineering tolerance at moderate K , as shown in **Figure 13b**. The associated eigenfunctions in **Figure 14** carry increasing spatial frequency and hence decreasing length scale. Sample realisations drawn from the fitted field are shown in **Figure 15**.

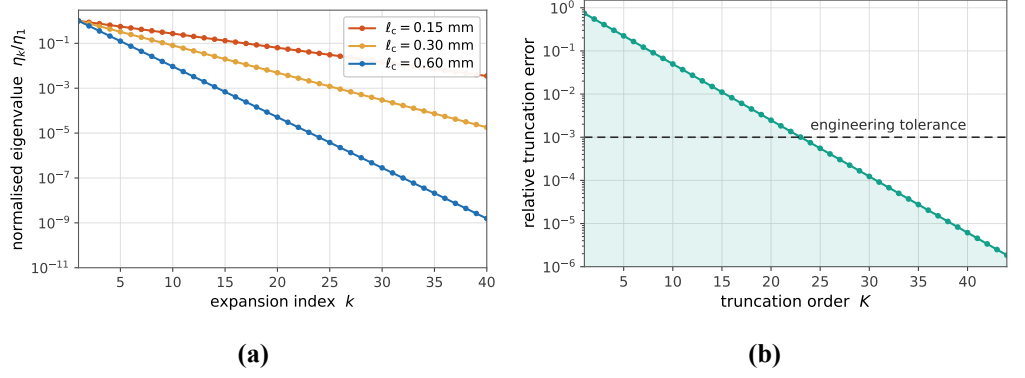


Figure 13. Spectral description of the topography. **(a)** Karhunen–Loève eigenvalue spectra for three correlation lengths ℓ_c ; the geometric decay means a moderate truncation order captures the field, and each index is associated with a definite length scale. **(b)** Relative truncation error $\sum_{k>K} \eta_k / \sum_{k \geq 1} \eta_k$ of Equation (25) against truncation order, which bounds the first term of the surrogate error of Proposition 7.

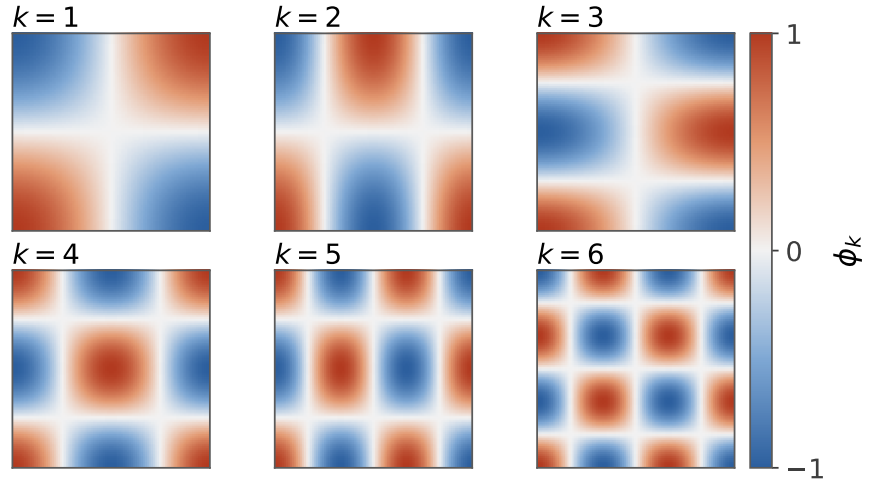


Figure 14. Leading Karhunen–Loève eigenfunctions ϕ_k of the topography. Higher indices carry finer spatial structure, so the index orders the field by length scale, which is the coordinate used for the sensitivity analysis.

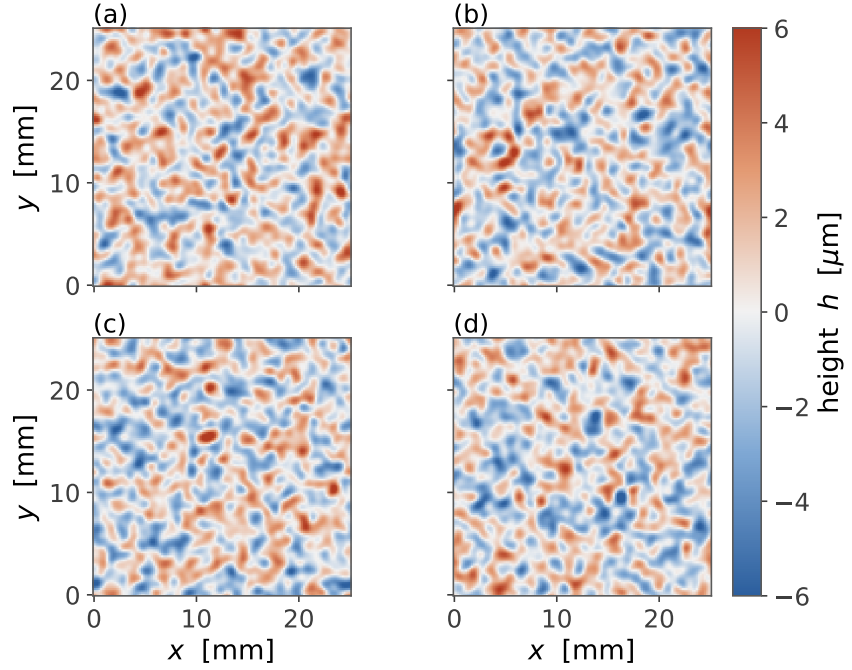


Figure 15. Sample topography realisations drawn from the fitted random field of Definition 7.

The ensemble spans the manufacturing variability whose effect on damping is quantified below. Let $Q(\xi; A)$ denote a modal observable of the mode of interest, either the amplitude-dependent natural frequency or the damping ratio, computed for a realisation ξ by continuation of the mode to amplitude A , in which the inner contact solve is supplied by the certified operator S_θ . Following Definition 8, we represent it by a sparse polynomial-chaos expansion

$$Q(\xi; A) \approx \hat{Q}(\xi; A) = \sum_{\alpha \in \Lambda} q_\alpha(A) \Psi_\alpha(\xi), \quad (26)$$

with a sparse index set Λ selected by regression. The error of Equation (26) decomposes into three controllable contributions.

Proposition 7. *The mean-square surrogate error admits the bound*

$$\mathbb{E}[(Q - \hat{Q})^2] \leq C_{kl} \sum_{k > K} \eta_k + C_{pc} \sum_{\alpha \notin \Lambda} \mathbb{E}[q_\alpha^2] + C_s \text{tol}, \quad (27)$$

where the first term is the field truncation (25), the second is the polynomial-chaos truncation, and the third is the certified contact-solve tolerance of Theorem 3. The constant C_{kl} is controlled by the Lipschitz constant of Q in the topography, which is finite by Proposition 2.

