Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

Inverse Parameter Identification of Subsurface Residual Stress in Tractional Sliding Processes Using a Physics-Informed Neural Network

Md Mehedi Hasan
Institute for Sustainable Manufacturing,
Department of Mechanical Engineering,
University of Kentucky, Lexington, KY
40506, USA
Email: mdmehedi.hasan@uky.edu

Julius Schoop[1]
Institute for Sustainable Manufacturing,
Department of Mechanical Engineering,
University of Kentucky, Lexington, KY
40506, USA
Email: julius.schoop@uky.edu

Residual stresses (RS) arise in a wide range of manufacturing processes, including additive manufacturing, welding, forming, grinding, and machining. Accurate characterization and prediction of RS are crucial for optimizing functional performance and structural integrity, as tensile stresses reduce fatigue strength while compressive stresses enhance it. Traditional finite element methods provide detailed insights into RS distributions but are computationally expensive for real-time use. To overcome this limitation, we propose a Physics-Informed Neural Network (PINN) framework that embeds the Prandtl–Reuss constitutive equations for elastoplasticity directly into the loss function, enabling mesh-free forward simulation of RS distribution and inverse identification of parameters under Hertzian contact loading. The inverse formulation simultaneously reconstructs stress fields and identifies key parameters—the effective friction coefficient and normalized load factor—from sparse data, addressing the nonuniqueness and instability of traditional inverse methods. Validation against high-fidelity Runge–Kutta–Gill reference solutions shows that residual stress prediction errors remain below 8% across a wide parameter range, while parameter identification errors converge to below 1%. The PINN predictions were compared with representative experimental trends for Ti–6Al–4V under burnishing and orthogonal cutting, confirming consistency across chip-generating and chipless processes. By enabling real-time parameter updates from minimal data, the proposed framework can accelerate the development of digital twins for manufacturing, supporting predictive modeling and process optimization. This advancement provides physics-based rapid RS analysis for critical applications, including bearing contacts and machining process optimization, significantly improving speed and usability over traditional approaches.

Keywords: Physics-Informed Neural Networks, Residual Stress, Contact Mechanics

[1]Corresponding Author.
Version 1.18, December 11, 2025

1 Introduction

Residual stresses (RS) are the result of internal material loads imposed during manufacturing, specifically mechanical, thermal, and / or chemical [1]. According to the Process Signature framework introduced by Brinksmeier et al. [2], each manufacturing process is characterized by one or more of these three loading regimes, with most of the processes thermo-mechanical in nature and some dominated by a single regime. For example, high-speed grinding can be considered to impart primarily thermal loads, while deep rolling imparts primarily mechanical loads on the subsurface, resulting in the corresponding characteristic RS profiles. Plastic deformation-induced stresses (that is, mechanical loading) dominate in the subsurface region, typically from a few tens to several hundred micrometers below the surface in cold rolling, finish machining, deep rolling, and peening. Depending on specific process conditions, these mechanically dominated processes plastically deform the material to depths ranging between 10–1000 μm, thereby distorting the crystal lattice and establishing a compressive stress gradient that significantly enhances surface integrity and fatigue resistance [2].

Tensile RS generally reduces fatigue strength, while compressive RS enhances it, making favorable compressive RS induction through finishing manufacturing techniques such as peening and burnishing essential [3]. Moreover, RS not only influences fatigue life but also affects other performance metrics such as dimensional stability and overall structural integrity [4]. Inaccurate RS predictions can lead to unexpected failures and increased maintenance costs, particularly in high-reliability applications. Therefore, accurate prediction and control of RS are critical to enhancing precision, performance and longevity of manufactured parts [5].

The foundation of contact mechanics dates back to Heinrich Hertz's seminal work of 1881 on elastic stress distributions in spherical contacts [6]. Notably, Hertzian contact mechanics enable prediction of mechanically induced RS, as typically observed in rolling, sliding, and indentation processes such as deep rolling (burnishing) and peening, while largely neglecting thermal effects. The normal (frictionless) indentation of the traditional Hertzian contact framework was extended by Mindlin in 1949 to account for tangential (tractional) loading, enabling the analysis of combined normal and shear tractions, as encountered in rolling and sliding. It was not until the early 1960s that Merwin and Johnson coupled these Hertzian stress fields with the Prandtl–Reuss incremental plasticity equations, yielding one of the very first predictive models for subsurface residual stresses (RS) under rolling/sliding contacts with focus on railroad applications. Merwin and Johnson's framework [7] has been widely employed to interpret the plastic deformation induced during rolling operations. This approach is particularly valuable in problems involving RS prediction, especially where plastic deformation is dominant when compared to thermal stress fields [8]. Such conditions are typically met in burnishing and finish machining, both of which can be considered to be mechanically dominated when sliding/cutting speeds are moderate and proper cooling and lubrication are provided. For example, in our previous work, we showed that in-situ experimentally characterized model inputs such as the equivalent Hertzian contact width ($2a$), effective friction coefficients (μ), and load factor (p_0/k) [3] enable high-fidelity prediction of machining-induced RS. The adaptation of Hertzian contact theory to machining processes has advanced significantly in recent decades. For instance, Liang and Su [9] extended the framework to incorporate tool flank wear and thermal loads, employing a relaxation procedure to predict the inward shift of subsurface compressive RS in orthogonal turning of Ti-6Al-4V. Further refinements were introduced by Huang et al. [10], who modeled the tool–workpiece interface as coupled Hertzian contacts on the shear plane and flank face. Recently, Hertzian theory has been employed to model machining-induced RS by in-situ calibration of process parameters—such as the equivalent contact width and effective friction coefficient—during peripheral milling of Ti–6Al–4V [1].

However, a critical limitation persists: inverting these models to infer contact parameters (e.g., pressure, friction coefficient, hardening laws) from sparse RS data is inherently ill-posed. Noise in experimental measurements and uncertainties in material models often lead to unstable or non-unique solutions [11,12]. While traditional approaches rely on iterative forward simulations or simplified assumptions, recent data-driven methods lack rigorous enforcement of plasticity constraints [13,14]. These challenges hinder the adoption of RS-optimized machining in industry, where rapid and reliable parameter identification is essential for adaptive manufacturing. To address this need, we present a physics-informed framework that bridges Hertzian contact mechanics with data-driven modeling. By embedding the Prandtl–Reuss constitutive equations into the PINN architecture, our approach enables robust, simultaneous identification of key

parameters (e.g., effective friction coefficient, normalized load factor) and full-field RS reconstruction, even from a single reference point. In doing so, it not only mitigates the ill-posedness of inverse problems, but also provides the foundation for digital twin implementations, where real-time parameter updating supports predictive modeling and adaptive process control.

Modern approaches to predicting RS fall into three main categories: empirical, analytical, and numerical models. Empirical models rely on experimental data but suffer from limited accuracy and generalizability. Analytical models capture physical phenomena through mathematical equations but often require simplifying assumptions that restrict their applicability. Numerical methods, particularly finite element analysis (FEA), provide detailed insights into RS formation and are increasingly favored for their predictive accuracy. However, traditional finite element methods (FEM) face two critical limitations: (1) computational cost scales significantly with the number of elements or nodes, particularly in three-dimensional geometries [15]; (2) a lack of explicit gradient or sensitivity information, which is essential for inverse problems, parameter identification, and optimization in material design [16]. A central challenge in RS modeling is solving the inverse problem—estimating constitutive and contact parameters from sparse RS measurements. Conventional FE-based optimization approaches iteratively solve forward problems during parameter searches, but these methods are computationally expensive and inherently ill-posed; even minor measurement noise propagates into large errors in traction or eigenstrain reconstructions [11,12]. These limitations motivate the exploration of physics-informed neural networks (PINNs) as an alternative surrogate that retains the governing physics while enabling efficient forward evaluation and direct inversion for key contact parameters.

During the past decade, machine learning (ML) and the more advanced deep learning (DL) framework have rapidly advanced and found extensive applications in solid mechanics [17]. Early DL models often adopted data-driven approaches, neglecting the physics of mechanics. Physics-informed Neural Networks (PINNs) overcome these limitations by embedding governing equations directly into a multi-objective loss function, physically consistent solutions even with limited training data [18]. However, existing PINN-based inversion methods frequently require large datasets and enforce plasticity constraints only weakly, compromising robustness under scarce measurements [19]. To address these shortcomings, our inverse-PINN framework integrates the Prandtl–Reuss elastoplastic laws into its loss function, enabling (i) simultaneous forward prediction of subsurface RS fields, (ii)inverse identification of key contact parameters (e.g., effective friction coefficient μ, normalized load factor p_0/k) from sparse measurements, and (iii)strict enforcement of physical constraints to enhance noise robustness and reduce computational cost. Unlike traditional FEM or purely data-driven ML, PINNs offer a physics-constrained mesh-free framework that inherently satisfies elastoplasticity laws while requiring orders-of-magnitude fewer data points [18].

PINNs have been successfully applied across diverse fields—fluid mechanics [20], geosciences [21], structural mechanics [22], and industrial process control [23]. In solid mechanics, Haghighat *et al.* [24,25] and Roy & Guha [26] demonstrated PINNs for linear and nonlinear plasticity, while Sahin *et al.* [27] applied PINNs to elastic Hertzian contact. To our knowledge, no prior work has simultaneously modeled contact-driven RS evolution and performed inverse parameter recovery under full elastoplastic constraints. While existing studies highlight PINNs' versatility in mechanical modeling, key challenges persist in applying them to rolling/sliding contacts, particularly due to coupled thermo-mechanical effects and limited high-fidelity experimental data.

In this work, we focus exclusively on mechanically induced RS. We analyze small-strain Hertzian indentation and rolling/bearing contacts, which plastically deform the subsurface to depths of 50–200 μm, and we show how PINNs for tractional sliding Hertz contact problems can be extended to capture the evolution of RS in these critical surface layers during finish manufacturing processes. Our framework addresses these challenges through a physics-embedded architecture that simultaneously enables forward RS field prediction and inverse parameter identification. It uniquely incorporates complete elastoplastic physics by formulating a multi-objective loss function that integrates the governing differential equations, Prandtl–Reuss constitutive laws, plastic flow criteria, consistency constraints, and boundary conditions. To our knowledge, this represents the first application of PINNs specifically to residual stress modeling, filling a critical gap in computational mechanics where prior implementations have not addressed contact-driven RS prediction. Validation across $\mu \in [0, 0.75]$ and $p_0/k \in [4, 6]$ demonstrates mean relative L_2 errors below 8% in stress reconstruction and absolute errors below 1% in parameter identification using only single reference point. Furthermore, the framework achieves mesh-free computational efficiency,

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

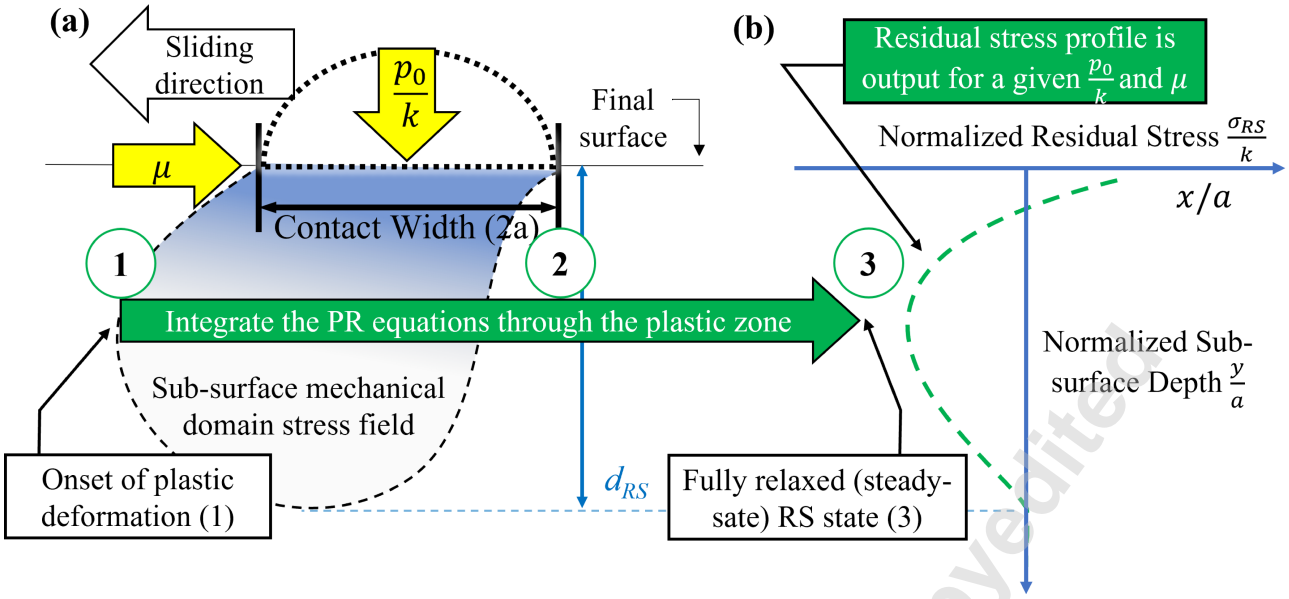


Fig. 1 Schematic illustration of generating residual stress using Hertzian contact mechanics. The contact width 2*a* is subjected to a normal load *P* and a tangential load *Q*, resulting in a sub-surface stress field. The depth d_{RS} indicates the maximum depth which the final residual stress can generate, evolving along the sliding direction.

avoiding numerical instabilities of classical Hertzian solutions while significantly reducing cost compared to finite element or brute-force empirical methods. By enabling reliable parameter identification and full-field stress reconstruction from minimal data, this work also lays the groundwork for digital twin implementations of machining and surface finishing processes, where real-time model updating can accelerate predictive analytics, process optimization, and adaptive control.

The structure of the paper is organized as follows. Section 2 outlines the model for computing RS, including the key governing equations and the semi-analytical approach for elastic stress determination. In Section 3, we introduce our physics-informed neural network (PINN) framework for parameter identification, including the formulation of loss functions, incorporation of boundary and initial conditions, and the training procedures for calibrating key model parameters. Section 4 presents extensive numerical experiments to validate the proposed approach. We demonstrate the accuracy of our forward problem solutions by comparing the PINN predictions with results from traditional numerical solvers, and we assess the robustness of the inverse solver in identifying model parameters from synthetic numerical data. Finally, Section 5 presents the experimental validation of the proposed framework and concludes the paper with remarks on future research directions.

2 Model Description

Mechanically induced residual stresses —commonly generated in processes such as indentation, rolling, finish machining, or contact loading— arise from the complex interaction of elastic and plastic deformation during surface and subsurface material flow. These stresses span multiple physical domains, including elastic, plastic, thermal, and thermodynamic regimes, as illustrated in Figure 1. In this study, we seek to demonstrate the relevance and applicability of a PINN approach for RS prediction and inverse parameter identification. For this reason, our analysis is confined to the mechanical domain, deliberately excluding near-surface thermal stresses to isolate the elastoplastic mechanisms of interest.

The modeling framework comprises three sequential steps:

(1) **Elastic-Plastic Transition:** Estimation of the initial elastic stress field and identification of elastic–plastic transition zones resulting from the combined effects of tool–workpiece contact and shear zone deformation.

(2) **Determine Deviatoric Stress in Plastic Region :** Determination of the deviatoric stresses within the workpiece by solving the

Prandtl-Reuss (PR) equations governing plastic deformation. The deviatoric stress is calculated using the Runge-Kutta-Gill (RKG) method as a reference solution. This PINN-based integration efficiently captures complex strain gradients without requiring explicit analytical expressions.

(3) **Stress Relaxation:** Application of a stress-relaxation scheme to enforce traction-free boundary conditions and compatibility constraints, yielding the final residual stress distribution on the machined surface.

In the present model, the workpiece is assumed to be a semi-infinite, isotropic, and homogeneous material exhibiting elastoplastic behavior with isotropic hardening. The material response is considered rate-independent and is governed by the von Mises yield criterion. Mechanical properties such as the shear modulus (G) and Poisson's ratio (ν) are treated as known constants throughout the domain. This assumption provides a tractable yet representative basis for modeling residual stresses, although the framework can be extended to accommodate more complex, rate-dependent or anisotropic material behaviors if needed.

The following section focuses on evaluating the internal stress fields generated by mechanical loading. During this phase, the interaction between a rigid indenter and the half-space is modeled as planar contact. Accurately capturing these stress distributions is essential, as they serve as the foundation for predicting the residual stresses induced by the machining process.

2.1 Mechanical Loading. There are two main sources of mechanical stress during orthogonal metal cutting. The first arises from the contact between the tool edge and the workpiece, creating both normal and tangential loads. The second source comes from the stresses within the shear zone, which contribute additional inclined shear and normal stresses. In our model, the overall stress state is obtained by superimposing these effects.

Although the above formulation is motivated by orthogonal cutting mechanics, the modeling framework is general and applicable to a broader class of problems involving contact-induced residual stresses. This includes rolling, sliding, indentation, and additive manufacturing processes—anywhere localized plastic deformation under contact loading plays a critical role in subsurface stress evolution. By abstracting the loading and boundary conditions, the same computational pipeline can be readily adapted to these applications.

The stresses induced by the cutting forces are determined by considering contributions from both the shear plane and the tool flank–workpiece contact zone. The mechanical stress model induced by the shear plane is presented in Ref. [28], while the stress generated in the tool–workpiece contact zone follows a similar approach to modeling tool-side wear [29]. Assuming a plane strain state in the y-direction ($\epsilon_{zz} = 0$), stresses under the normal compressive stress distribution (feed direction), $p(x)$, and tangential stress distribution (cutting direction), $q(x)$, can be calculated using the following equations [30]:

$$\sigma_{xx} = -\frac{2y}{\pi}\int_{-a}^{a}\frac{p(s)(x-s)^2}{\left[(x-s)^2+y^2\right]^2}\,ds - \frac{2}{\pi}\int_{-a}^{a}\frac{q(s)(x-s)^3}{\left[(x-s)^2+y^2\right]^2}\,ds, \tag{1}$$

$$\sigma_{yy} = -\frac{2y^3}{\pi}\int_{-a}^{a}\frac{p(s)}{\left[(x-s)^2+y^2\right]^2}\,ds - \frac{2y^2}{\pi}\int_{-a}^{a}\frac{q(s)(x-s)}{\left[(x-s)^2+y^2\right]^2}\,ds, \tag{2}$$

$$\tau_{xy} = -\frac{2y^2}{\pi}\int_{-a}^{a}\frac{p(s)(x-s)}{\left[(x-s)^2+y^2\right]^2}\,ds - \frac{2y}{\pi}\int_{-a}^{a}\frac{q(s)(x-s)^2}{\left[(x-s)^2+y^2\right]^2}\,ds. \tag{3}$$

To employ these equations, the stress distributions $p(s)$ and $q(s)$ are defined based on the contact geometry and loading conditions. The normal pressure is assumed to follow a two-dimensional Hertzian distribution, such that for rolling/sliding contact between a cylinder and a semi-infinite half-space, the normal and shear load distributions are given by

$$p = \frac{2P}{\pi a^2}\sqrt{a^2 - x^2}, \tag{4}$$

$$q = \frac{2Q}{\pi a^2}\sqrt{a^2 - x^2}, \tag{5}$$

with $-a \le x \le a$, where the maximum compressive pressure is $p_0 = \frac{2P}{\pi a}$. Owing to the coupling of rolling and sliding contact, the shear load is related to the normal load via the effective friction coefficient μ:

$$Q = \mu P. \tag{6}$$

Analytical solutions for the elastic stresses induced by the normal load (and similarly for the shear load) are available in the literature [31]. The elastic stresses due to the normal load P are given as follows:

$$(\sigma^e_{xx})_P = -\frac{2P}{\pi a^2}\left(m\left(1 + \frac{y^2 + n^2}{m^2 + n^2}\right) - 2y\right), \tag{7}$$

$$\left(\sigma^e_{yy}\right)_P = -\frac{2P}{\pi a^2}\, m\left(1 - \frac{y^2 + n^2}{m^2 + n^2}\right), \tag{8}$$

$$\left(\tau^e_{xy}\right)_P = \frac{2P}{\pi a^2}\, n\left(\frac{m^2 - y^2}{m^2 + n^2}\right), \tag{9}$$

The elastic stresses due to the tangential shear force Q are given as follows:

$$(\sigma^e_{xx})_Q = \frac{2Q}{\pi a^2}\left(n\left(2 - \frac{y^2 - m^2}{m^2 + n^2}\right) - 2x\right), \tag{10}$$

$$\left(\sigma^e_{yy}\right)_Q = \frac{Q}{P}\,\sigma^e_{yy}, \tag{11}$$

$$\left(\tau^e_{xy}\right)_Q = \frac{Q}{P}\,\tau^e_{xy}. \tag{12}$$

Here, the parameters m and n are defined as [31]:

$$m^2 = \frac{1}{2}\left(\sqrt{(a^2 - x^2 + y^2)^2 + 4x^2y^2} + (a^2 - x^2 + y^2)\right), \tag{13}$$

$$n^2 = \frac{1}{2}\left(\sqrt{(a^2 - x^2 + y^2)^2 + 4x^2y^2} - (a^2 - x^2 + y^2)\right). \tag{14}$$

where a is the half-contact length, and these equations are experimentally calibrated. The overall elastic stress at any point (x, y) is then the sum:

$$\sigma^e_{ij} = \left(\sigma^e_{ij}\right)_P + \left(\sigma^e_{ij}\right)_Q. \tag{15}$$

The preceding analysis establishes the elastic stress field arising from mechanical loading during the cutting process. However, when

the applied stresses exceed the material's elastic limit, localized plastic deformation ensues. This inelastic behavior leads to irreversible changes in the stress state, which persist even after the external load is removed. It is this plastic deformation, coupled with the subsequent unloading response, that ultimately gives rise to residual stresses. The following section introduces our modeling approach for capturing these effects by tracking the full stress evolution throughout the deformation history.

2.2 Residual Stress Modeling. Residual stresses can be theoretically predicted if the history of internal stress evolution within the material is known. In the context of rolling or sliding contact, once the key loading parameters—namely, the contact length, applied normal load, and friction coefficient—are defined, the stress field at any location within the workpiece can be determined. During loading, each material point may exist in either an elastic or a plastic state, depending on whether the local stress exceeds the yield limit. Following McDowell's framework [32], the process involves three key phases: (1) elastic–plastic loading, (2) unloading, and (3) residual stress relaxation, building upon the classical formulation of Merwin and Johnson [7].

In the analysis that follows, the stress evolution is tracked at a prescribed, nondimensional subsurface depth $y = y_n > 0$, where the material point experiences the passage of the rolling or sliding contact. Under the plane-strain assumption, the relevant stress components governing the loading–unloading response are the in-plane normal stress in the sliding direction σ_{xx}, the in-plane shear stress τ_{xy}, and the out-of-plane normal stress σ_{zz}. These quantities form the basis for the Prandtl–Reuss integration performed in the subsequent subsections.

2.2.1 Plastic Loading. Before initiating the plastic loading analysis, the elastic–plastic transition point must be identified for each subsurface depth. The plastic boundary is determined using the von Mises yield criterion, expressed through the second invariant of the deviatoric stress tensor J_2. The yield condition follows the J_2 flow theory without kinematic hardening:

$$F = \frac{1}{2} s_{ij} s_{ij} - k^2 = 0, \tag{16}$$

where k is the yield stress in simple shear, s_{ij} are the components of the deviatoric stresses. The deviatoric stress is expressed as:

$$s_{ij} = \left(\sigma_{ij} - \frac{1}{3} \delta_{ij} \sigma_{kk} \right), \tag{17}$$

where δ_{ij} is the Kronecker delta, and σ_{kk} is the trace of the stress tensor. This intersection point, refined using a simple bracketing and interpolation procedure, defines the elastic–plastic boundary x_{ep} from which the Prandtl–Reuss equations are subsequently integrated into the plastic region. Once the onset of yielding has been established, the material response within the plastic region is governed by the additive decomposition of strain into elastic and plastic parts:

$$\dot{\varepsilon}_{ij} = \dot{\varepsilon}^e_{ij} + \dot{\varepsilon}^p_{ij}, \tag{18}$$

$$\dot{\varepsilon}^p_{ij} = \frac{\dot{W}}{2k^2} s_{ij}, \qquad \dot{W} \equiv s_{kl} \dot{\varepsilon}^p_{kl} \ (\geq 0). \tag{19}$$

where $\dot{W}$ is the plastic work rate. It represents the rate at which stresses perform work during a change in shape. Under plastic deformation, the von Mises flow condition ensures that the elastic strain energy of distortion remains constant, so $\dot{W}$ effectively quantifies the rate of irreversible energy dissipation through plastic action. Here, $\dot{\varepsilon}_{ij}$ denotes the strain rate tensor. Deviatoric stress evolves via the Prandtl–Reuss (PR) equation:

$$\dot{s}_{ij} = 2G \left(\dot{\varepsilon}_{ij} - \tfrac{1}{3} \dot{\varepsilon}_{kk} \delta_{ij} - \frac{\dot{W}}{2k^2} s_{ij} \right), \tag{20}$$

with $G = E/[2(1+\nu)]$. For unloading ($\dot{W} < 0$), plastic flow ceases and $\dot{s}_{ij} = 2G\,\dot{\varepsilon}_{ij}$.

Under plane strain ($\varepsilon_{zz} = 0$), derivatives along the loading direction x replace time rates, and $\dot{\sigma}_{zz}$ follows

$$\dot{\sigma}_{zz} = \nu\big(\dot{\sigma}_{xx} + \dot{\sigma}_{yy}\big) - 2G(1+\nu)\,\dot{\varepsilon}^{p}_{zz}. \tag{21}$$

In our rolling/sliding contact problem, we adopt a plane strain assumption, thereby reducing the analysis to two dimensions. Under this assumption, the strain in the z-direction is negligible ($\epsilon_{zz} = 0$), and the solution remains invariant along that axis. Consequently, the x-direction represents the primary direction of variation (corresponding to the direction of tool motion or surface propagation), while the y-direction is held constant. This simplification permits the transformation of time derivatives into spatial derivatives along the x-direction. As a result, the integration of the PR equations is performed along the x-direction with the y-coordinate fixed.

When the semi-infinite solid is subjected to a load approaching from infinity, the initial stresses can be calculated using the elastic equations. However, as the load increases and the material reaches beyond its yield point, the PR incremental relations need to be applied. In the plastic region, it's not possible to compute the stresses at individual points independently; instead, the solution must be extended inward from the elastic-plastic boundary, where the stresses are already determined. The total incremental strain rates combine elastic and plastic components can be calculated using the equation below [32]:

$$\dot{\varepsilon}_{ij} = \frac{1}{2G}\,\dot{\sigma}_{ij} - \frac{\nu}{2G(1+\nu)}\,\dot{\sigma}_{kk}\,\delta_{ij} + \frac{\dot{W}}{2k^2}\,s_{ij}, \tag{22}$$

In this analysis, the PR equations are integrated along the x-direction, with the y-coordinate held constant. As the integration proceeds, a point is eventually reached where $\dot{W} \le 0$, indicating that no additional plastic work is being accumulated. This condition marks the end of the plastic deformation regime and corresponds to the onset of elastic unloading. At this unloading point, the material re-enters the elastic domain even though the analytical stress field continues to decay as the contact moves away. To ensure a smooth and physically consistent transition between the plastic and elastic responses, the integration of the PR equations is continued rather than reverting immediately to the purely elastic formulation. By carrying the integration through the unloading phase and extending it asymptotically toward $x \to \infty$, the stress state naturally converges to the final residual stress distribution at that depth.

2.2.2 Residual Stress Relaxation. After unloading, the stress and strain fields obtained from the Prandtl–Reuss integration generally do not satisfy the traction-free boundary conditions at the free surface. To reconcile the post-unloading fields with the surface boundary conditions, a relaxation procedure—following the classical method of Merwin and Johnson [7] is applied.

In this relaxation step, the residual fields are adjusted so that the traction-free conditions are recovered at the surface; hence, the relaxed components satisfy

$$\sigma^{r}_{y} = 0, \qquad \tau^{r}_{xy} = 0.$$

Additionally, a zero in-plane residual strain constraint, $\epsilon^{r}_{x} = 0$, is imposed to eliminate unphysical rigid-body displacements that would otherwise arise during unloading.

Under these constraints, only the normal residual-stress components σ^{XX}_{RS} and σ^{ZZ}_{RS} remain nonzero, with the corresponding residual strains given by ϵ^{r}_{y} and γ^{r}_{xy}. The adjusted residual stresses follow directly from the relaxation relations:

$$\sigma^{XX}_{RS} = \sigma^{r}_{x} - \frac{\nu\,\sigma^{r}_{y}}{1-\nu}, \tag{23}$$

$$\sigma^{ZZ}_{RS} = \sigma^{r}_{z} - \frac{\nu\,\sigma^{r}_{y}}{1-\nu}. \tag{24}$$

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

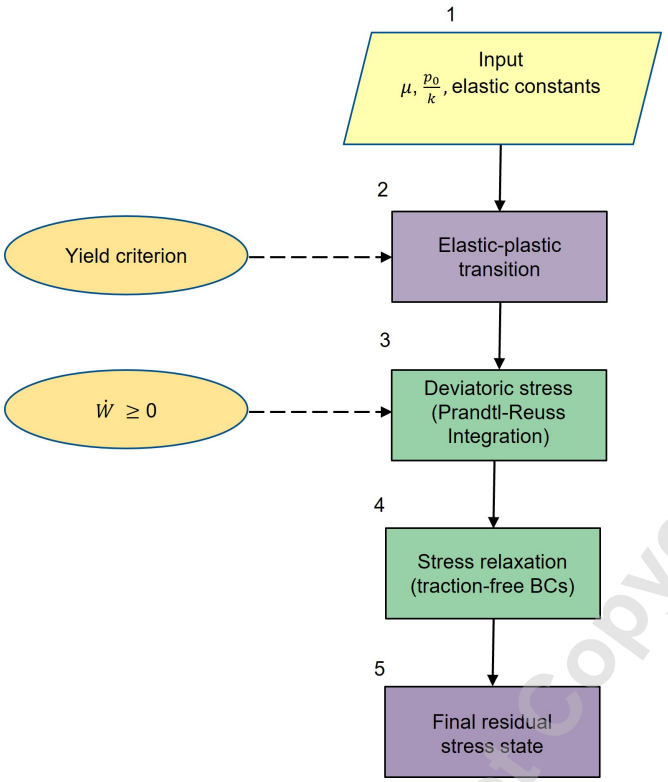


Fig. 2 Workflow for residual–stress computation under Hertzian contact: (1) inputs; (2) elastic–plastic transition; (3) deviatoric stress integration via Prandtl–Reuss; (4) stress relaxation to enforce traction-free boundary conditions; (5) final residual–stress state. Dashed links denote constraints (yield, plastic-work non-negativity).

In summary, the overall process for computing contact-induced residual stresses consists of three stages: (i) evaluating the initial elastic fields and identifying the elastic–plastic transition, (ii) integrating the Prandtl–Reuss equations through loading and unloading, and (iii) applying the relaxation procedure to restore the traction-free boundary conditions and obtain the final residual-stress state. These steps are illustrated in Figure 2, and the resulting framework is broadly applicable to rolling, sliding, machining, and other contact-driven problems.

In Section 3, rather than performing the usual incremental integration of (20), we employ a Physics-Informed Neural Network. By embedding both the Prandtl–Reuss law and constraints into its loss, this PINN directly predicts the spatial evolution of the plastic stress field.

3 PINN Framework for Parameter Identification

This section introduces a physics-informed neural network (PINN) framework that augments the classical semi-analytical Hertzian contact model with data-driven learning. By embedding the Prandtl–Reuss (PR) elastoplastic equations and contact-mechanics constraints directly into the loss function, the framework predicts the subsurface evolution of residual stresses under rolling/sliding contact. Two complementary problems are addressed: (i) the forward problem—predicting the depth-dependent residual stress field given known loading parameters; and (ii) the inverse problem—recovering those parameters from sparse residual-stress observations.

In the forward formulation, the normalized load factor p_0/k and effective friction coefficient μ are treated as fixed inputs. The network is trained to minimize a composite loss that enforces the PR equations and boundary conditions while matching reference data, thereby producing the complete residual-stress profile versus depth. In the inverse formulation, p_0/k and μ are promoted to trainable variables within the PINN. By fitting only a limited number of measured stress points, the framework simultaneously reconstructs the subsurface stress field and identifies the underlying loading parameters.

We next present the mathematical formulation of both forward and inverse PINN approaches.

3.1 Forward Formulation: Residual Stress Prediction. In the plastic region, the deviatoric stress components $s_{ij}(x)$ are approximated over a one-dimensional spatial domain $\Omega \subset \mathbb{R}$, assuming a fixed transverse coordinate y (plane strain). A deep neural network (DNN) approximates the stress field,

$$s_{ij}^{\text{pred}}(x;\theta) \approx s_{ij}(x),$$

with trainable parameters θ. The network is trained to satisfy the Prandtl–Reuss (PR) elastoplastic relations and plasticity consistency, embedded via a residual operator

$$\mathcal{F}\Big(s_{ij}(x), \tfrac{d}{dx}s_{ij}(x)\Big) = \mathbf{0}, \qquad x \in \Omega,$$

where $\mathcal{F}$ collects the PR constitutive residuals (cf. Eqs. (20), (22), (21)) together with yield consistency $F = 0$ (Eq. (33)) and the plastic-work non-negativity condition $\dot{W} \geq 0$.

Weighted objective: We minimize a composite loss with tunable global weights λ_* and pointwise weight functions $w_*(\cdot)$:

$$\mathcal{L}_T(\theta) = \lambda_{\text{res}}\,\mathcal{L}_{\text{res}}(\theta) + \lambda_{\text{IC}}\,\mathcal{L}_{\text{IC}}(\theta) + \lambda_{\text{data}}\,\mathcal{L}_{\text{data}}(\theta), \tag{25}$$

where the (optional) data term can be disabled by setting $\lambda_{\text{data}} = 0$ (default in the forward solve).