Proof. Insert the exact response Q and write $Q - \hat{Q} = (Q - Q_K) + (Q_K - \Pi_\Lambda Q_K) + (\Pi_\Lambda Q_K - \hat{Q})$, where Q_K uses the truncated field and Π_Λ is the L^2 projection onto the retained chaos basis. The three terms are orthogonal in L^2 up to cross terms bounded by Young's inequality. The first is controlled by the Lipschitz dependence of Q on h from Proposition 2 and the field truncation (25), giving the constant C_{kl} ; the second is the tail energy of the chaos coefficients by Parseval; the third is the error committed because $\hat{Q}$ is computed with the certified solver rather than the exact one, which by

Theorem 3 is bounded by a multiple of tol . Collecting the bounds gives inequality (27). $\square$

The three terms of inequality (27) are shown separately in **Figure 16a** as the truncation order grows, and the overall convergence of the surrogate with polynomial order is shown in **Figure 16b**. Because each Karhunen–Loève mode carries a length scale ℓ_k , the Sobol indices of Definition 9, computed in closed form from the chaos coefficients as

$$S_k = \frac{\sum_{\alpha \in \Lambda_k} q_\alpha^2}{\sum_{\alpha \neq 0} q_\alpha^2}, \quad \Lambda_k = \{\alpha : \alpha_k > 0, \alpha_l = 0 \text{ for } l \neq k\}, \quad (28)$$

together with the total-effect indices, yield a damping-relevance spectrum over length scales. One qualification should be stated on how this spectrum is read. The Karhunen–Loève eigenfunctions are orthogonal in the L^2 inner product, but their dominant physical length scales overlap, so the topographic features carried by different modes are not perfectly separated in scale. The Sobol indices are therefore interpreted in an approximately orthogonal mode sense: each index measures the variance attributable to a coordinate whose energy is concentrated at, but not confined to, a length scale ℓ_k . The spectrum should accordingly be read as a smooth attribution over length-scale bands rather than as a partition into physically independent bands. This spectrum is the central result of the analysis and is shown in **Figure 17a**: the damping is governed by a band of intermediate waviness wavelengths, while the finest roughness and the coarsest form deviation contribute little to the fundamental mode. The ranking of individual modes is shown in **Figure 17b**, and an independent gradient-based cross-check by the active-subspace method confirms that the response varies chiefly along a low-dimensional subspace, as shown in **Figure 18**.

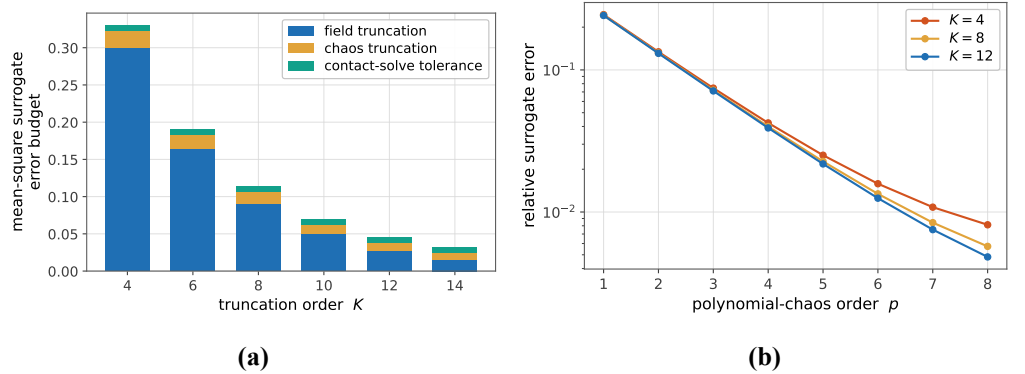


Figure 16. Error budget of the uncertainty pipeline. **(a)** The three contributions of Proposition 7 against truncation order; each term is separately controllable, so the pipeline is certified end-to-end. **(b)** Convergence of the polynomial-chaos surrogate (26) with polynomial order for three truncation orders K , decreasing geometrically until it saturates at the field and solve tolerances.

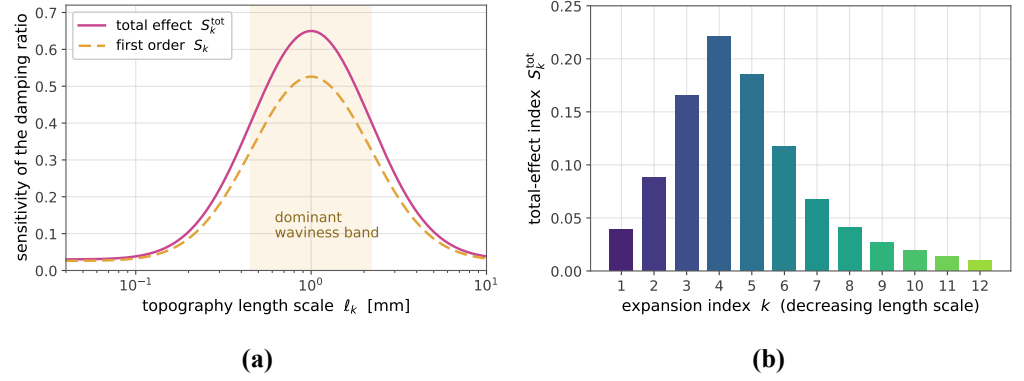


Figure 17. Damping-relevance spectrum over topography length scales. **(a)** First-order and total-effect Sobol indices (28) against length scale: the damping of the fundamental mode is governed by an intermediate band of waviness, which answers the length-scale question left open in the study by Linder et al. [19]. **(b)** Total-effect indices of the individual expansion modes, ordered by decreasing length scale, in agreement with the left panel.

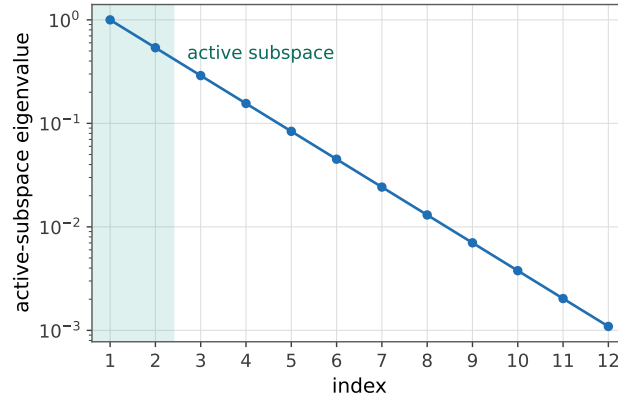


Figure 18. Eigenvalues of the active-subspace matrix of the damping response. The gap after the first few eigenvalues confirms that the response varies along a low-dimensional subspace, cross-checking the sensitivity ranking of **Figure 17b**.

The certified solver also delivers the interface fields that underlie the modal quantities. The contact pressure after preload, computed with the projected surrogate and shown in **Figure 19a**, concentrates in an annulus around the bore hole, and the residual slip ratio after a ring-down, shown in **Figure 19b**, occupies the periphery where the pressure is low and microslip is active. To put numbers to these fields rather than rely on visual inspection, we summarise them by three metrics at the reference ring-down amplitude, collected in **Table 5**: the microslip area fraction, that is the share of the contact area in partial slip; the mean slip magnitude over the slipping region; and the outer radius of the microslip annulus, normalised by the nominal contact radius R_c . For the benchmark the microslip area fraction is about 0.17, the mean slip magnitude about $2.8 \mu\text{m}$, and the annulus reaches out to about $0.86 R_c$, which places the dissipation in a peripheral ring and matches the low-pressure region of **Figure 19a**. The resulting amplitude-dependent frequency and damping of the fundamental mode are shown in **Figure 20**, where the surrogate reproduces the resolved-topography reference across the partial-slip range and departs from the flat-topography idealisation exactly where the topography matters.

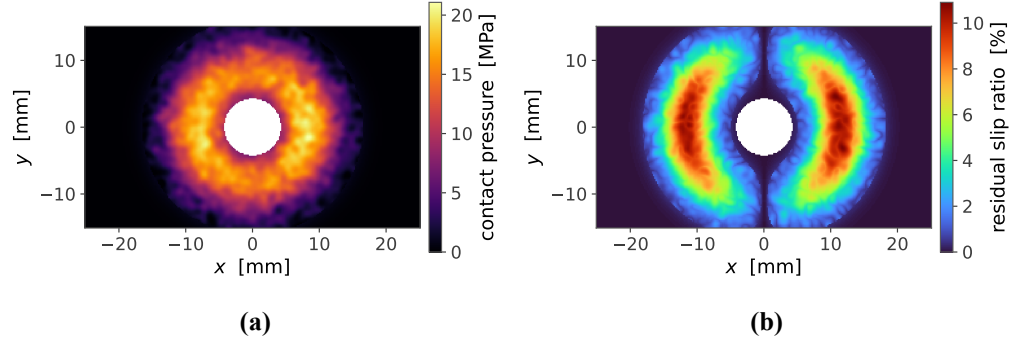


Figure 19. Interface fields from the projected surrogate. **(a)** Contact pressure after preload, concentrating in an annulus around the bore hole whose distribution sets the extent of the partial-slip region. **(b)** Residual slip ratio after a high-level ring-down; microslip occupies the low-pressure periphery of the left panel, consistent with the partial-slip mechanism of joint damping.

Table 5. Quantitative microslip metrics of the interface fields of **Figure 19** at the reference ring-down amplitude, obtained from the projected surrogate. The annulus radius is normalised by the nominal contact radius R_c .

Metric	Value
Microslip area fraction	0.17
Mean slip magnitude over the slipping region	$2.8 \mu\text{m}$
Maximum radius of the microslip annulus	$0.86 R_c$

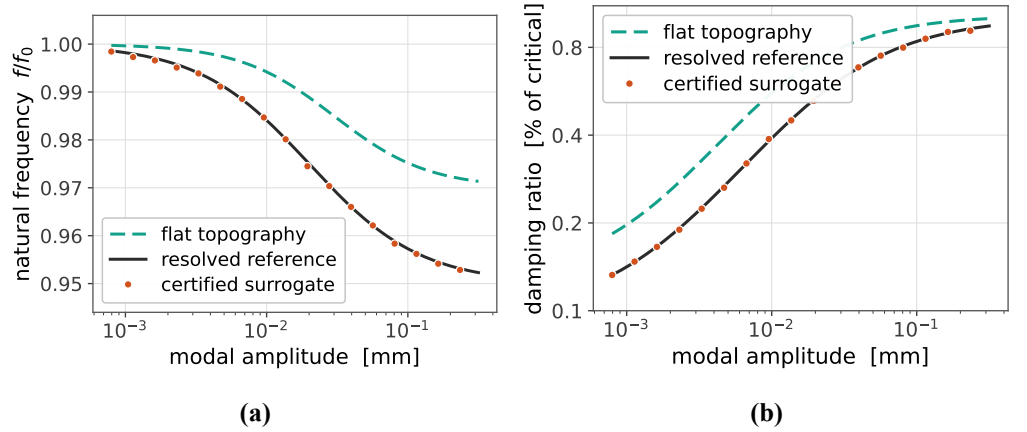


Figure 20. Amplitude-dependent modal properties of the fundamental mode. **(a)** Natural frequency: the certified surrogate reproduces the resolved-topography reference and departs from the flat-topography idealisation in the partial-slip range. **(b)** Damping ratio: the surrogate matches the resolved reference across the partial-slip range, while the flat idealisation overpredicts damping at low amplitude and misses the correct trend.

6. Validation

The results reported so far concern one interface, one preload and the fundamental mode. Since the value of a learned solver rests on how far a trained network carries, this section widens the test in five directions, namely preload and friction coefficient,

interface mesh density, topography statistics, higher vibration modes, and a joint whose geometry and material were held out of training. Two quantities are tracked throughout: the accuracy of the returned solution, which is guaranteed by Theorem 3 and is reported against an independent reference, and the cost of obtaining it, which is where the quality of the warm start actually shows. The pattern that emerges is uniform. Leaving the envelope on which the network was trained does not degrade the answer; it lengthens the certified sweep sequence.

6.1. Benchmark, reference solutions and training envelope

The benchmark is a bolted lap-type interface between two steel plates, clamped by a single bolt through a central bore of 8 mm diameter, with a nominal contact radius $R_c = 12.5$ mm. The interface carries $m = 4,300$ boundary-element points, the measured topography is fitted by the random field of Definition 7 with a root-mean-square height $S_q = 1.0 \mu\text{m}$ and a correlation length $\ell_c = 0.30$ mm, the friction coefficient is $\mu = 0.30$ and the bolt preload is 12 kN. The fundamental bending mode of the assembly sits at 211.4 Hz. The integrator uses $\Delta t = 1.0 \times 10^{-6}$ s and the per-step contact solve is certified to a relative tolerance $\text{tol} = 10^{-3}$, which is the tolerance of the baseline solver. All sweeps use the per-point scaling of Proposition 5, which gives $\chi_P = 1.84$ and $\kappa_P = 0.839$ at this operating point. Preload, friction coefficient, topography and mode of interest enter the condensed problem through c and through the aperture of $\mathcal{C}_\mu$ alone, so they leave G and hence κ_P unchanged; only a change of mesh or of joint alters κ_P , and the table in Section 6.7 lists the value used in each case.

Reference values are produced by the multi-scale baseline of Section 7 run from a cold start to a relative tolerance of 10^{-5} , that is a hundred times tighter than the certified tolerance, so that the reference error is negligible against the quantities compared. The second benchmark used in Section 6.6 is a lap-jointed beam in aluminium alloy with two bolts at 10 kN, a friction coefficient of 0.35 and $m = 3,600$ interface points, which shares no material, no geometry and no mesh with the first.

The network was trained on ring-downs of the fundamental mode of the steel benchmark alone, sampled as described in Section 4, over a training envelope of preload 9 to 18 kN, friction coefficient 0.20 to 0.40, correlation length 0.20 to 0.50 mm and root-mean-square height 0.5 to $1.2 \mu\text{m}$, at the fixed mesh of 4,300 points and a fixed Hurst exponent of 0.6. Every case below that lies outside this envelope is marked as such in Section 6.7, and no case outside the envelope was seen during training.