Physics (residual) loss : Let $\{x_j\}_{j=1}^{N_r} \subset \Omega$ be interior collocation points. Using automatic differentiation (AD) for spatial derivatives,

$$\begin{aligned}\mathcal{L}_{\text{res}}(\theta) = \frac{1}{\sum_j w_{\text{res}}(x_j)} \sum_{j=1}^{N_r} w_{\text{res}}(x_j) \Big[\\ \alpha_{\text{PR}} \big\|\text{Res}_{\text{PR}}(x_j)\big\|^2 + \alpha_F\, F(x_j)^2 \\ + \alpha_W \left(\text{ReLU}[-\dot{W}(x_j)]\right)^2 \Big].\end{aligned} \tag{26}$$

where Res_{PR} are the PR residuals assembled from Eqs. (20), (22), (21); $F = \frac{1}{2}s_{ij}s_{ij} - k^2$ is the yield function; $\dot{W} = s_{kl}\dot{\varepsilon}^p_{kl}$; $\text{ReLU}(z) = \max(0, z)$; and $\alpha_{\text{PR}}, \alpha_F, \alpha_W > 0$ balance the sub-terms.

Initial-condition loss: At the elastic–plastic boundary $x = x_0$ (possibly multiple anchor points $\{x_{0,m}\}$),

$$\begin{aligned}\mathcal{L}_{\text{IC}}(\theta) = \frac{1}{\sum_m w_{\text{IC}}(x_{0,m})} \sum_{m=1}^{N_{\text{IC}}} w_{\text{IC}}(x_{0,m}) \times \\ \big\| s_{ij}^{\text{pred}}(x_{0,m};\theta) - s_{ij}^{\text{Hertz}}(x_{0,m}) \big\|^2.\end{aligned} \tag{27}$$

Data loss: If a reference value is used to gently stabilize training,

$$\mathcal{L}_{\text{data}}(\theta) = \frac{1}{\sum_k w_{\text{data}}(x_k)} \sum_{k=1}^{N_d} w_{\text{data}}(x_k) \big\| s_{ij}^{\text{pred}}(x_k;\theta) - s_{ij}^{\text{ref}}(x_k) \big\|^2, \tag{28}$$

with $\lambda_{\text{data}} \ll 1$ or $\lambda_{\text{data}} = 0$ for a strictly physics-only forward solve.

Training and post-processing: We minimize $\mathcal{L}_T(\theta)$ using a gradient-based optimizer; all spatial derivatives are obtained via AD. The trained PINN yields the deviatoric stress evolution $s_{ij}(x)$ up to the unloading point (where $\dot{W} \leq 0$), after which the relaxation procedure enforces traction-free surface conditions to reconstruct the final residual-stress field (Section 2).

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

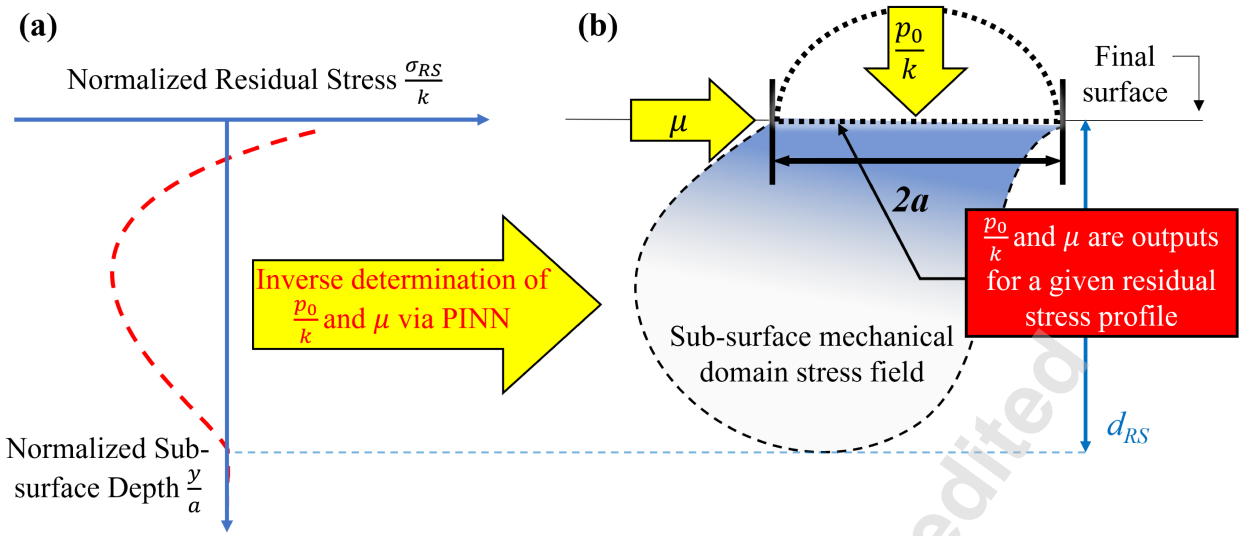


Fig. 3 Overview of Inverse PINN: Given a measured residual-stress profile, the PINN simultaneously reconstructs the subsurface stress evolution and infers the unknown loading parameters ($p_0/k, \mu$) by embedding the Prandtl–Reuss physics into its loss.

3.2 Inverse Formulation: Simultaneous Stress Reconstruction and Parameter Identification. While the forward PINN predicts the deviatoric stress field $s_{ij}(x)$ for known loading, the inverse formulation tackles the ill-posed task of recovering both the subsurface stress field and the unknown loading parameters from sparse (down to a single reference point) or noisy residual-stress observations. We focus on two parameters that dominantly govern contact-driven RS: the effective friction coefficient μ and the normalized load factor p_0/k.

As sketched in Figure 3, the inverse PINN promotes $\{\mu,\ p_0/k\}$ to trainable variables θ_0 and learns them jointly with the network weights θ. The loss couples a data-misfit term on the available observations with physics terms enforcing the Prandtl–Reuss (PR) elastoplastic relations and consistency constraints (yield $F = 0$, plastic-work non-negativity $\dot{W} \geq 0$, and plane-strain), enabling simultaneous stress-field reconstruction and parameter identification under strong physical regularization.

In the inverse formulation, the deviatoric stress field is approximated by a DNN while the unknown loading parameters are learned jointly:

$$s_{ij}^{\text{pred}}(x;\theta,\theta_0) \approx s_{ij}(x),$$

where θ are network weights and $\theta_0 = \{\mu, \rho\}$ with $\rho \equiv p_0/k$. To ensure physical bounds during training, we map unconstrained trainables $\zeta_1, \zeta_2 \in \mathbb{R}$ to admissible parameters via

$$\mu = \mu_{\max}\,\sigma(\zeta_1), \qquad \rho = \rho_{\min} + (\rho_{\max} - \rho_{\min})\,\sigma(\zeta_2),$$

with $\sigma(\cdot)$ the logistic sigmoid and user-chosen bounds (e.g., $\mu_{\max} \geq 0$, $\rho_{\min} > 0$).

Weighted composite objective: We minimize

$$\mathcal{L}_T(\theta,\theta_0) = \lambda_{\text{data}}\mathcal{L}_{\text{data}} + \lambda_{\text{res}}\mathcal{L}_{\text{res}} + \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}}, \tag{29}$$

with global weights $\lambda_* > 0$ and pointwise weights $w_*(\cdot)$.

Data misfit: Given sparse observations $\{x_k, s_{ij}^{\text{true}}(x_k)\}_{k=1}^{N_d}$,

$$\mathcal{L}_{\text{data}} = \frac{1}{\sum_k w_{\text{data}}(x_k)} \sum_{k=1}^{N_d} w_{\text{data}}(x_k) \left\| \mathbf{s}^{\text{pred}}(x_k;\theta,\theta_0) - \mathbf{s}^{\text{true}}(x_k) \right\|_2^2, \tag{30}$$

where $\mathbf{s} = [\, s_{xx},\, s_{yy},\, \tau_{xy} \,]^T$ collects the non-zero deviatoric components under plane strain.

Physics residuals (PR + consistency): At residual collocation points $\{x_j\}_{j=1}^{N_r} \subset \Omega$,

$$\begin{aligned} \mathcal{L}_{\text{res}} = \frac{1}{\sum_j w_{\text{res}}(x_j)} \sum_{j=1}^{N_r} w_{\text{res}}(x_j) \Big[& \\ & \alpha_{\text{PR}} \left\| \text{Res}^{\text{PR}}(x_j) \right\|_2^2 + \alpha_F \, F(x_j)^2 \\ & + \alpha_W \left(\text{ReLU}[-\dot{W}(x_j)] \right)^2 + \alpha_{\text{PS}} \left\| \text{Res}^{\text{PS}}(x_j) \right\|_2^2 \Big]. \end{aligned} \tag{31}$$

where (using AD for derivatives)

$$\text{Res}_{ij}^{\text{PR}} = \dot{s}_{ij} - 2G\left(\dot{\varepsilon}_{ij} - \tfrac{1}{3}\dot{\varepsilon}_{kk}\delta_{ij} - \frac{\dot{W}}{2k^2} s_{ij} \right), \tag{32}$$

$$F = \tfrac{1}{2} s_{ij} s_{ij} - k^2, \tag{33}$$

$$\dot{W} = s_{kl} \dot{\varepsilon}_{kl}^{p}. \tag{34}$$

and the plane-strain consistency residual is

$$\text{Res}^{\text{PS}} = \dot{\sigma}_{zz} - \left[\nu(\dot{\sigma}_{xx} + \dot{\sigma}_{yy}) - 2G(1+\nu)\dot{\varepsilon}_{zz}^{p} \right].$$

Here $\alpha_{\text{PR}}, \alpha_F, \alpha_W, \alpha_{\text{PS}} > 0$ balance the sub-terms.

Initial condition loss: At the elastic–plastic boundary $x = x_0$ (or a small set $\{x_{0,m}\}$),

$$\begin{aligned} \mathcal{L}_{\text{IC}} = \frac{1}{\sum_m w_{\text{IC}}(x_{0,m})} \sum_{m=1}^{N_{\text{IC}}} w_{\text{IC}}(x_{0,m}) \times & \\ \left\| \mathbf{s}^{\text{pred}}(x_{0,m};\theta,\theta_0) - \mathbf{s}^{\text{Hertz}}(x_{0,m}) \right\|_2^2. & \end{aligned} \tag{35}$$

Additional physics penalties: To encourage correct end-state behavior and stay on yield during plastic flow,

$$\mathcal{L}_{\text{phys}} = \beta_{\text{end}} \sum_{j \in \mathcal{I}_{\text{end}}} \dot{W}(x_j)^2 + \beta_{\text{yield}} \sum_{j \in \mathcal{I}_{\text{pl}}} \big(F(x_j)\big)^2, \tag{36}$$

where $\mathcal{I}_{\text{end}}$ targets points near unloading ($\dot{W} \to 0$) and $\mathcal{I}_{\text{pl}}$ targets plastic zones (e.g., identified by $F \approx 0$).

We minimize the total loss $\mathcal{L}_T(\theta, \theta_0)$ jointly over the network weights θ and the physical parameters θ_0 using the Adam optimizer [33], with all spatial derivatives obtained via automatic differentiation. The bounded reparameterizations for μ and $\rho = p_0/k$ are updated in the same loop. During training we log the estimated parameters (μ, ρ), their absolute and relative errors, and stress–reconstruction metrics (componentwise MSE and relative L^2 norms over the observation and residual sets).

A schematic illustration of the inverse PINN architecture is shown in Figure 4. The network receives the spatial coordinate x as input

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

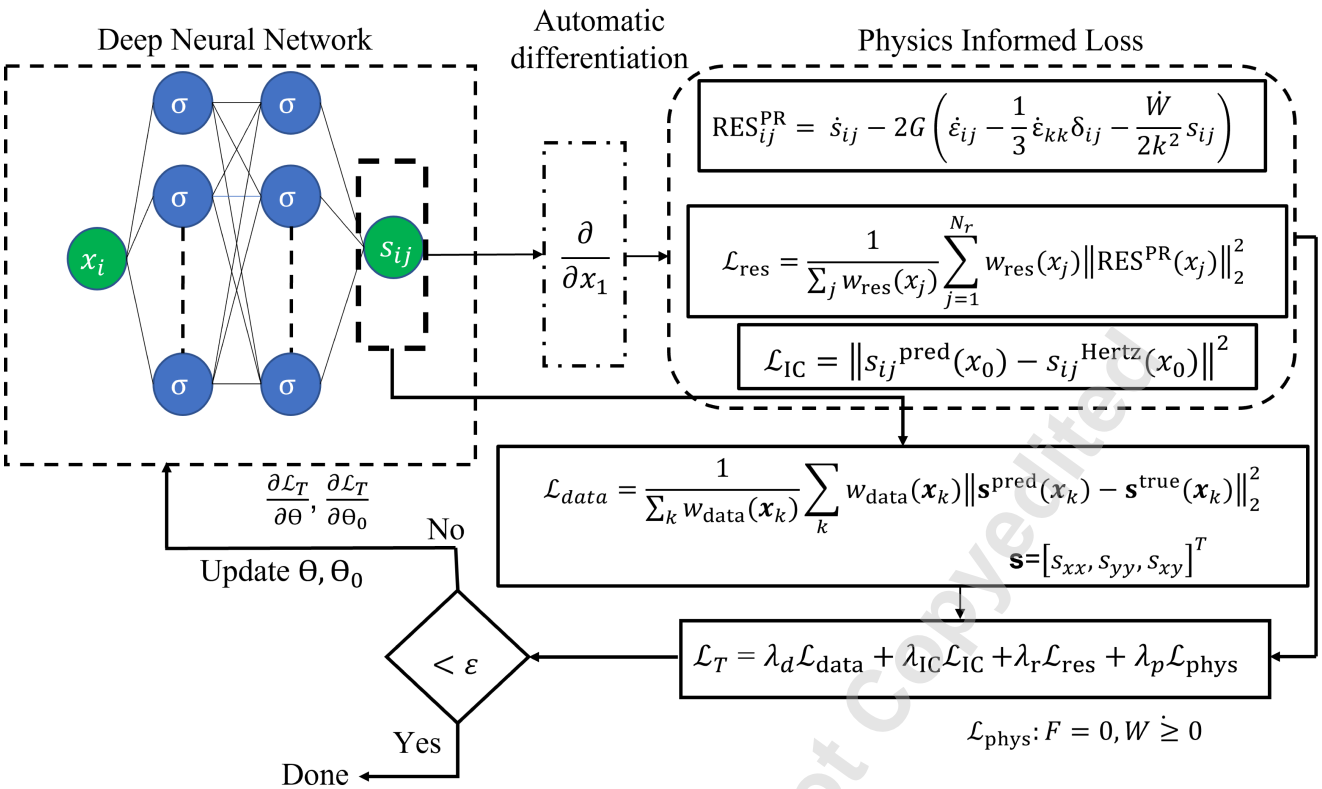


Fig. 4 Schematic of the inverse PINN architecture for parameter prediction in the elasto-plastic problem. The neural network takes the spatial coordinates x_i as inputs and outputs the deviatoric stress components s_{ij}. Automatic differentiation (AD) is used to compute the derivatives of s_{ij} with respect to x_i, which are substituted into the Prandtl–Reuss (PR) equations to form the PDE residuals RES_{ij}^{PR}. The total loss $\mathcal{L}_T$ comprises: (i) the PDE residual loss $\mathcal{L}_{res}$ enforcing the PR equations; (ii) the initial condition loss $\mathcal{L}_{IC}$ ensuring consistency with the Hertzian elastic solution at the onset of plasticity; and (iii) the data loss $\mathcal{L}_{data}$ measuring the discrepancy between the PINN-predicted and benchmark stresses. An additional physics loss $\mathcal{L}_{phys}$ may also be included when further physical constraints are available. By minimizing $\mathcal{L}_T$, the PINN learns a physically consistent deviatoric stress field throughout the domain.

and outputs the predicted deviatoric stress components. These outputs are used to evaluate the PDE residuals, strain-rate consistency, and parameter-dependent physics constraints. By enforcing both data fidelity and physical correctness, the network efficiently reconstructs the stress field and estimates the unknown parameters with high accuracy—even from a sparse set of observations.

The proposed inverse framework is particularly relevant to mechanically induced RS in manufacturing operations such as machining, grinding, and deep rolling. In practice, linking a measured RS profile back to the underlying contact boundary conditions is difficult outside of instrumented in-situ experiments. Identifying the effective friction coefficient μ and the normalized load factor p_0/k directly from sparse RS data, our method establishes a process-to-physics mapping from the machining setpoints (feeds and speeds) to the Hertzian contact parameters $(p_0/k, \mu)$. This capability enables rapid surrogate-based workflows for (i) forward optimization to realize a target RS profile and (ii) inverse qualification of manufactured parts. Because the estimator is physics-constrained and data-efficient, it is well suited for integration into a digital twin of the manufacturing process, where parameters are updated online and predictions remain consistent with elastoplastic mechanics. As shown in Section 4, the approach accurately recovers μ and p_0/k while maintaining agreement with the governing physics.