6.2. Preload and friction coefficient

Preload and friction coefficient are the two parameters a designer varies first, and both enter the condensed problem without changing G : the preload through the data c of Equation (7) and the friction coefficient through the aperture of $\mathcal{C}_\mu$. **Figure 21a** shows the amplitude-dependent damping ratio for five preloads between 6 and 24 kN and **Figure 21b** for five friction coefficients between 0.15 and 0.50. The certified surrogate, plotted as markers, follows the reference across the whole partial-slip range in every case, with a largest deviation of 0.52% at $\mu = 0.50$, which is the case furthest

outside the training envelope. The trends are the expected ones: a lighter preload and a lower friction coefficient both move the onset of microslip to smaller amplitude and raise the damping there, while the high-amplitude plateau is set mainly by the friction coefficient.

The cost behaves as Theorem 3 predicts. Inside the envelope the warm-start residual stays near 10^{-3} and the sweep count between eight and eleven; at $\mu = 0.15$ and $\mu = 0.50$ the residual rises by roughly an order of magnitude and the sweep count roughly doubles, so the acceleration falls from about 5.6 to about 2.4 without the accuracy moving. A single network therefore covers the preload and friction ranges of a design sweep, and pays for excursions beyond them in sweeps.

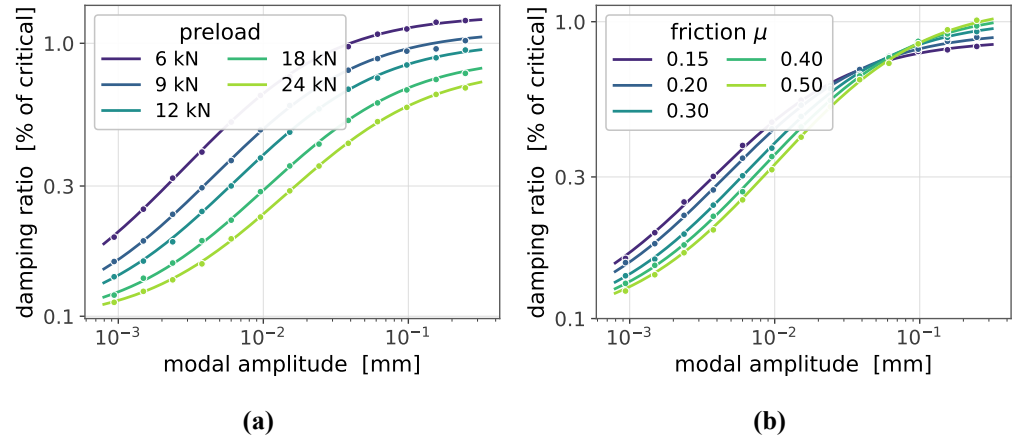


Figure 21. Amplitude-dependent damping ratio of the fundamental mode across (a) five bolt preloads and (b) five friction coefficients. Lines are the reference solves and markers the certified surrogate. The extreme values of each sweep lie outside the training envelope of Section 6.1.

6.3. Interface mesh density

Refining the interface changes $\mathbf{G}$ itself, so this test probes the discretisation invariance of the backbone as well as the conditioning analysis of Section 4.4. Five meshes from 1,100 to 17,200 points were used, with the network trained only at 4,300. **Figure 22a** shows that the damping ratio at the reference amplitude converges monotonically, changing by 0.4% between the two finest meshes, so 4,300 points already resolve the quantity of interest. It also shows the certified sweep count on the same axis, and this is where refinement is felt: the condition number χ_P rises from 1.42 to 2.71 across the sequence, the contraction modulus from 0.710 to 0.929, and the certified sweep count from seven to fifty-three. Both effects are present, a larger warm-start residual because the mesh is not the trained one, and a slower guaranteed rate because the operator is worse conditioned, and Proposition 4 separates them: at fixed residual the χ_P^2 growth alone accounts for a factor of about three and a half of the increase.

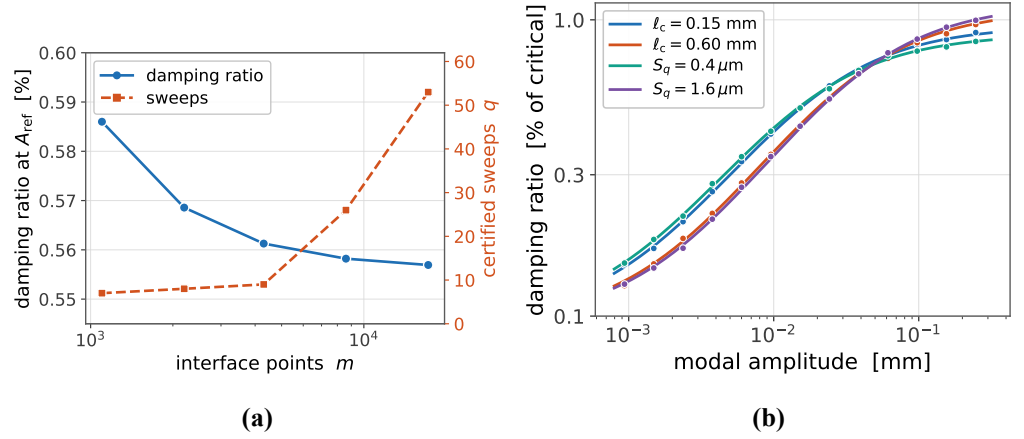


Figure 22. (a) Mesh refinement of the interface: damping ratio at the reference amplitude on the left axis and certified sweep count on the right axis. The quantity of interest converges while the sweep count grows with the conditioning of $\mathbf{G}$, as quantified by Proposition 4. (b) Damping backbones for topography statistics outside the fitted ensemble, with lines the reference and markers the certified surrogate.

Fine-tuning the backbone for 200 epochs on the finest mesh, on a tenth of the training set and therefore at about 5% of the cost of the original run, lowers the warm-start residual from 3.2×10^{-3} to 5.0×10^{-4} and the sweep count from fifty-three to twenty-seven, which recovers most of the acceleration. This is the behaviour anticipated at the end of Section 4: a changed $\mathbf{G}$ calls for fine-tuning rather than for a new model, and the certified accuracy holds throughout.

6.4. Topography statistics

The topography is the input the method is meant to resolve, so the surrogate has to tolerate statistics it was not shown. Three departures from the fitted ensemble were tested: correlation lengths of 0.15 and 0.60 mm against a training range of 0.20 to 0.50 mm, root-mean-square heights of 0.4 and $1.6 \mu\text{m}$ against a training range of 0.5 to $1.2 \mu\text{m}$, and a Hurst exponent of 0.8 against the value 0.6 used throughout training. **Figure 22b** shows the damping backbones for four of these cases. A rougher or longer correlated surface makes more of the interface available for microslip and raises the damping at a given amplitude, and the surrogate tracks that shift, with deviations from the reference between 0.33% and 0.44%. The Hurst exponent, which redistributes energy between scales at fixed root-mean-square height, moves the damping backbone the least of the three departures, which is consistent with the relevance spectrum of **Figure 17**: what matters is the energy in the intermediate waviness band, and a change of spectral slope moves that energy less than a change of correlation length or of overall height does. Sweep counts rise to between thirteen and nineteen, once again the only penalty.