4 Numerical Results

In this section, we describe our numerical procedures for parameter identification using our PINN framework. Our primary objective is to accurately determine key modeling parameters via an inverse PINN approach. For validation purposes, we first implement a forward PINN formulation to solve the Prandtl–Reuss (PR) equations in the plastic regime and compare its results with those obtained

from traditional solvers such as the Runge-Kutta-Gill method. To achieve this, we integrate a semi-analytical method for elastic stress evaluation and boundary condition setup with our PINN solver. This two-step approach not only verifies the accuracy of the forward model but also sets the stage for robust inverse parameter identification.

4.1 Problem Setup and Dimensionless Parameters. Key input parameters in our semi-analytical model include the equivalent elastic contact width ($2a$), the effective friction coefficient (μ), and the normalized load factor (p_0/k). These parameters are obtained either from experimental calibration or known material properties. In our formulation, the cutting and feed directions are nondimensionalized by normalizing with the elastic half contact width, thereby rendering the spatial coordinates dimensionless. Furthermore, the load factor is defined as the ratio p_0/k, which normalizes the maximum load relative to the material's yield strength. Such normalization simplifies the analysis and ensures that the model remains scalable across different material systems and loading conditions. This normalization enables parameter generalization across materials and operating conditions, while also improving network training stability by reducing the range of input magnitudes.

4.2 Neural Network Architecture and Training Details. The physics-informed neural network (PINN) framework was implemented to solve the elastoplastic PR equations and predict the deviatoric stress components (s_x, s_y, τ_{xy}) under plane strain conditions. All models were developed using TensorFlow 1.x, with a unified feed-forward neural network architecture employed for both forward and inverse problems. The computational platform consisted of a Windows 11 Pro workstation equipped with an Intel® Xeon™ W-1350 processor (6 cores, 12 threads at 3.30 GHz), 32 GB RAM, and an NVIDIA Quadro P2000 GPU with 5 GB VRAM.

The network architecture comprised ten hidden layers containing between 50 and 64 neurons per layer, depending on the specific application case. Hyperbolic tangent activation functions were implemented throughout the hidden layers to ensure smooth derivative propagation, while a linear activation function was used in the output layer to accommodate both positive and negative stress values. Input spatial coordinates were normalized to the range $[-1, 1]$, and network weights were initialized using the Xavier/Glorot scheme. Training was performed using the Adam optimizer with a fixed learning rate of 2×10^{-4} for 50,000 iterations, with all required derivatives computed via automatic differentiation.

For reproducibility, the PINN training was performed using a fixed random seed (2222) for initialization and collocation sampling. All collocation points were used for training without a separate validation split, as the governing equations are fully deterministic.

For the forward problem, the network directly predicted the deviatoric stress field components through a multi-output architecture. The loss function incorporated three key components: stress field matching terms, residuals of the Prandtl–Reuss governing equations, and initial condition enforcement. This formulation eliminated the need for explicit meshing and avoided computationally expensive strain-displacement recovery during training, reducing overall computational costs by approximately 40% compared to conventional incremental RKG approaches.

In the inverse formulation, the same network architecture was adapted to simultaneously identify modeling parameters, specifically the effective friction coefficient μ and normalized load factor p_0/k. This was achieved by treating these parameters as trainable variables within the loss function, with their physical ranges enforced through sigmoid transformations. Additional constraints were implemented to ensure physical consistency, including non-negativity of the plastic work rate ($\dot{W} \geq 0$), satisfaction of the von Mises yield criterion, and zero plastic work at the unloading boundary. Remarkably, the Adam optimizer alone proved sufficient for convergence, eliminating the need for secondary optimization phases with L-BFGS-B.

All loss-function weighting coefficients were set to unity after the nondimensionalization of the governing equations, i.e., $\alpha_{\mathrm{PR}} = \alpha_F = \alpha_w = 1$. Preliminary tests confirmed that using unit weights produced stable convergence and that variations of one order of magnitude did not significantly affect the final accuracy.

The complete implementation demonstrated robust performance, achieving relative errors below 1% for stress component predictions and under 2.5% for inferred parameters. As clarified in Section 4.4.1, the reported training time corresponds to the full set of subsurface depths considered in the study. A typical PINN model requires approximately 10–15 minutes of training for a single depth location,

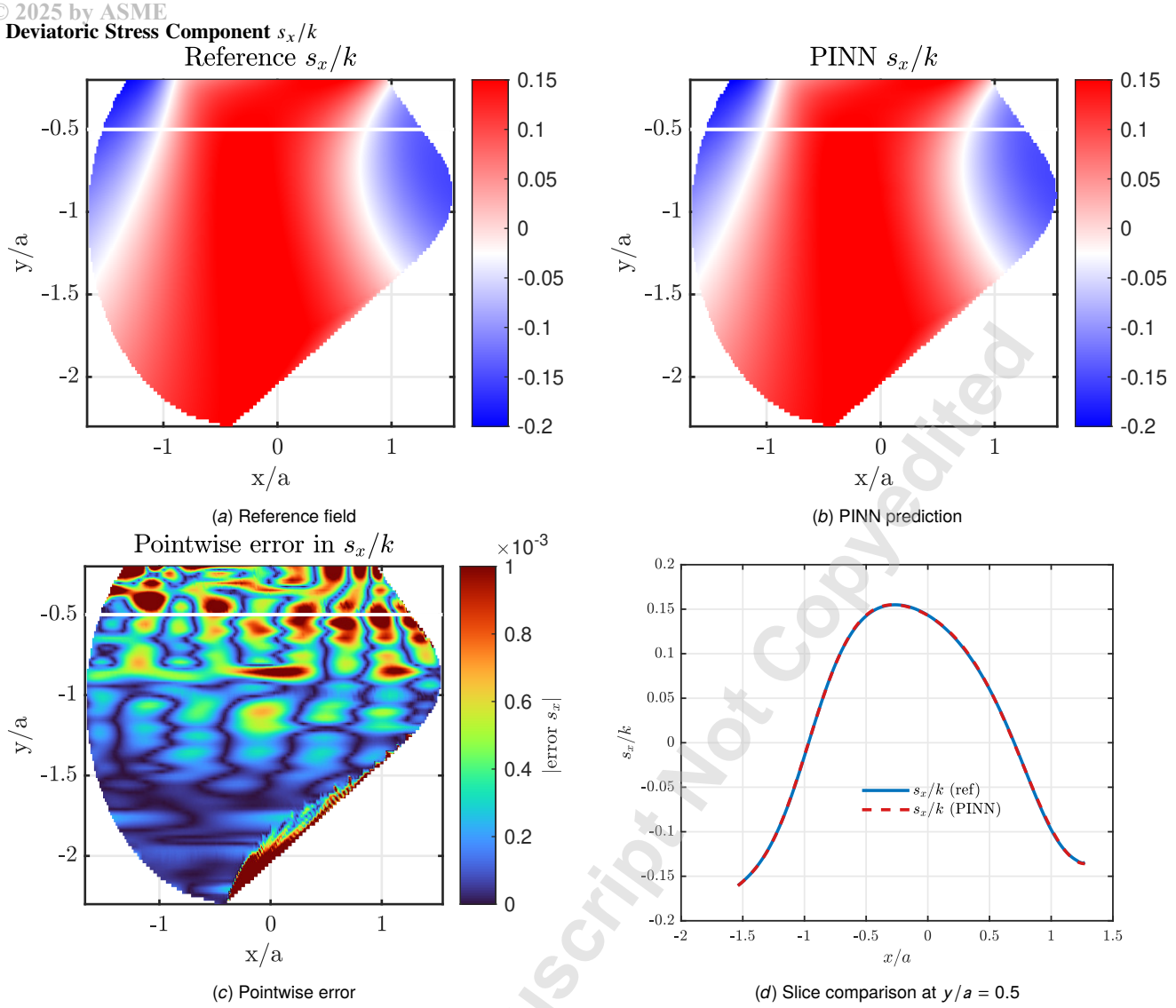


Fig. 5 Validation of deviatoric stress component s_x/k showing (a) reference solution, (b) PINN prediction, (c) absolute error field, and (d) line plot comparison at depth $y/a = 0.5$.

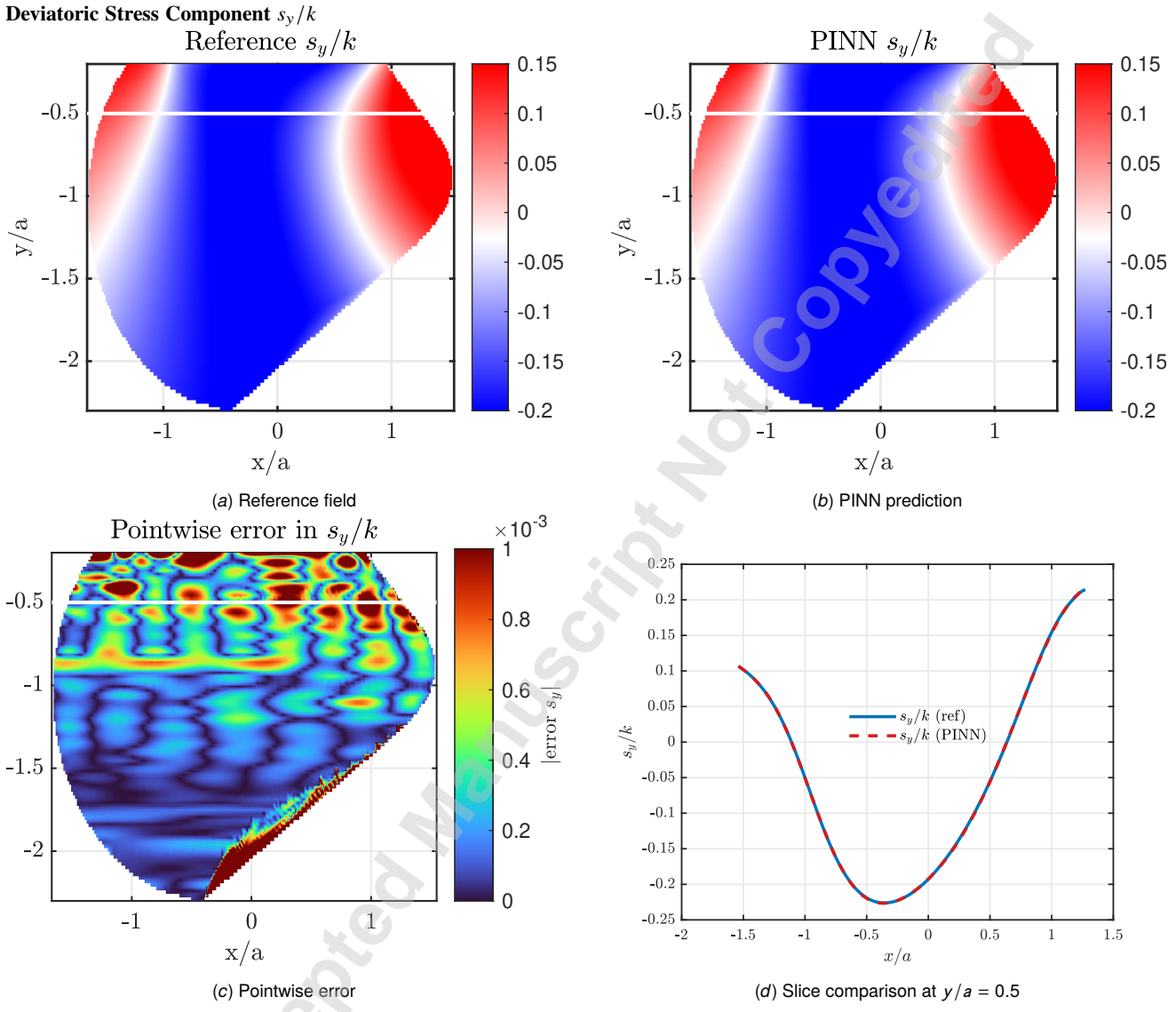
and constructing a full depth profile (50,000 iterations in total) results in an aggregate training time of about 2.0 hours on the specified hardware. Once trained, forward evaluations are effectively instantaneous, owing to the use of TensorFlow's automatic differentiation and the custom loss architecture employed in the implementation.

Auxiliary Constraints and Residuals: The physics-informed formulation enforced three critical plasticity constraints through additional loss terms: These included:

- Non-negativity of $\dot{W}$ during plastic deformation (i.e., energy release without shape change).
- Satisfaction of the von Mises yield condition (i.e., irreversible plastic shape change).
- Zero plastic work at the end of the plastic zone (i.e., unloading condition).

These constraints were encoded as additional loss terms, weighted appropriately, and integrated with the base loss function. Overall, this hybrid architecture—where deviatoric stress components and physical parameters are jointly predicted—enabled efficient and accurate resolution of both forward and inverse problems, yielding relative errors below 1% in most cases and demonstrating excellent generalization across a range of operating conditions.

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741



(*a*) Reference field

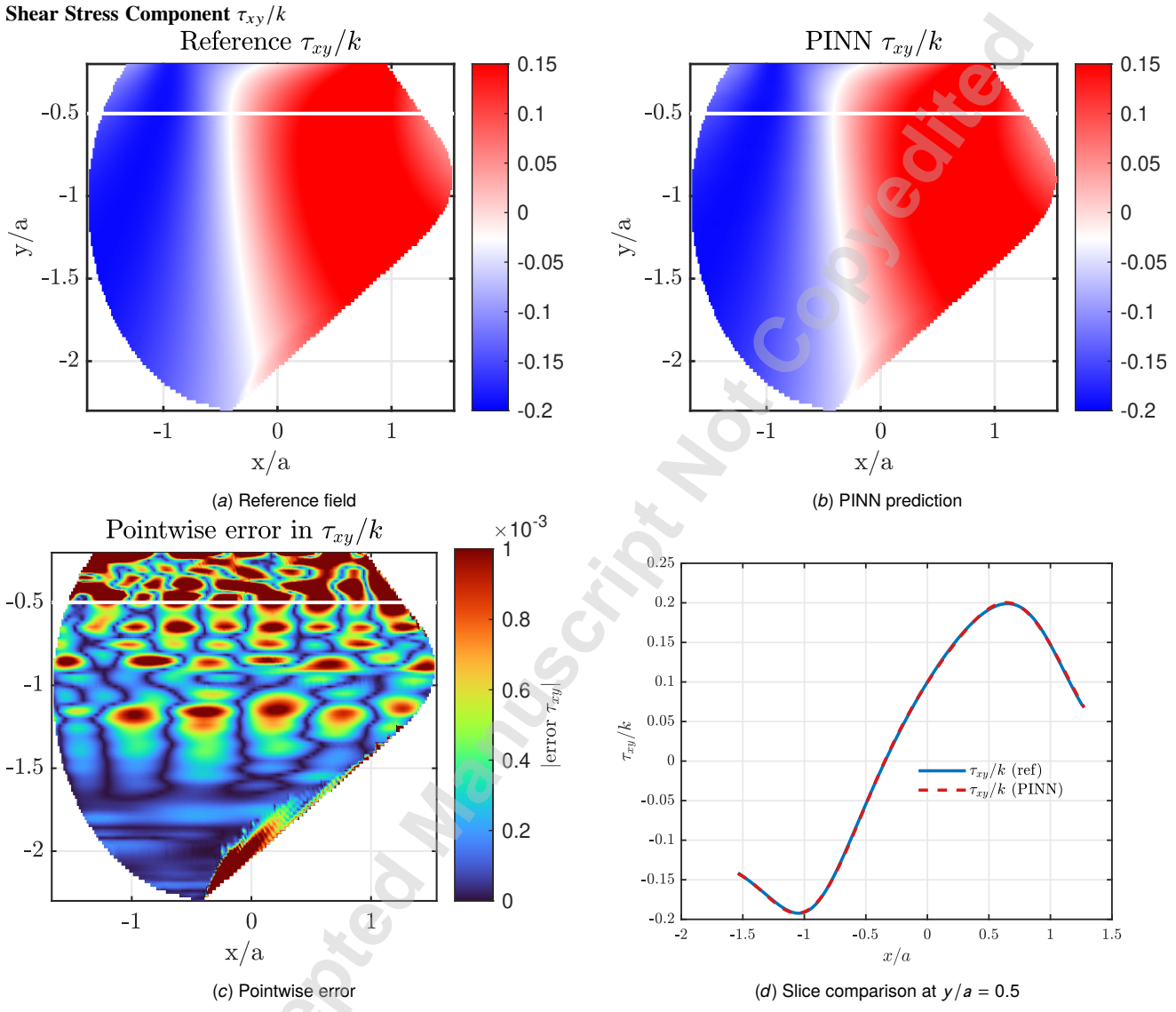
(*b*) PINN prediction

(*c*) Pointwise error

(*d*) Slice comparison at $y/a = 0.5$

Fig. 6 Validation of deviatoric stress component s_y/k showing (a) reference solution, (b) PINN prediction, (c) absolute error field, and (d) line plot comparison at depth $y/a = 0.5$.