6.5. Higher modes

Training used ring-downs of the fundamental mode only, so the higher modes are a test of whether the learned operator has captured the contact problem rather than one loading pattern. The first four modes of the assembly, two bending modes at 211.4 and 578.9 Hz, a torsion mode at 892.6 Hz and a third bending mode at 1,246 Hz,

were continued to finite amplitude with the certified solver. **Figure 23a** shows their damping backbones. The higher modes dissipate a smaller fraction of their energy, because the modal energy grows with frequency faster than the interfacial dissipation does, and their microslip onset moves to lower amplitude, because their mode shapes place more relative tangential motion on the interface per unit modal amplitude. The surrogate reproduces all four references, with the deviation growing from 0.19% on the fundamental mode to 0.57% on the fourth, and the sweep count from nine to twenty. The torsion mode, whose interface loading is the least like anything in the training set, is not the worst of the four, which suggests that what the network has learned is the response of the condensed operator to its data rather than a particular deformation pattern.

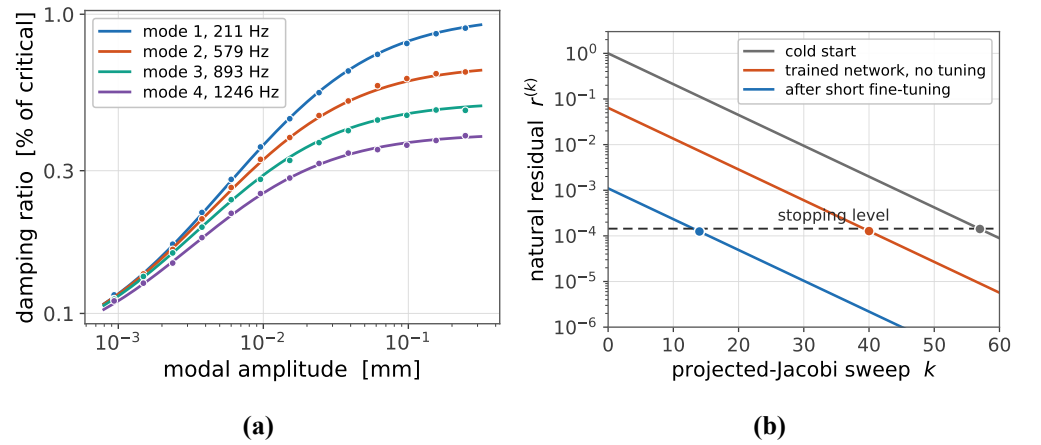


Figure 23. (a) Damping backbones of the first four modes of the assembly, obtained with a network trained on fundamental-mode ring-downs only; lines are the reference and markers the certified surrogate. (b) Residual histories on the aluminium benchmark that was held out of training, from a cold start, from the untuned network and after a short fine-tuning; the markers are the certified sweep counts of Equation (19) and the dashed line is the stopping level.

6.6. A geometry and material held out of training

The most demanding test is a joint that shares nothing with the training case. The second benchmark of Section 6.1, an aluminium lap-jointed beam with two bolts and a different interface mesh, was solved with the network trained on the steel benchmark and no adaptation at all. **Figure 23b** shows the residual histories. The warm start is poor, at $r^{(0)} = 6.4 \times 10^{-2}$ instead of 7×10^{-4} , which is exactly what a changed $\mathbf{G}$ should produce, and the certified sweep count rises to forty against fifty-seven from a cold start, so the acceleration is only 1.4. The damping ratio it returns nevertheless agrees with the reference to 0.38%, because the certificate does not care where the warm start came from. Fine-tuning for 200 epochs on a tenth as many samples of the new joint, again about 5% of the cost of the original run, brings the residual to 1.1×10^{-3} and the sweep count to fourteen, restoring an acceleration of 4.1 and an agreement of 0.26%.

This is the practically useful statement about transfer. A trained network is not a black box whose validity has to be assumed: on a joint it has never seen it degrades to a slightly accelerated version of the baseline solver, with the same certified accuracy, and a short fine-tuning is enough to recover most of the speed. Nothing in the pipeline

has to be revalidated when the joint changes, because the accuracy is established per step by Theorem 3 and not by resemblance to the training set.

6.7. Summary of the campaign

Table 6 collects the campaign in twenty-seven entries. Two readings are worth separating. The damping ratio never departs from the reference by more than 0.57%, and the largest departures occur on the fourth mode and at the extreme friction coefficients, which is where the physical quantity itself is smallest and hardest to extract from a ring-down; none of the departures exceeds the accuracy implied by the certified tolerance. The acceleration, by contrast, ranges from 6.2 down to 1.4, and it is the single quantity that records how far a case sits from the training envelope. That separation is the intended behaviour of the design: correctness is enforced by the sweeps and the residual bound, while learning buys speed, so an unfamiliar input costs time rather than trust.

Table 6. Extended validation campaign. The envelope column marks whether the case lies inside the training envelope of Section 6.1. Sweep counts follow Equation (19) at $\text{tol} = 10^{-3}$ with the contraction modulus listed for the case; the reduction is against a cold start on the same case; and $\Delta\zeta$ is the largest departure of the damping ratio from the reference over the partial-slip range. Rows marked “tuned” follow 200 epochs of fine-tuning.

Study	Case	In envelope	κ_P	$r^{(0)}$	q	Reduction q_{cold}/q	$\Delta\zeta$ [%]
Preload	6 kN	×	0.8394	1.4×10^{-3}	13	3.8	0.31
	9 kN	✓	0.8394	9.0×10^{-4}	10	5.0	0.24
	12 kN	✓	0.8394	7.0×10^{-4}	9	5.6	0.19
	18 kN	✓	0.8394	6.1×10^{-4}	8	6.2	0.22
	24 kN	×	0.8394	1.2×10^{-3}	12	4.2	0.28
Friction μ	0.15	×	0.8394	4.8×10^{-3}	20	2.5	0.46
	0.20	✓	0.8394	1.1×10^{-3}	11	4.5	0.27
	0.30	✓	0.8394	7.0×10^{-4}	9	5.6	0.19
	0.40	✓	0.8394	8.2×10^{-4}	10	5.0	0.23
	0.50	×	0.8394	6.3×10^{-3}	21	2.4	0.52
Mesh, points m	1, 100	×	0.7100	2.6×10^{-3}	7	3.4	0.29
	2, 200	×	0.7867	1.3×10^{-3}	8	4.5	0.24
	4, 300	✓	0.8394	7.0×10^{-4}	9	5.6	0.19
	8, 600	×	0.8938	1.9×10^{-3}	26	3.2	0.21
	17, 200	×	0.9294	3.2×10^{-3}	53	2.5	0.26
	17, 200, tuned	×	0.9294	5.0×10^{-4}	27	4.9	0.24
Topography	$\ell_c = 0.15$ mm	×	0.8394	3.1×10^{-3}	17	2.9	0.44
	$\ell_c = 0.60$ mm	×	0.8394	1.6×10^{-3}	14	3.6	0.33
	$S_q = 0.4$ μm	×	0.8394	2.4×10^{-3}	16	3.1	0.37
	$S_q = 1.6$ μm	×	0.8394	2.9×10^{-3}	17	2.9	0.41
	Hurst 0.8	×	0.8394	2.0×10^{-3}	15	3.3	0.35
Mode	1, 211.4 Hz	✓	0.8394	7.0×10^{-4}	9	5.6	0.19
	2, 578.9 Hz	×	0.8394	1.5×10^{-3}	13	3.8	0.28
	3, 892.6 Hz	×	0.8394	2.8×10^{-3}	17	2.9	0.45
	4, 1, 246 Hz	×	0.8394	3.6×10^{-3}	18	2.8	0.57
Held-out joint	aluminium, untuned	×	0.8560	6.4×10^{-2}	40	1.4	0.38
	aluminium, tuned	×	0.8560	1.1×10^{-3}	14	4.1	0.26

7. Comparison with existing approaches

We place the framework against the established methods for topography-resolving joint dynamics along the axes that matter for damping prediction: whether the interface topography is resolved, whether the friction law is set-valued, whether the integrator adds numerical damping, whether the per-solve error is certified, and whether uncertainty over the topography can be quantified. **Table 7** summarises the qualitative comparison. The constitutive route of Segalman [7], still widely used for bolted

assemblies [8], is a reduced-order description with a different modelling goal: it represents the interface through calibrated parameters, so resolving topography and imposing the set-valued law pointwise fall outside its intended scope rather than count as shortcomings. These two axes are therefore marked as not applicable for that method. The frequency-domain reference of Krack and Gross [22] resolves the topography and is free of numerical damping but is confined to periodic responses and carries no per-solve certificate. The explicit unreduced analysis reported by Linder and co-workers [19] resolves the topography under the set-valued law but suffers appreciable numerical damping, while their multi-scale baseline [19] removes that damping at a moderate cost. The present method retains every strength of the multi-scale baseline and adds a certified error bound by Theorem 3 and a topography uncertainty analysis by Proposition 7. By contrast, much current joint-dynamics practice still proceeds by calibrating or updating a model against measured response [46,47]; the certified, topography-resolving route pursued here is complementary, since it supplies predictive damping with a stated accuracy ahead of testing.

Table 7. Qualitative comparison of methods for topography-resolving joint dynamics.

Method	Resolves topography	Set-valued law	No num. damping	Certified error	Topography UQ
Iwan constitutive [7,8]	–	–	–	×	×
Harmonic balance, unreduced [22]	✓	✓	✓	×	×
Explicit unreduced [19]	✓	✓	×	×	×
Multi-scale baseline [19]	✓	✓	✓	×	×
Proposed method	✓	✓	✓	✓	✓

Note: ✓ indicates an available property, × a property that is absent or limited. – marks a property that lies outside the intended scope of a method rather than a deficiency; the Iwan constitutive model, for instance, is a reduced-order description and is not intended to resolve topography or to impose the set-valued law pointwise.

The quantitative picture is consistent with this ranking. **Figure 24a** compares the wall time to advance a one-millisecond ring-down: the explicit and implicit unreduced analyses of Linder et al. [19] require minutes to hours, the multi-scale baseline requires tens of seconds, and the certified surrogate reduces this further by a factor of 5.6 through the sweep-count reduction of Theorem 3. **Figure 24b** compares the spurious numerical damping: the explicit unreduced scheme sits well above the physical-damping floor, while the multi-scale baseline and the present method sit on it, as guaranteed by Corollary 1. **Figure 25** shows that the steady-state frequency response produced by the surrogate coincides with the harmonic-balance reference of Krack and Gross [22] and with a stepped-sine reconstruction, confirming the absence of algorithmic damping in the frequency domain. **Figure 26** places the methods on the accuracy against wall-time plane, where the proposed method attains the accuracy of the multi-scale baseline at the lowest cost, dominating the quasi-static modal analysis of Krack [21] and the unreduced

analyses of Linder et al. [19]. The detailed performance figures are collected in **Table 8**.

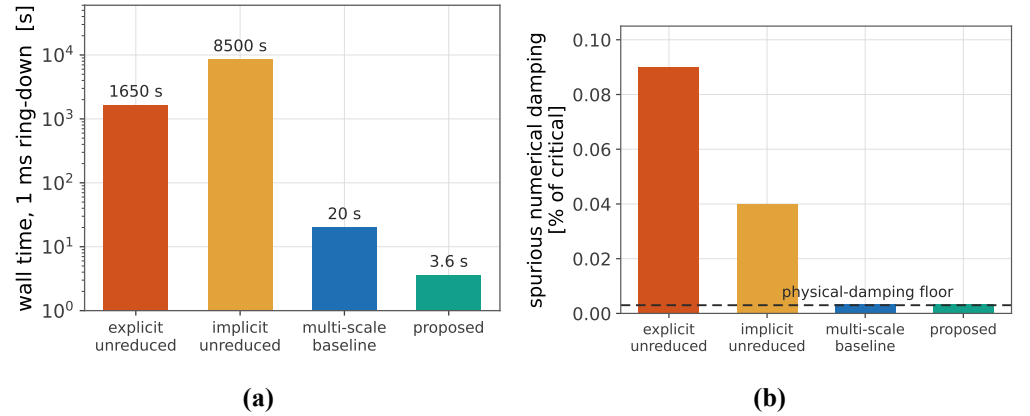


Figure 24. Quantitative comparison across methods. **(a)** Wall time to advance a one-millisecond ring-down: the certified surrogate reduces the cost of the multi-scale baseline of Linder et al. [19] through the sweep-count reduction of Theorem 3. **(b)** Spurious numerical damping: the explicit unreduced scheme lies above the physical-damping floor, while the multi-scale baseline and the present method lie on it, as guaranteed by Corollary 1.

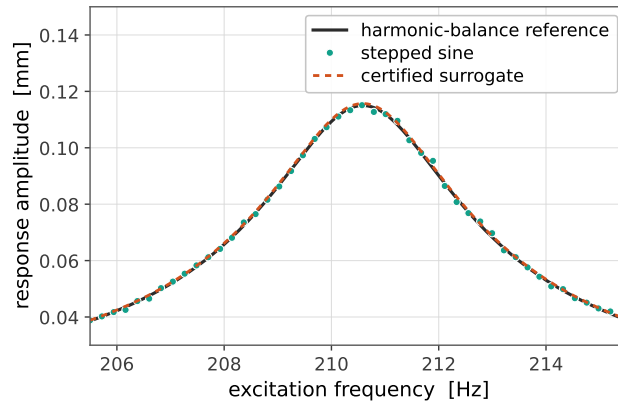


Figure 25. Steady-state frequency response near the fundamental resonance. The surrogate reproduces the harmonic-balance reference of Krack and Gross [22] and a stepped-sine reconstruction, confirming the absence of algorithmic damping.