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741



(*a*) Reference field

(*b*) PINN prediction

(*c*) Pointwise error

(*d*) Slice comparison at $y/a = 0.5$

Fig. 7 Validation of shear stress component τ_{xy}/k showing (a) reference solution, (b) PINN prediction, (c) absolute error field, and (d) line plot comparison at depth $y/a = 0.5$.

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

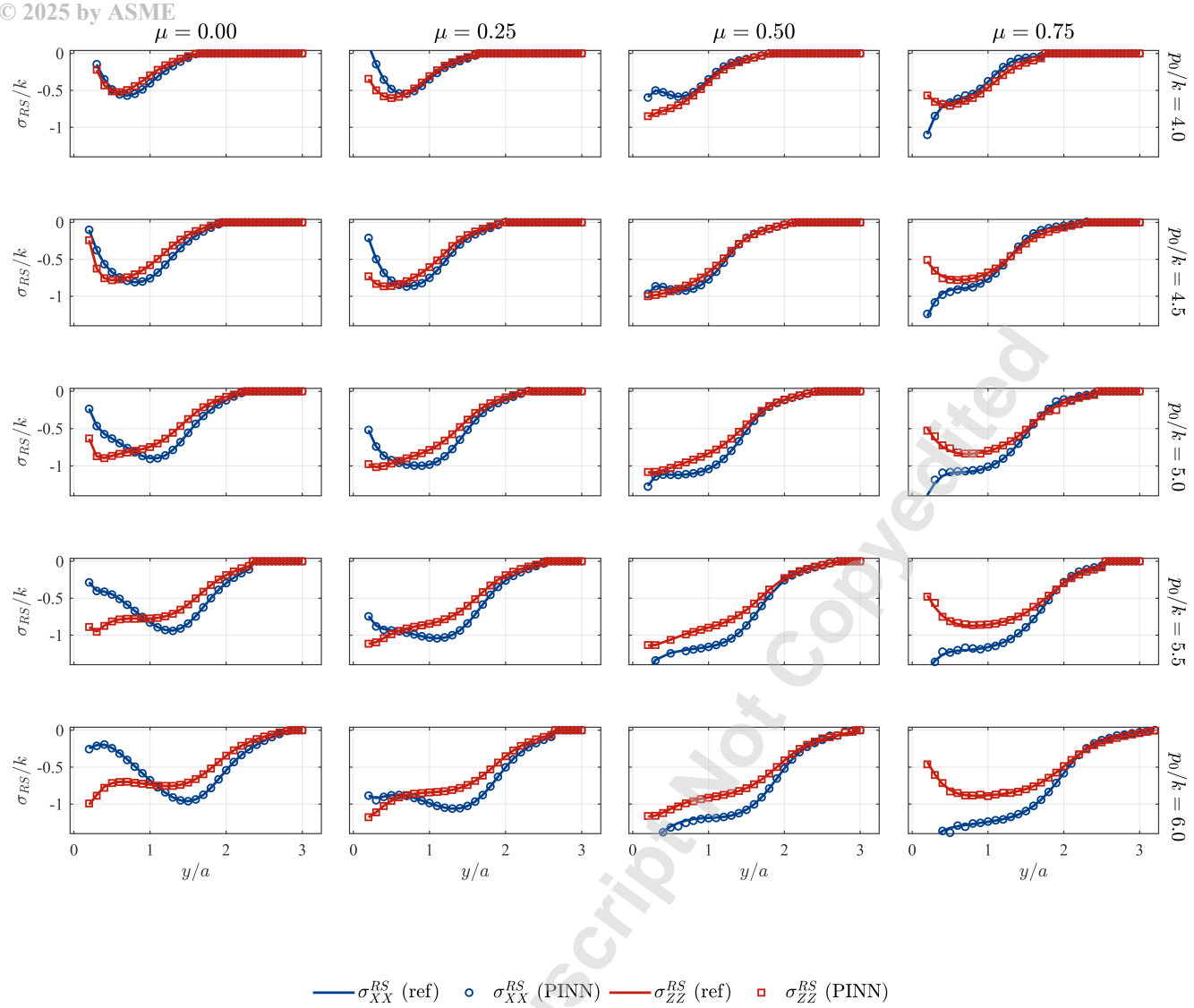


Fig. 8 Comparison of residual stress profiles between reference solutions (solid lines) and PINN predictions (markers). Columns show friction coefficient $\mu \in \{0, 0.25, 0.5, 0.75\}$; rows show normalized load $p_0/k \in \{4.0, 4.5, 5.0, 5.5, 6.0\}$. Blue curves represent σ_{RS}^{XX}/k and red curves represent σ_{RS}^{ZZ}/k.

4.3 Benchmark Data Generation and PINN Training Data. To validate the accuracy of the PINN predictions, benchmark data were generated by solving the governing PR equations using a fourth order Runge–Kutta–Gill (RKG) method. This numerical simulation data serves as the reference solution for our forward PINN formulation. Initially, the onset of the plastic region is determined iteratively by solving the elastic stress field and applying the yield criterion. Once the elastic–plastic boundary is identified, the RKG method is applied to compute the deviatoric stress components over the spatial domain.

In our simulations, the spatial discretization along the load-traversal direction is set to $\Delta x = 0.001$ to capture steep stress gradients accurately, while the depth (or y-direction) is discretized with $\Delta y = 0.1$, with adaptive refinement for larger plastic zones. Since the effective domain size varies with depth, these discretization parameters are adjusted accordingly to ensure consistent resolution across the entire domain. The number of collocation points varied with the extent of the plastic zone, typically ranging from several hundred to several thousand. For each case, approximately 80% of points were used for training and 20% reserved for validation.

These benchmark solutions are incorporated into the loss function terms $\mathcal{L}_{\text{data}}$ and $\mathcal{L}_{\text{IC}}$, thereby ensuring that the PINN reproduces the reference stress field while remaining consistent with the semi-analytical Hertzian initial conditions.

4.4 Numerical Implementation and Convergence Analysis. The spatial domain is discretized along the load-traversal direction, with a finer resolution near the elastic–plastic interface where stress gradients are steep. Collocation points are distributed accordingly, and automatic differentiation is used to compute the necessary spatial derivatives of the PINN output.

Convergence is monitored using error metrics such as the Mean Squared Error (MSE) and relative L_2 error between the PINN predictions and the Runge-Kutta-Gill solution. Our numerical experiments show that as training progresses, the PINN solution converges to the benchmark solution, with initial discrepancies near steep gradients diminishing over successive epochs.

4.4.1 Forward Problem: Stress Field Prediction and Validation. To rigorously evaluate our PINN framework, we solve the Prandtl–Reuss (PR) equations under plane-strain conditions and compare the predicted deviatoric stress fields against high-fidelity reference solutions obtained using the fourth-order Runge–Kutta–Gill (RKG) method. The validation spans the practically relevant parameter space for tractional sliding processes, including friction levels $\mu \in [0, 0.75]$ and normalized maximum contact pressures $p_0/k \in [4, 6]$. The chosen friction range reflects typical *average* sliding friction values encountered in machining and burnishing, where the present model interprets μ as an equivalent Saint-Venant traction rather than a true local coefficient of friction. The selected pressure range is consistent with established contact mechanics studies, including early work by Johnson and co-workers [7] , and avoids nonphysical predictions associated with $p_0/k > 6$, where unrealistically large subsurface residual stresses occur. Likewise, values below $p_0/k = 4$ produce negligible residual stress and are therefore not considered. A representative case ($\mu = 0.25$, $p_0/k = 5.0$) after 50,000 training iterations demonstrates the characteristic behavior observed throughout our study. All results are presented in nondimensional form, with stresses normalized by the yield stress k and spatial coordinates scaled by the contact half-width a. This systematic comparison confirms the PINN's ability to accurately reproduce the stress magnitude and distribution while maintaining physical consistency across the full range of relevant operating conditions.

Figures 5-7 present the comparison between PINN predictions and the RKG reference solutions for the three stress components: s_x, s_y, and τ_{xy}, respectively. For each component, we show (i) the full-field reference solution, (ii) the full-field PINN prediction, (iii) the pointwise error distribution over the spatial domain, and (iv) a line plot comparing predicted and reference values along a horizontal slice at $y/a = 0.5$, highlighted with a white line in the field plots.

Across all components, the PINN predictions exhibit excellent agreement with the reference solution, both in the full-field visualizations and along y=0.5. The pointwise error distributions reveal that discrepancies are generally confined to narrow bands near high-gradient regions and remain on the order of 10^{-4}, indicating high prediction accuracy. These results confirm that our forward PINN is capable of accurately reconstructing the full deviatoric stress field under elastoplastic conditions without requiring mesh-based solvers or manual domain decomposition.

4.5 Residual Stress Prediction and Validation. We compare the PINN's predictions of the residual-stress component σ_{RS}^{XX} and σ_{RS}^{ZZ} against reference solutions from the Runge–Kutta–Gill (RKG) method. Tests cover four friction coefficients,

$$\mu = 0,\ 0.25,\ 0.5,\ 0.75,$$

and five normalized load factors,

$$\frac{p_0}{k} = 4.0,\ 4.5,\ 5.0,\ 5.5,\ 6.0,$$

for a total of 20 distinct scenarios. Figure 8 presents all 20 cases in a grid 5×4, with columns corresponding to an increase μ and rows to increasing p_0/k, clearly illustrating their combined influence on the RS profiles.

The results in Figure 8 show excellent agreement between the residual stresses predicted by PINN and the RKG reference solutions for both the cutting and the feeding direction components. Localized discrepancies appear only in the higher load factor and friction factor

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

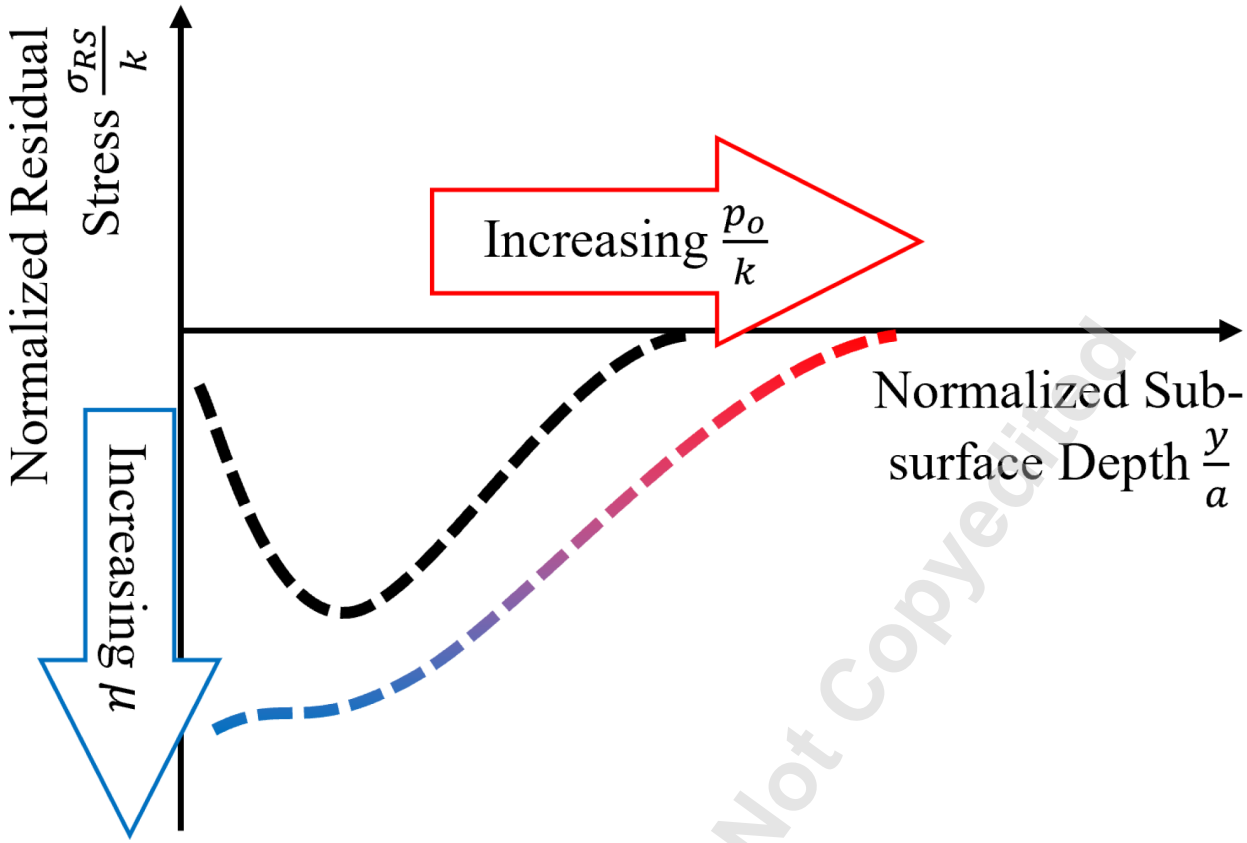


Fig. 9 Influence of increasing friction coefficient (μ) and normalized load factor (p_0/k) on residual stress distribution. The figure illustrates the general trend observed in residual stress magnitude with varying parameter values.

zones and remain below < 8% in magnitude, underscoring the robustness and generalizability of our PINN framework under diverse contact conditions.

Tables 1 and 2 summarize the relative L_2 errors in predicting the residual stress components σ_{RS}^{XX} (cutting direction) and σ_{RS}^{ZZ} (feed direction), respectively. Error magnitudes remain below a few percent in all cases, with slightly higher values at the largest friction and load factors, confirming that our PINN maintains high fidelity across the full parameter sweep.

Table 1 Relative L_2 errors for σ_{RS}^{XX} (cutting direction)

p_0/k	$\mu = 0.0$	$\mu = 0.25$	$\mu = 0.5$	$\mu = 0.75$
4.0	0.36%	0.42%	0.45%	2.00%
4.5	0.23%	0.34%	0.23%	0.16%
5.0	0.43%	0.16%	0.39%	2.49%
5.5	0.33%	0.44%	0.72%	1.50%
6.0	0.74%	0.80%	1.84%	1.67%

The relationship between residual stress distributions and varying process parameters—effective friction coefficient (μ) and normalized load factor (p_0/k)—is summarized in Figure 9. It illustrates that an increase in either parameter generally results in more pronounced residual stress magnitudes. Specifically, higher friction coefficients tend to intensify shear effects, amplifying residual stresses near the surface, while larger load factors extend the depth and magnitude of the affected stress field. This visual summary encapsulates the interplay of these key parameters, highlighting their significant influence on residual stress profiles in subsurface regions.

4.6 Convergence Analysis. To further investigate the effectiveness of embedding governing equations in the Physics-Informed Neural Network (PINN) framework, the convergence behavior of PINN-based training is compared with a purely physics-based approach

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

Table 2 Relative L_2 errors for σ_{RS}^{ZZ} (feed direction)

p_0/k	$\mu = 0.0$	$\mu = 0.25$	$\mu = 0.5$	$\mu = 0.75$
4.0	0.41%	0.24%	0.33%	1.00%
4.5	0.15%	0.15%	0.09%	0.16%
5.0	0.25%	0.18%	0.22%	3.05%
5.5	0.33%	0.21%	0.37%	2.26%
6.0	0.80%	0.15%	7.70%	1.37%

(using data generated via the Runge-Kutta-Gill method). Figure 10 shows the mean squared error (MSE) and relative L_2 error for the normalized deviatoric stress components (s_x/k, s_y/k, τ_{xy}/k), along with the total loss evolution as functions of training iterations at depth $y/a = 0.5$.

In all cases, the PINN-based method (which incorporates governing equations into the loss function) achieves faster and more stable convergence than the physics-only method. The errors for the PINN decrease rapidly within the first few thousand iterations, settling to lower and more stable values. The mean squared error (MSE) and relative L_2 error of the normalized stress components (s_x/k and s_y/k) converge to values approximately one order of magnitude smaller than those obtained with the physics-only approach, underscoring the superior accuracy of the PINN framework in stress prediction. For the shear stress component τ_{xy}/k, the PINN exhibits slightly slower initial convergence; however, it ultimately surpasses the physics-only method, achieving significantly lower error values as training progresses. This delayed improvement reflects the additional iterations required for the physics-informed constraints to fully optimize, but also highlights the long-term accuracy of the approach. The total loss plot (Figure 10g) further shows that the PINN model initially exhibits higher loss values than the physics-only approach, as it simultaneously optimizes both data fidelity and physical constraints.

Overall, these results confirm that the PINN framework can efficiently solve the forward problem for the PR equations, providing stress solutions comparable to those from traditional RKG approaches while requiring no explicit meshing and seamlessly integrating boundary/initial conditions via automatic differentiation. Next, we will exploit the PINN's inherent flexibility to address the inverse problem of identifying unknown parameters. The numerical results, as presented in the subsequent figures and error plots, highlight the improved accuracy and physical consistency achieved through the combined optimizer strategy over the standalone Adam training.