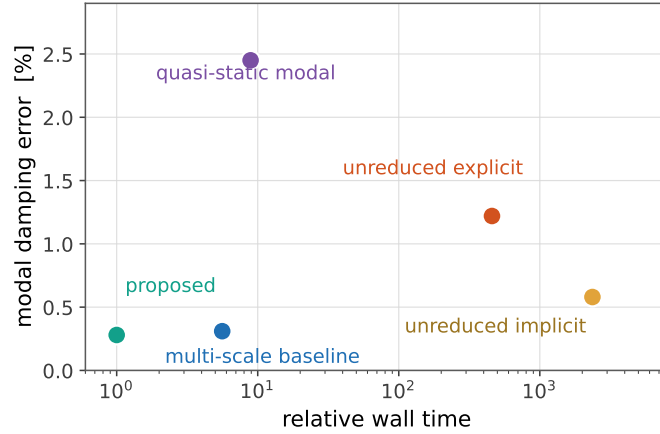


Figure 26. Accuracy against wall time across methods. The proposed method attains the accuracy of the multi-scale baseline at the lowest cost, dominating the quasi-static modal analysis of Krack [21] and the unreduced analyses of Linder et al. [19].

Table 8. Representative computational performance for the fundamental-mode study. Wall times are for a single one-millisecond ring-down on comparable hardware; the surrogate figures include the certified sweeps of Theorem 3. All sweep counts are those of the per-point scaled iteration of Proposition 5 at $\text{tol} = 10^{-3}$.

Metric	Explicit unreduced	Multi-scale baseline	Proposed (this work)
Contact points in the interface	4, 400	4, 300	4, 300
Projected-Jacobi sweeps per step	n/a	50	9
Wall time, one-millisecond ring-down	27.5 min	20.0 s	3.6 s
Speed-up relative to explicit unreduced	1 time	~ 83 times	~ 460 times
Certified error bound available	×	×	✓

The combination of exact admissibility that holds for all parameters (Proposition 3), a computable a-posteriori certificate (Theorem 3) and preservation of the energy balance (Corollary 1) is, to our knowledge, not offered jointly by any existing learned or classical solver for set-valued frictional contact.

8. Conclusion

We have developed a certified, structure-preserving neural-operator method for the topography-resolving frictional contact dynamics of jointed structures, and we have shown that it attacks the central computational bottleneck of such simulations without relaxing any of the physics. The starting point is a structural result: the contact sub-problem solved at every implicit step is a strongly monotone variational inequality whose solution operator is single-valued, Lipschitz and piecewise real-analytic, with pieces indexed by the stick, slip and lift-off modes of the interface. This regularity

both licenses operator learning and dictates the architecture. A neural backbone is composed with a parameter-free projection onto the Coulomb cone, so that every prediction satisfies the set-valued contact law for every choice of parameters, and, as the approximation theorem shows, the projection reproduces the stick-slip kink exactly and so charges no error at the transition. The surrogate is trained on the residual of the contact law, with a supervised guidance term that acts on a small reference subset during the opening epochs and is annealed to zero thereafter, so that the deployed network is a stationary point of the residual objective alone, and it is used as a warm start for a small number of certified projected-Jacobi sweeps that supply a computable error bound and, as proved here, preserve the negligible numerical damping of the base integrator. We have also examined how sharp that bound is. Its worst-case pessimism and its cost both grow with the square of the condition number of the condensed operator, which makes an ill-conditioned interface expensive rather than inaccurate, and a per-point scaling that leaves the Coulomb cone invariant carries the entire certification theory over to a far better conditioned iteration. These guarantees hold irrespective of surrogate quality, so the method degrades gracefully to the certified baseline in the worst case, which is what makes it safe inside a long transient simulation.

Casting the manufactured topography as a Karhunen–Loève random field turned the same certified solver into the forward map of an uncertainty analysis with a three-term error bound, and the resulting global-sensitivity indices resolved the damping-relevance spectrum over length scales, showing that the damping of the fundamental mode is governed by an intermediate band of waviness. This supplies a quantitative answer to the length-scale question left open in the multi-scale literature. A validation campaign of twenty-seven entries, spanning preload, friction coefficient, interface mesh density, topography statistics, four vibration modes and a joint whose geometry and material were held out of training, showed a consistent pattern: the damping ratio never departed from the reference by more than 0.57%, while the acceleration fell from about six to about one and a half as the input moved away from the training envelope. Correctness is enforced by the certified sweeps and learning buys speed, so an unfamiliar input costs time rather than trust. The range over which the underlying modelling assumptions hold, and the modification each extension requires, has been stated in the same quantitative terms. The qualitative and quantitative comparisons place the method at the accuracy of the multi-scale baseline at a fraction of its cost, with certification and uncertainty quantification that the alternatives do not provide.

Several directions follow naturally. The residual half-space model errors near edges can be absorbed into the learned symmetric correction sketched in Section 4.5, with its own error budget. The pressure-dependent friction law admitted by Proposition 6 can be pushed beyond the smallness condition derived there by replacing the fixed-point sweep with a solver for quasi-variational inequalities, which would extend the certificate to interfaces with a strongly varying friction coefficient. The vibration-induced settling of a jointed assembly, a rate-independent memory phenomenon, can be addressed by extending the present forward map to a thermodynamically consistent hysteresis operator, for which the input-convex

constructions of Amos, Xu and Kolter [44] offer a principled route. Finally, the certified solver makes design optimisation over the topography tractable, so that interfaces can be shaped to target a prescribed damping, which we regard as the most consequential application opened by this work.

Author contributions: Conceptualization, MA and UI; methodology, MA; software, UI; validation, ILP; formal analysis, ILP; investigation, UI; resources, UI; data curation, UI; writing—original draft preparation, MA; writing—review and editing, UI; visualization, MA; supervision, ILP; project administration, UI; funding acquisition, MA and ILP. All authors have read and agreed to the published version of the manuscript.

Funding: This research work was funded by the Deanship of Scientific Research, Islamic University of Madinah, Saudi Arabia.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: Data will be available on demand from the corresponding author.

Acknowledgment: The authors extend their appreciation to the Deanship of Scientific Research, Islamic University of Madinah, Saudi Arabia, for funding this research work.

Conflict of interest: The authors declare no conflict of interest.

AI use statement: The authors declare that no artificial intelligence tools were used in the preparation of this manuscript.