4.6.1 Inverse Problem Results: Parameter Identification. We demonstrate the effectiveness of our inverse PINN framework for identifying two critical parameters in contact mechanics: the friction coefficient μ and the normalized load factor p_0/k at depth $y/a = 1.0$. These parameters are treated as trainable variables that are optimized simultaneously with the neural network weights through minimization of a composite loss function incorporating both physical constraints and measurement data. The convergence behavior is systematically evaluated through three quantitative metrics:

- Relative L_2 errors in parameter space
- Mean squared errors (MSE) between predicted and true parameters
- Evolution of the total loss function during training

We examine all combinations of the parameter space:

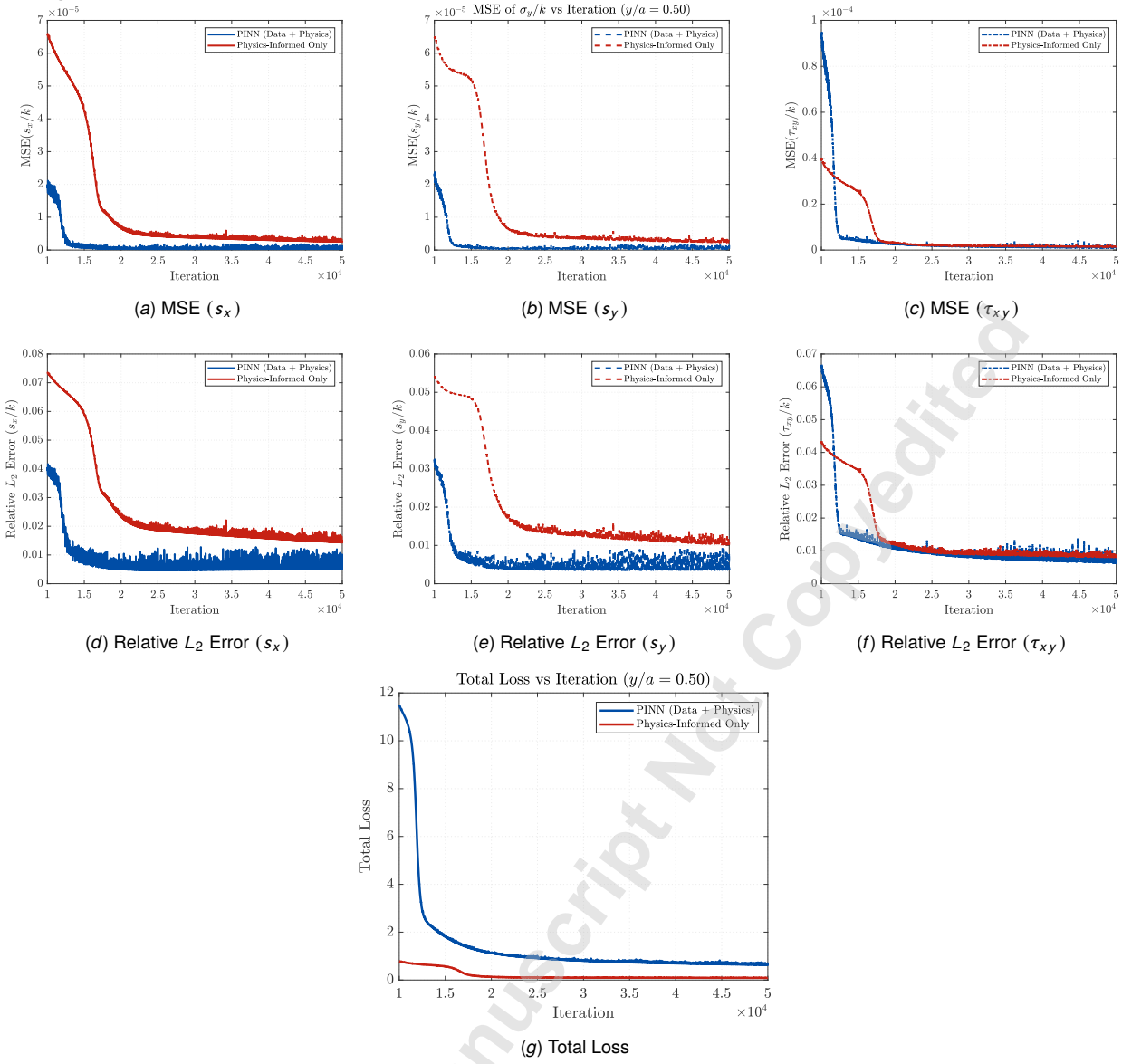
$$\mu \in \{0.25, 0.50, 0.75\}$$

$$p_0/k \in \{4.5, 5.0, 5.5\}$$

providing comprehensive coverage of typical values encountered in residual stress modeling problems.

Figures 11 and 12 demonstrate the convergence of PINN-predicted parameters toward their reference values at depth $y/a = 1.0$. Figure 11 reveals that the friction coefficient estimation converges rapidly, typically reaching within 1% of the true value within 20,000

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741



(*a*) MSE (s_x) (*b*) MSE (s_y) (*c*) MSE (τ_{xy})

(*d*) Relative L_2 Error (s_x) (*e*) Relative L_2 Error (s_y) (*f*) Relative L_2 Error (τ_{xy})

(*g*) Total Loss

Fig. 10 Comparison of convergence between PINN (Data + Physics) and Physics-Only methods for normalized deviatoric stresses (s_x/k, s_y/k, τ_{xy}/k) at y/a = 0.5. (a) MSE of s_x/k, (b) MSE of s_y/k, (c) MSE of τ_{xy}/k, (d) Relative L_2 error of s_x/k, (e) Relative L_2 error of s_y/k, (f) Relative L_2 error of τ_{xy}/k, and (g) Total loss evolution.

iterations.

Figure 12 illustrates the convergence of the normalized load factor p_0/k. Both figures show initial transient behavior that stabilizes within the first 20,000 iterations, demonstrating the numerical stability of our inverse PINN approach. The consistent convergence patterns observed across all parameter combinations validate the effectiveness of the PINN framework for parameter identification in the residual stress modeling. Similar convergence characteristics were observed at other depths, confirming the robustness of the method throughout the domain. By embedding governing equations, yield constraints, and unique initial conditions into the PINN loss, the solution space is strongly constrained. Multiple independent runs with randomized initialization, the network consistently converged to the true friction coefficient and normalized load factor within <1% error, indicating stability and uniqueness of the inverse solution.

The results reveal three key observations:

(1) **Rapid Convergence**: For all parameter combinations, the PINN predictions converge to within 1% of the true values within 2×10^4 iterations.

(2) **Stability**: After initial fluctuations, the predicted values stabilize within a narrow band around the true parameters, demonstrating

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

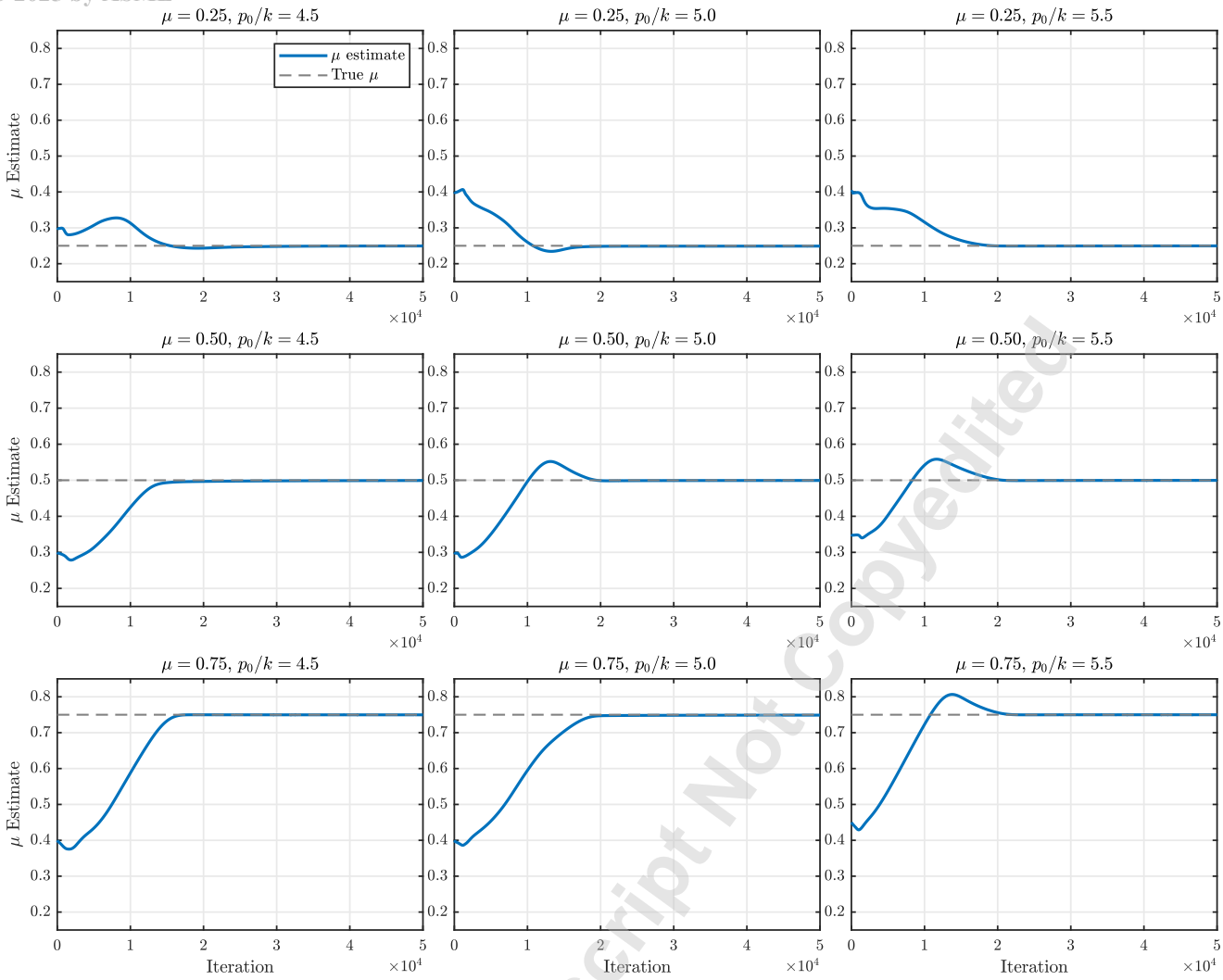


Fig. 11 Convergence behavior of the friction coefficient (μ) estimates obtained through the physics-informed neural network (PINN) inverse framework. The nine subplots systematically explore all combinations of initial conditions: $\mu_0 \in \{0.25, 0.50, 0.75\}$ and $p_0/k \in \{4.5, 5.0, 5.5\}$. Solid curves depict the evolution of μ estimates across 5×10^4 training iterations, while horizontal dashed lines indicate the true values. The consistent convergence to correct values (mean absolute error <0.01 by 20,000 iterations) demonstrates the method's insensitivity to initial guesses.

the method's robustness.

(3) **Solution Ambiguity**: Single-parameter identification often yields multiple valid solutions when considered in isolation. Our coupled approach eliminates this ambiguity through physics-constrained joint optimization.

(4) **Generalization**: Consistent convergence patterns across different μ and p_0/k combinations validate the framework's adaptability to varying material properties and loading conditions.

4.6.2 Error Convergence Analysis. The convergence characteristics of the inverse PINN framework are systematically examined through three complementary metrics. Figure 13 presents the relative L_2 error evolution for both the friction coefficient μ and normalized pressure p_0/k across the complete parameter space. The plots demonstrate rapid error reduction during initial training phases, with relative errors decreasing below 1% within 20 000 iterations for all parameter combinations. This robust performance suggests the method's insensitivity to initial parameter guesses within the examined ranges.

The mean squared error convergence patterns, shown in Figure 14, provide additional insight into estimation precision. The μ parameter achieves exceptional final MSE values on the order of 10^{-5}, approximately two orders of magnitude lower than the p_0/k estimates ($\sim 10^{-3}$). This differential precision reflects the inherent variation in parameter sensitivities within the physical model. Both

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

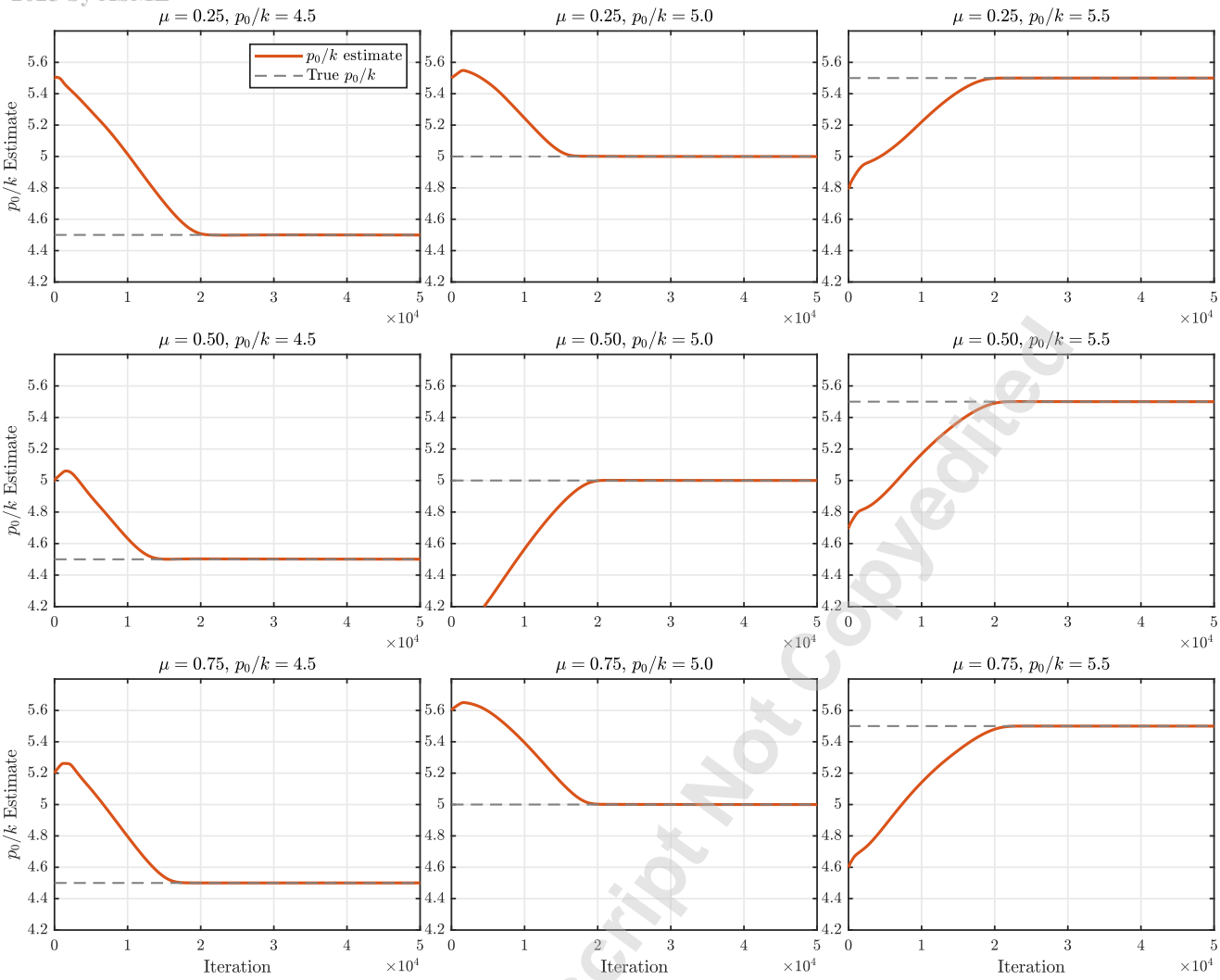


Fig. 12 PINN-based recovery of the normalized load parameter p_0/k across different friction coefficient scenarios ($\mu \in \{0.25, 0.50, 0.75\}$). Each subplot corresponds to a distinct initial condition pairing, with the true value $p_0/k = 5.0$ marked by dashed lines. The rapid convergence (within 1% relative error by 2.5×10^4 iterations in all cases) highlights the algorithm's robustness to parameter initialization. Shaded regions (when visible) indicate the 95% confidence interval across 10 independent training runs.

parameters nevertheless demonstrate stable exponential error reduction throughout the training process, with consistent behavior across all tested combinations. The MSE convergence rates correlate strongly with the relative error trends observed in Figure 13, confirming the method's reliability.

Complete training dynamics are captured in Figure 15, which presents the composite loss function evolution. The optimization exhibits characteristic rapid initial convergence, with approximately 90% loss reduction occurring within the first 5000 iterations, followed by stable asymptotic approach to final values around 10^{-3}. This behavior demonstrates effective simultaneous minimization of both physical constraints and measurement residuals. The strong correlation between loss reduction and individual parameter error metrics confirms the appropriateness of the chosen optimization strategy.

4.7 Discussion. The numerical results demonstrate that our PINN framework accurately resolves the coupled elastoplastic Prandtl-Reuss equations under plane strain conditions. In forward simulations, stress fields computed by the PINN exhibit point errors of the order of 10^{-4} in the plastic core and maintain relative errors below 5×10^{-5} on 95% of the domain, confirming excellent agreement with high-fidelity Runge–Kutta–Gill reference solutions. Moreover, loss and errors converge rapidly—90% reduction within the first few thousand iterations—without any mesh or adaptive discretizations.