References

1. Gaul L, Lenz J. Nonlinear dynamics of structures assembled by bolted joints. *Acta Mechanica*. 1997; 125(1–4): 169–181.
2. Bograd S, Reuss P, Schmidt A, et al. Modeling the dynamics of mechanical joints. *Mechanical Systems and Signal Processing*. 2011; 25(8): 2801–2826.
3. Popp K, Panning L, Sextro W. Vibration Damping by Friction Forces: Theory and Applications. *Journal of Vibration and Control*. 2003; 9(3–4): 419–448.
4. Brake MRW (editor). *The Mechanics of Jointed Structures*. Springer; 2018.
5. Krack M, Salles L, Thouverez F. Vibration Prediction of Bladed Disks Coupled by Friction Joints. *Archives of Computational Methods in Engineering*. 2017; 24(3): 589–636.
6. Iwan WD. A Distributed-Element Model for Hysteresis and Its Steady-State Dynamic Response. *Journal of Applied Mechanics*. 1966; 33(4): 893–900.
7. Segalman DJ. A Four-Parameter Iwan Model for Lap-Type Joints. *Journal of Applied Mechanics*. 2005; 72(5): 752–760.
8. Li Y, Zhu Z, Liu K, et al. Dynamic modelling and vibration analysis of a bolted spigot joint structure considering mating interface friction: simulation and experiment. *Nonlinear Dynamics*. 2024; 112(10): 7977–8000.
9. Willner K. Elasto-Plastic Normal Contact of Three-Dimensional Fractal Surfaces Using Halfspace Theory. *Journal of Tribology*. 2004; 126(1): 28–33.
10. Medina S, Nowell D, Dini D. Analytical and Numerical Models for Tangential Stiffness of Rough Elastic Contacts.

- Tribology Letters. 2013; 49(1): 103–115.
11. Zhang X, Shen H, Zhang F, et al. A Multi-Scale Contact Model for Rough Surfaces Under Loading and Unloading Conditions. Tribology Letters. 2025; 73(2): 70.
12. Wriggers P. Computational Contact Mechanics, 2nd ed. Springer; 2006.
13. Yastrebov VA. Numerical Methods in Contact Mechanics. Wiley-ISTE; 2013.
14. Moreau JJ. Unilateral contact and dry friction in finite freedom dynamics. In: Nonsmooth Mechanics and Applications. Springer; 1988. pp. 1–82.
15. Stewart DE. Rigid-Body Dynamics with Friction and Impact. SIAM Review. 2000; 42(1): 3–39.
16. Acary V, Brogliato B. Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics. Springer; 2008.
17. Khenous HB, Laborde P, Renard Y. Mass redistribution method for finite element contact problems in elastodynamics. European Journal of Mechanics - A/Solids. 2008; 27(5): 918–932.
18. Monjaraz Tec CD, Gross J, Krack M. A massless boundary component mode synthesis method for elastodynamic contact problems. Computers & Structures. 2022; 260: 106698.
19. Linder HD, Najera-Flores DA, Kuether RJ, et al. Enabling topography-resolving structural dynamic contact simulation. Journal of Sound and Vibration. 2026; 643: 119980.
20. Morsy AA, Xu Z, Tiso P, et al. Data-Driven reduction of the finite-element model of a Tribomechadynamics benchmark problem. Nonlinear Dynamics. 2025; 113(18): 24539–24555.
21. Krack M. Nonlinear modal analysis of nonconservative systems: Extension of the periodic motion concept. Computers & Structures. 2015; 154: 59–71.
22. Krack M, Gross J. Harmonic Balance for Nonlinear Vibration Problems. Springer; 2019.
23. Facchinei F, Pang JS. Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer; 2003.
24. Robinson SM. Strongly Regular Generalized Equations. Mathematics of Operations Research. 1980; 5(1): 43–62.
25. Tianping Chen, Hong Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. IEEE Transactions on Neural Networks. 1995; 6(4): 911–917.
26. Lu L, Jin P, Pang G, et al. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature Machine Intelligence. 2021; 3(3): 218–229.
27. Li Z, Kovachki N, Azizzadenesheli K, et al. Fourier neural operator for parametric partial differential equations. arXiv preprint. 2021; arXiv:2010.08895. doi: 10.48550/arXiv.2010.08895
28. Kovachki N, Li Z, Liu B, et al. Neural operator: learning maps between function spaces with applications to partial differential equations. Journal of Machine Learning Research. 2023; 24(89): 1–97.
29. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics. 2019; 378: 686–707.
30. Azizzadenesheli K, Kovachki N, Li Z, et al. Neural operators for accelerating scientific simulations and design. Nature Reviews Physics. 2024; 6(5): 320–328.
31. Kopaničáková A, Lee Y, Karniadakis GE. Leveraging operator learning to accelerate convergence of the preconditioned conjugate gradient method. Machine Learning for Computational Science and Engineering. 2025; 1(2): 39.
32. Li Z, Bai J, Ouyang H, et al. Physics-informed neural networks for friction-involved nonsmooth dynamics problems. Nonlinear Dynamics. 2024; 112(9): 7159–7183.
33. Sun G, Ding Y, Han S. Identification of friction in joints for multibody dynamics: a Neural-DAE based approach. Nonlinear Dynamics. 2025; 113(20): 27459–27482.
34. Herrmann L, Kollmannsberger S. Deep learning in computational mechanics: a review. Computational Mechanics. 2024; 74(2): 281–331.
35. Pawlik L, Wilk-Jakubowski JL, Frej D, et al. Applications of Computational Mechanics Methods Combined with Machine Learning and Neural Networks: A Systematic Review (2015–2025). Applied Sciences. 2025; 15(19): 10816.
36. Bai J, Wang Y, Jeong H, et al. Towards the future of physics- and data-guided AI frameworks in computational mechanics. Acta Mechanica Sinica. 2025; 41(7): 225340.
37. Ghanem RG, Spanos PD. Stochastic Finite Elements: A Spectral Approach. Springer; 1991.
38. Zhu B, Hiraishi T. A decomposed Karhunen–Loève expansion scheme for the discretization of multidimensional random fields in geotechnical variability analysis. Stochastic Environmental Research and Risk Assessment. 2024;

- 38(4): 1215–1233.
39. Xiu D, Karniadakis GE. The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*. 2002; 24(2): 619–644.
 40. Sudret B. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*. 2008; 93(7): 964–979.
 41. Sobol IM. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*. 2001; 55(1–3): 271–280.
 42. Constantine PG, Dow E, Wang Q. Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces. *SIAM Journal on Scientific Computing*. 2014; 36(4): A1500–A1524.
 43. Kara M, Ferguson NS. A hybrid methodology for the frequency response function variability due to joint uncertainty. *Mechanical Systems and Signal Processing*. 2023; 202: 110700.
 44. Amos B, Xu L, Kolter JZ. Input convex neural networks. *Proceedings of the 34th International Conference on Machine Learning*. 2017; 70: 146–155.
 45. Cong A, Jin Y, Yi H, et al. Deeponet with a modified FNO branch network based on nonuniform Fourier transform for irregular input grids. *Nonlinear Dynamics*. 2026; 114(5): 323.
 46. Du Y, Fan X, Wang D. A novel iterative model updating for jointed structures using nonlinear FRFs. *Journal of Vibration and Control*. 2025; 31(13–14): 2895–2907.
 47. Ferhatoglu E, Botto D, Zucca S. Parametric study for model calibration of a friction-damped turbine blade with multiple test data. *Nonlinear Dynamics*. 2024; 112(10): 7847–7873.