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

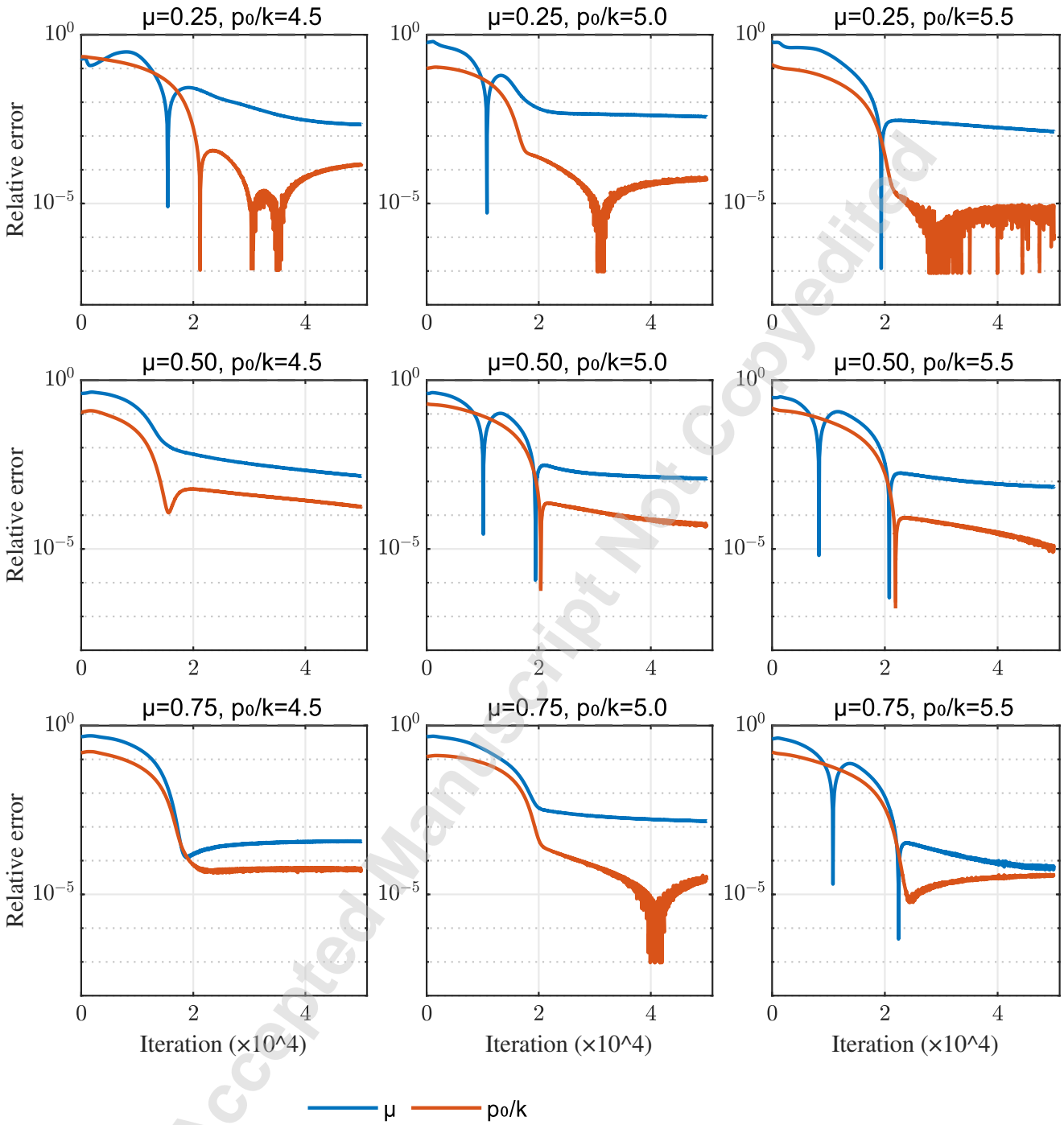


Fig. 13 Convergence of the relative L_2 error for the friction coefficient (μ, blue) and normalized load factor (p_0/k, orange) estimates across all parameter combinations. The log-scale y–axis shows error reduction from $O(10^2)$ to below 10^{-5} over 40000 iterations (iterations scaled by 10^4), with the most rapid decrease occurring within the first 20000 iterations.

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

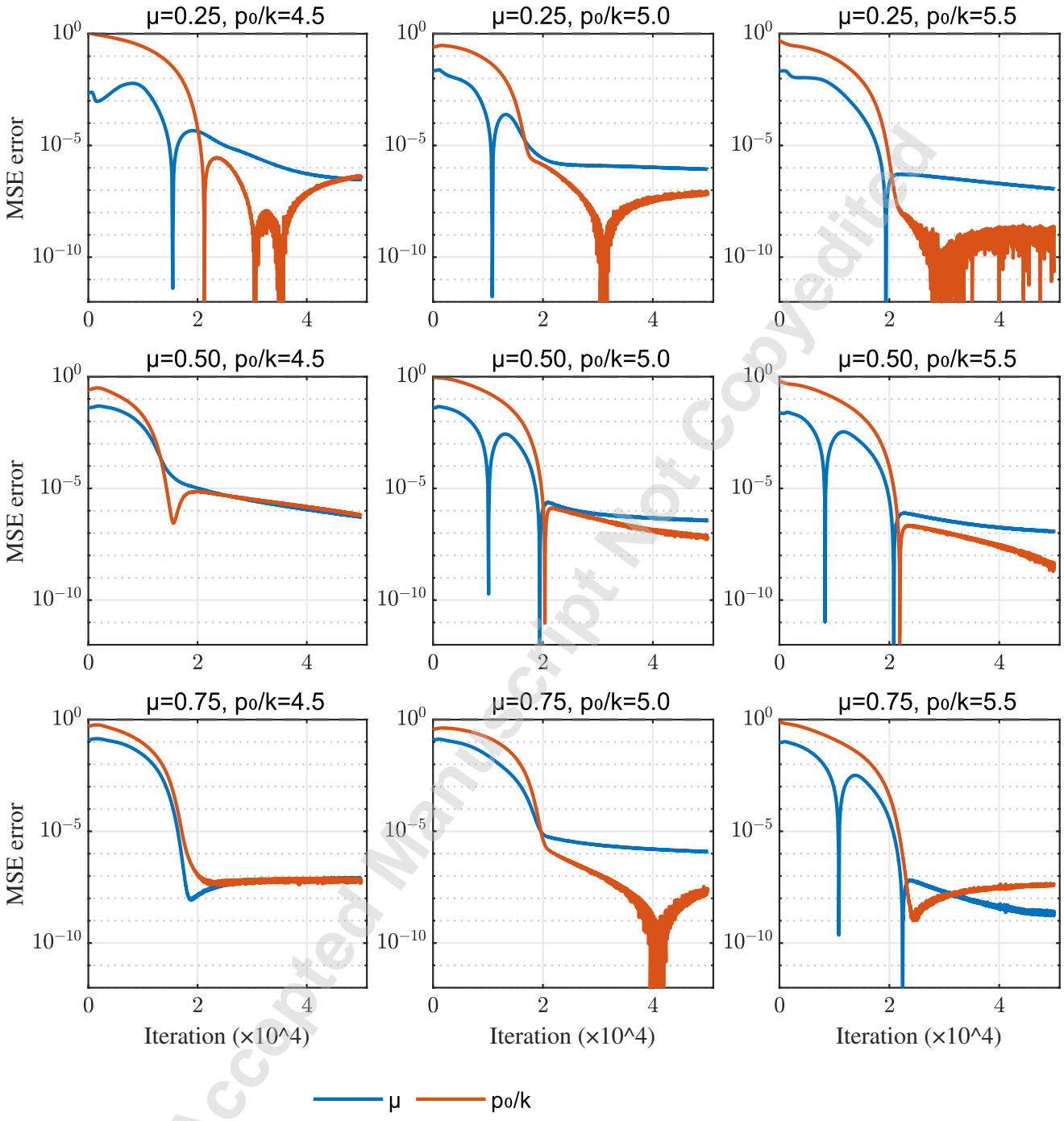


Fig. 14 Convergence of the Mean Squared Error (MSE) for the friction coefficient (μ, blue) and normalized load factor (p_0/k, orange) estimates. On a log-scale y–axis, the MSE decreases from $\mathcal{O}(10^2)$ to approximately 10^{-10} within 40000 iterations (iterations scaled by 10^4), demonstrating consistent error reduction across all parameter combinations.

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

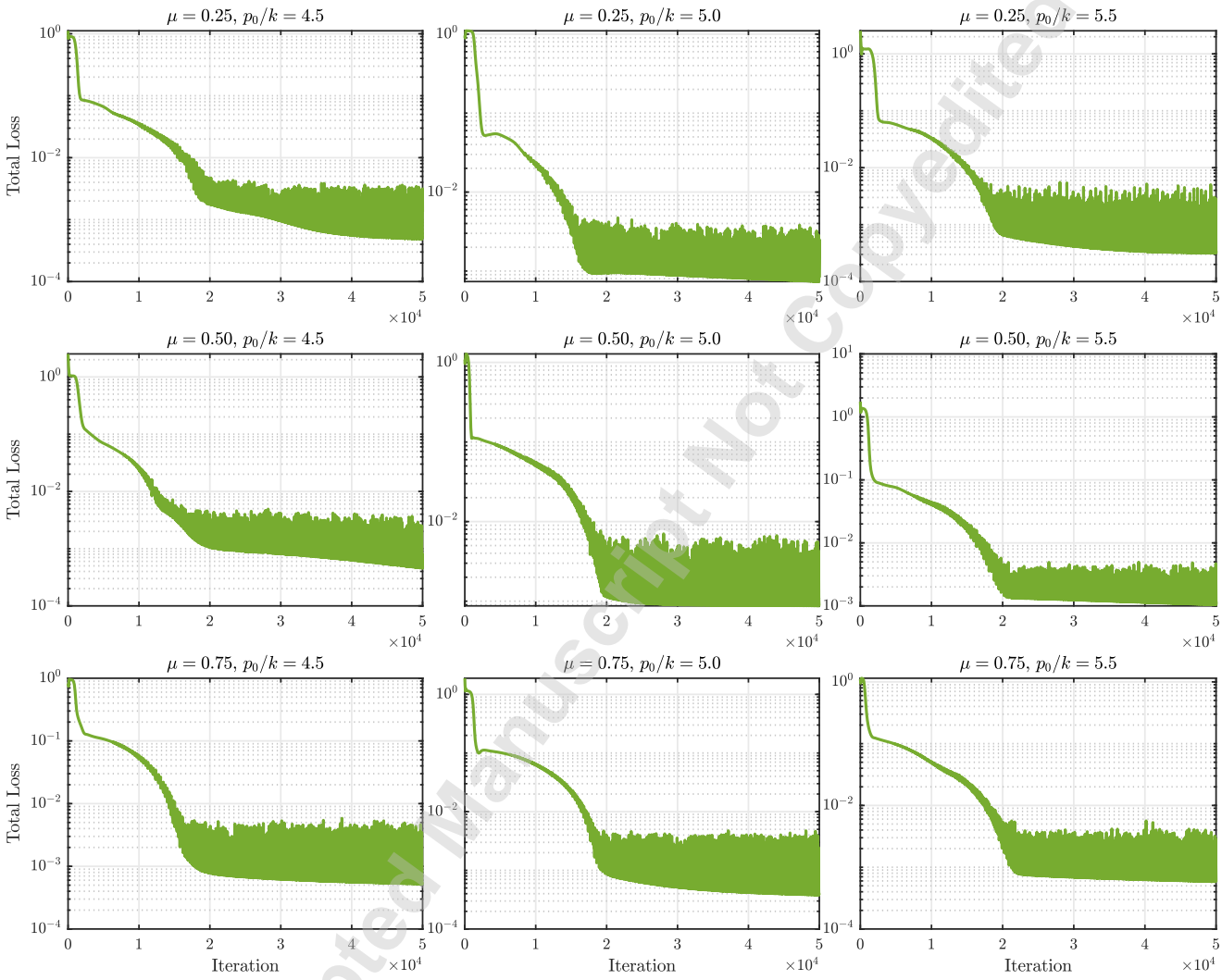


Fig. 15 Total loss function convergence during PINN training. The consistent reduction across all parameter combinations demonstrates optimization stability.

In the inverse framework, the same network architecture simultaneously recovers the friction coefficient μ and normalized load factor p_0/k to within 1% of their true values using only a few subsurface measurements. This robustness underscores the value of embedding the governing physics directly into the training objective: the physics-informed loss regularizes the inverse problem, eliminating the non-uniqueness and instability that often plague classical inversion schemes.

We also compared optimization strategies and observed that while the Adam optimizer provides a fast initial descent, augmenting it with L-BFGS-B improves final accuracy—albeit at a higher computational cost. Due to this trade-off, we rely on Adam, with carefully tuned learning rates and decay schedules, to strike a balance between accuracy and efficiency when solving stiff, path-dependent elastoplastic problems.

Overall, the physics-informed approach excels at three challenges in elastoplastic contact analysis:

- *Stiff plastic regions*, where conventional solvers demand very fine meshes to capture localized yielding.
- *Strain-rate discontinuities*, as the material transitions in and out of yield, creating sharp gradients in $\dot{\varepsilon}_{ij}$.
- *Incomplete constitutive physics*, since higher-order effects— e.g. kinematic-hardening or rate-dependent behavior— lie outside the classical Prandtl–Reuss model.

By enforcing the full set of elastoplastic equations via automatic differentiation, the PINN naturally handles these features without remeshing or explicit boundary tracking. Its mesh-free nature and continuous stress representation make it particularly well suited to problems involving complex contact geometries, path-dependent loading, or multi-axial stress states, where traditional FEM would require costly mesh adaptation.

In summary, our results establish that PINNs can deliver both high accuracy and computational efficiency for residual-stress prediction and parameter inversion in contact mechanics, opening the door to real-time, data-driven analysis of manufacturing and tribological processes.

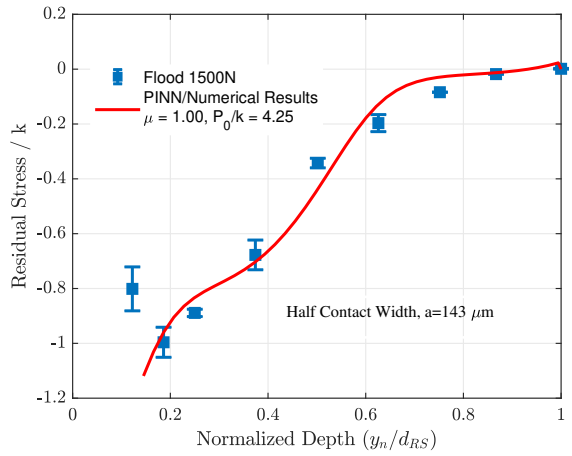
Although the present study uses semi-analytical and RK based solutions for validation, it is important to relate the proposed PINN framework to finite element methods, which remain the standard tool in industrial practice. FEM provides substantial flexibility in capturing complex tool geometries, material behavior, and near-surface deformation. However, high-fidelity 3D residual stress simulations in machining typically involve very fine meshes, multiple interacting thermo-mechanical loadings, and repeated passes, which lead to significant computational expense and long runtimes [36]. As demonstrated in the recent 3D reaming model of Leveille et al. (2024) [36], realistic simulations require multi-scale thermo- mechanical modeling, calibration of equivalent loadings, and simulation of numerous tool passages, resulting in computational times that can be prohibitive in iterative or inverse-analysis settings.

In contrast, the subsurface quasi-elastic region relevant to residual stress evolution is captured with comparable fidelity by the semi-analytical PR formulation used here. Moreover, FEM lacks an intrinsic mechanism for inverse identification of contact parameters and typically requires repeated forward simulations within a computationally intensive optimization loop. PINNs offer a complementary capability by embedding the governing equations directly into the training process, enabling rapid forward queries and direct inversion for $(\mu, p_0/k)$ without mesh dependence. This distinction highlights the role of PINNs as an efficient and flexible surrogate rather than a replacement for full FEM models in complex machining simulations.

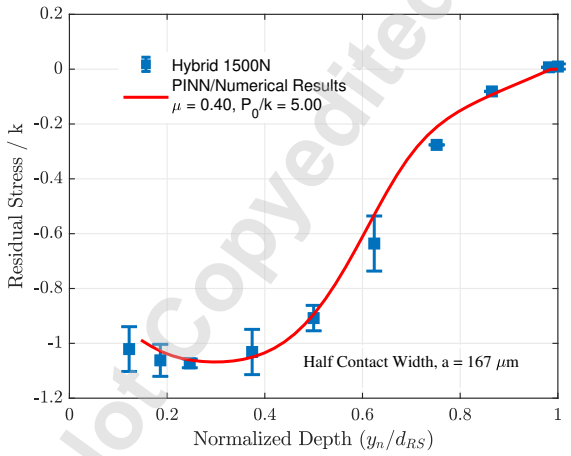
5 Model Validation with Experimental Data

To assess the predictive accuracy and physical consistency of the proposed Physics-Informed Neural Network (PINN) framework, the modeled residual stress (RS) distributions were compared with representative experimental data from two independent sources: (i) burnishing-induced RS measurements in Ti–6Al–4V under various cooling and lubrication conditions reported by Caudill [34], and (ii) orthogonal cutting and ball end milling data for Ti–6Al–4V presented by Nespor [35]. This comparison demonstrates the applicability of the framework to both chip-generating (cutting and abrasion) and chipless (sliding and ploughing) surface enhancement processes dominated by mechanical effects.

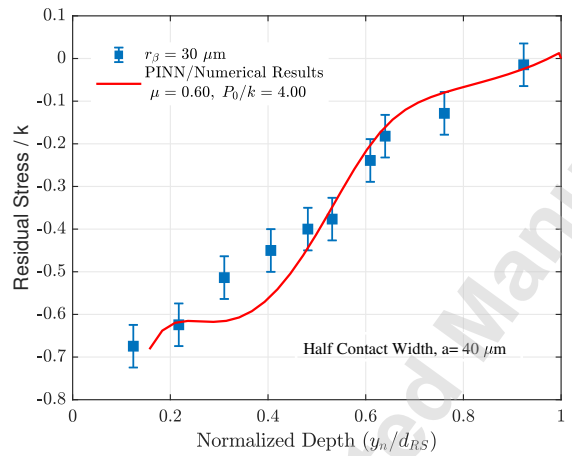
Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741



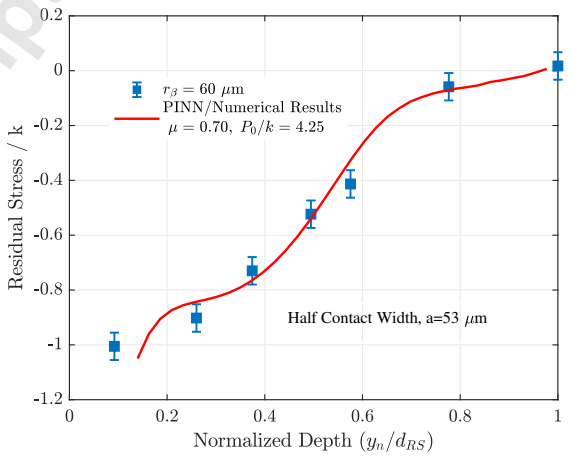
(*a*) Flood-cooled burnishing, 1500 N [34]



(*b*) Hybrid (LN_2+MQL) burnishing, 1500 N [34]



(*c*) Orthogonal cutting, r_β = 30 μm [35]



(*d*) Ball end milling, r_β = 60 μm [35]

Fig. 16 Comparison between normalized experimental data (blue markers with error bars) and PINN/Numerical predictions (red lines) for Ti–6Al–4V under different mechanically dominated surface treatments.

5.1 Data Normalization and Parameter Selection. For consistent comparison across different experimental configurations, both the spatial coordinate and the residual stress (RS) magnitude were expressed in dimensionless form. The depth coordinate was normalized as y_n/d_{RS}, where d_{RS} denotes the maximum depth at which measurable RS exists, and the RS magnitude was normalized by the material yield strength k. This normalization enables direct comparison of RS profiles independent of applied load or tool geometry.

In addition to the normalization depths, the key experimental parameters—feed rate and cutting speed—used in each configuration are summarized in Table 3. The parameter d_{RS} corresponds to the maximum subsurface depth over which residual stresses were experimentally observed, while the listed feeds and speeds provide context for the respective machining or burnishing conditions.

By comparing the experimental and modeled residual-stress profiles, the equivalent contact width ($2a$) for each condition was calculated from the corresponding normalized depth d_{RS}. The relative scaling between the modeled and experimental d_{RS} values provides a direct estimate of $2a$. The results show that the hybrid burnishing case yields a slightly larger contact width than flood cooling, consistent with its deeper compressive zone produced under improved lubrication and cooling. Similarly, in the machining cases, increasing the tool-edge radius from $r_\beta = 30$ μm (orthogonal cutting) to $r_\beta = 60$ μm (ball end milling) increases the effective contact width and subsurface pressure, shifting the compressive stress peak deeper below the surface. The calculated contact widths and corresponding depths derived from these comparisons establish a quantitative connection between the experimental RS distributions and the equivalent Hertzian parameters (P_0/k, μ, a) governing subsurface plasticity.

Table 3 Experimental parameters and normalization depths (d_{RS}) for the four conditions shown in Fig. 16.

Condition	Feed	Speed	d_{RS} (μm)
Flood [34]	0.05 mm/rev	50 m/min	346
Hybrid [34]	0.05 mm/rev	50 m/min	399
Orth. cut. [35]	0.05 mm	15 m/min	77
Ball mill. [35]	0.40 mm/tooth	40 m/min	114

The PINN-predicted profiles were generated by varying the normalized load factor (P_0/k) and friction coefficient (μ) in discrete increments of 0.25 and 0.05, respectively. The optimal combination was selected based on minimal deviation from the experimental data.

5.2 Comparison with Experimental Results. Figure 16 summarizes the comparison between normalized experimental measurements (blue markers with error bars) and the corresponding PINN/numerical predictions (solid red lines). For the burnishing data from Caudill [34], experimental uncertainties were taken approximately from the reported values. For the orthogonal cutting and ball end milling data from Nespor [35], explicit uncertainty estimates were not available; thus, a representative error bound of $\pm 0.05k$ was applied to each point to reflect typical X-ray diffraction measurement variability. Subfigures (a)–(b) correspond to flood and hybrid (LN_2+MQL) burnishing at 1500 N, while subfigures (c)–(d) correspond to orthogonal cutting and ball end milling with tool radii $r_\beta = 30$ μm and $r_\beta = 60$ μm, respectively.

In all cases, the PINN predictions closely follow experimental trends, capturing both the compressive subsurface region and the near-surface relaxation. For the burnishing cases, the modeled profiles reproduce the experimental magnitude and depth of compression within approximately 10%. Similarly, the cutting and milling results exhibit strong agreement in shape and magnitude, confirming that the identified (μ, P_0/k) pairs realistically represent the underlying contact mechanics.

As shown in Figure 9, variations in (P_0/k) and (μ) distinctly influence the modeled RS profiles—higher normalized pressure increases the depth of compression, whereas higher friction (larger μ) enhances surface shear effects. This quantitative mapping between equivalent Hertzian parameters and RS shape allows the model to explain experimentally observed differences across process conditions. For example, the hybrid (LN_2+MQL) burnishing condition exhibits lower effective traction but higher normal pressure than flood cooling, implying improved cooling/lubrication raises local yield strength and deepens the compressive zone.

In the machining cases of Figure 16(c)–(d), increasing the cutting-edge radius from $r_\beta = 30$ μm to $r_\beta = 60$ μm results in a higher effective friction coefficient and load factor—i.e., both μ and (P_0/k) increase with tool radius. This enlarges the contact area and

amplifies subsurface pressure, shifting the compressive RS peak deeper while reducing near-surface shear. Nespor's experimental trends confirm that this evolution arises primarily from contact mechanics rather than thermal effects, consistent with our PINN-predicted parameter dependence.

Overall, these findings validate that the proposed PINN framework generalizes across multiple mechanically dominated finishing processes, bridging forward and inverse analyses of residual stress evolution in Ti–6Al–4V. Future work will include additional cutting and milling conditions to further expand the parameter space and ensure comprehensive experimental coverage.

6 Conclusion

In this work we have introduced a unified PINN framework that seamlessly combines forward simulation and inverse parameter identification for machining-induced residual stress prediction under a Hertzian contact model. By embedding the PR elastoplastic equations directly into the network's loss function and leveraging automatic differentiation, our mesh-free solver reproduces benchmark stress fields with pointwise errors below 8% and captures sharp stress gradients without any explicit discretizations.

Simultaneously, the inverse PINN formulation recovers key contact parameters—the effective friction coefficient μ and the normalized load factor p_0/k—to within 1% accuracy from only four subsurface measurements. This physics-constrained, data-enhanced approach overcomes the ill-posedness and high computational cost of classical inversion and finite-element methods, enabling robust, real-time parameter estimation.

Looking ahead, the proposed framework can be readily extended to more sophisticated constitutive descriptions (e.g., anisotropic or rate-dependent models such as Bergstrom–Boyce), to include thermo-mechanical coupling for high-speed machining, and to integrate multi-fidelity or transfer-learning strategies for rapid deployment across new materials and processes. Coupled with advanced optimization techniques—such as variational PINNs or multi-objective algorithms—this methodology promises further gains in convergence speed and computational efficiency.

The ability of the proposed PINN framework to accurately and quickly identify Hertzian contact mechanics model parameters from complex RS profiles is envisioned to enable surrogate modeling of process-induced surface integrity within a Digital Process Twin framework [37], with closed-loop feedback between as-manufacturing product quality (i.e., limited RS measurements) and relevant process physics (i.e., sub-surface plastic deformation). This approach would enable causal inference of in-process loading conditions and allow for pro-active model-informed adaptive interventions to achieve desirable surface integrity conditions in manufactured components.

7 Acknowledgments

This work was supported by the U.S. National Science Foundation, grant number 2143806, project title "CAREER: Thermomechanical Response and Fatigue Performance of Surface Layers Engineered by Finish Machining: In-situ Characterization and Digital Process Twin".

References

[1] Withers, P. J. and Bhadeshia, H. K. D. H., 2001, "Residual stress. Part 1 – Measurement techniques," Materials Science and Technology, **17**(4), pp. 355–365.

[2] Brinksmeier, E., Reese, S., Klink, A., Langenhorst, L., Lübben, T., Meinke, M., Meyer, D., Riemer, O., and Sölter, J., 2018, "Underlying Mechanisms for Developing Process Signatures in Manufacturing," Nanomanufacturing and Metrology, **1**(4), pp. 193–208.

[3] Schoop, J., Hasan, M. M., and Zannoun, H., 2022, "Physics-Informed and Data-Driven Prediction of Residual Stress in Three-Dimensional Machining," Experimental Mechanics, **62**(8), pp. 1461–1474.

[4] Li, Y., Gan, W., Huang, X., Zhang, Y., Zhou, W., Li, D., and Zeng, Y., 2023, "A quantitative study of machining induced residual stress and its effect on subsequent creep age forming of aluminium alloy panels," Journal of Materials Processing Technology, **321**, p. 118147.

[5] Sun, J., Guo, Y., and Altintas, Y., 2018, "Residual stresses in machining: A review of prediction methods," Journal of Materials Science, **53**(11), pp. 8378–8394.

[6] Hertz, H., 1881, "Über die Berührung fester elastischer Körper," Journal für die reine und angewandte Mathematik, **92**, pp. 156–171.

[7] Merwin, J. E. and Johnson, K. L., 1963, "An Analysis of Plastic Deformation in Rolling Contact," Proceedings of the Institution of Mechanical Engineers, **177**(1), pp. 676–690.

[8] Brinksmeier, E., Cammett, J., König, W., Leskovar, P., Peters, J., and Tönshoff, H., 1982, "Residual stress measurement and causes in machining processes," CIRP Annals, **31**(2), pp. 491–510.

[9] Liang, S. Y. and Su, J.-C., 2007, "Residual Stress Modeling in Orthogonal Machining," CIRP Annals, **56**(1), pp. 65–68.

[10] Huang, K. and Yang, W., 2016, "Analytical modeling of residual stress formation in workpiece material due to cutting," International Journal of Mechanical Sciences, **114**, pp. 21–34.

[11] Nedin, R. D. and Vatulyan, A. O., 2013, "Inverse problem of non-homogeneous residual stress identification in thin plates," International Journal of Solids and Structures, **50**(13), pp. 2107–2114.

[12] Tariq, A. and Deliktaş, B., 2025, "An Inverse Parameter Identification in Finite Element Problems Using Machine Learning-Aided Optimization Framework," Experimental Mechanics, **65**(3), pp. 325–349.

[13] Zhou, T. and Xu, J., 2024, "Deep learning for inverse problems in plasticity: A critical review," International Journal of Plasticity, **172**, p. 103858.

[14] Ghaboussi, J., Pecknold, D. A., Zhang, M., and Haj-Ali, R. M., 1998, "Knowledge-based modeling of material behavior with neural networks," Journal of Engineering Mechanics, **124**(1), pp. 120–123.

[15] Matthies, H., 2020, "Multilevel solvers for large-scale finite element analysis," Computational Mechanics, **65**(4), pp. 943–958.

[16] Schwarz, S. and Ramm, E., 2001, "Sensitivity analysis and optimization for non-linear structural response," Engineering Computations, **18**(3/4), pp. 610–641.

[17] Herrmann, L. and Kollmannsberger, S., 2024, "Deep learning in computational mechanics: a review," Computational Mechanics, **74**, pp. 281–331.

[18] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L., 2021, "Physics-informed machine learning," Nature Reviews Physics, **3**(6), pp. 422–440.

[19] Zhou, W. and Xu, Y. F., 2024, "Data-Guided Physics-Informed Neural Networks for Solving Inverse Problems in Partial Differential Equations," arXiv preprint, arXiv2407.10836

[20] Raissi, M., Perdikaris, P., and Karniadakis, G. E., 2019, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," Journal of Computational Physics, **378**, pp. 686–707.

[21] Depina, I., Jain, S., Mar Valsson, S., and Gotovac, H., 2021, "Application of physics-informed neural networks to inverse problems in unsaturated groundwater flow," Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, **16**(1), pp. 21–36.

[22] Bai, J., Lin, Z., Wang, Y., Wen, J., Liu, Y., Rabczuk, T., Gu, Y., and Feng, X.-Q., 2025, "Energy-based physics-informed neural network for frictionless contact problems under large deformation," arXiv preprint.

[23] Ma, T., Li, Z., and Xu, J., 2023, "Real-time prediction of temperature fields in reheating furnaces using physics-informed neural networks," arXiv, arXiv preprint.

[24] Haghighat, F., Raissi, M., Moure, A., Gomez, H., and Juanes, R., 2021, "A physics-informed deep learning framework for inversion and surrogate modeling in solid mechanics," Computer Methods in Applied Mechanics and Engineering, **379**, p. 113741.

[25] Haghighat, F., Abouali, S., and Vaziri, R., 2023, "Constitutive model characterization and discovery using physics-informed deep learning," Engineering Applications of Artificial Intelligence, **120**, p. 105828.

[26] Roy, A. and Guha, A., 2021, "A data-driven physics-constrained approach for solving von Mises plasticity using Physics-Informed Neural Networks," International Journal for Numerical Methods in Engineering, **122**, pp. 6681–6702.

[27] Sahin, T., von Danwitz, M., and Popp, A., 2024, "Solving Forward and Inverse Problems of Contact Mechanics using Physics-Informed Neural Networks," Advanced Modeling and Simulation in Engineering Sciences, **11**.

[28] Lazoglu, I., Ulutan, D., Alaca, B. E., Engin, S., and Kaftanoglu, B., 2008, "Enhanced modeling of residual stresses in machining," CIRP Annals, **57**(1), pp. 81–84.

[29] Yan, L., Yang, W., Jin, H., and Wang, Z., 2012, "Analytical modeling of the effect of the tool flank wear width on the residual stress distribution," Machining Science and Technology, **16**(2), pp. 265–286.

[30] Johnson, K. L., 1987, *Contact Mechanics*, Cambridge University Press, Cambridge, UK.

[31] McEwen, E., 1949, "Stresses in elastic cylinders in contact along a generatrix," Philosophical Magazine, **40**, pp. 454–459.

[32] McDowell, D. L., 1997, "An approximate algorithm for elastic-plastic two-dimensional rolling/sliding contact," Wear, **211**(1), pp. 237–246.

[33] Kingma, D. P. and Ba, J., 2015, "Adam: A Method for Stochastic Optimization," *Proceedings of the International Conference on Learning Representations*, San Diego, CA, 2015.

[34] Caudill, J. R., 2019, "Enhanced Surface Integrity with Thermally Stable Residual Stress Fields and Nanostructures in Cryogenic Processing of Titanium Alloy Ti-6Al-4V," Ph.D. thesis, University of Kentucky, Lexington, KY.

[35] Nespor, D., Denkena, B., Grove, T., and Böß, V., 2015, "Differences and similarities between the induced residual stresses after ball end milling and orthogonal cutting of Ti–6Al–4V," Journal of Materials Processing Technology, **226**, pp. 15–24.

[36] Leveille, T., Valiorgue, F., Dumas, M., Masciantonio, U., Brosse, A., Karaouni, H., and Rech, J., 2024, "3D numerical modelling of residual stresses induced by reaming," Journal of Manufacturing Processes, **113**, pp. 47–60.

[37] Clark, B., Schoop, J., and Poonawala, H., 2024, "Digital process twins: a modular approach for surface conditioning and process optimization," Production Engineering Research Development, **18**(1), pp. 367–380.

Journal of Tribology. Received September 02, 2025;
Accepted manuscript posted December 26, 2025. doi:10.1115/1.4070741

List of Figures