ENGINEERING High-Performance Manufacturing
https://doi.org/10.2738/ENGHPM.2026.0002

Original Research

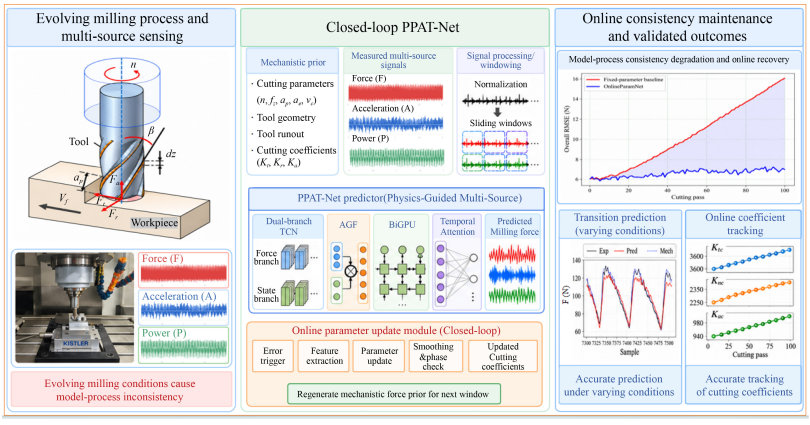
Closed-Loop Mitigation of Process Consistency Degradation in Evolving Milling Processes

Ping Deng[1], Ruping Song[1], Tianran Liu[1], Qingyu Chi[1], Qile Bo[1], Shuaifei Jia[1], Xu Li[1], Te Li[1], Yongqing Wang[1,2], Haibo Liu (✉)[1]

1 State Key Laboratory of High-performance Precision Manufacturing, Dalian University of Technology, Dalian 116024, China
2 Intelligent Manufacturing Longcheng Laboratory, Changzhou 213164, China

© The Author(s) 2026. This article is published by Higher Education Press.

GRAPHICAL ABSTRACT



HIGHLIGHTS

- Process feedback preserves mechanistic model consistency as milling conditions evolve.
- A runout-aware force model separates geometry from cutting-force coefficients.
- Mechanistic force priors guide attention-based multistep prediction.
- Model-process discrepancies trigger constrained updates and prior regeneration.

ABSTRACT During milling, the evolution of tool condition, material response, and machine-tool dynamics progressively undermines the consistency of fixed-parameter mechanistic models with the actual process. Existing physics–data hybrid approaches typically rely on static mechanistic priors and lack an effective mechanism for preserving model consistency during process evolution. To address this limitation, this study proposes PPAT-Net, which updates the mechanistic representation through process feedback to maintain alignment with the evolving milling process. Milling force is used as an observable response to quantify model–process discrepancy and validate the proposed paradigm. A runout-aware, parameter-separable mechanistic model provides physically interpretable priors for process prediction, while feedback assesses model consistency and adapts key parameters under physical constraints. Over 100 consecutive cutting passes at fixed nominal parameters, the fixed-model error increased from 5.82 to 16.16 N, whereas the proposed adaptation maintained it near 6.8 N. In a continuous three-stage varying-condition experiment, the framework remained stable across operating transitions and accommodated substantial parameter shifts. The results demonstrate that the proposed approach mitigates model-consistency degradation while improving prediction accuracy and cross-condition generalization without compromising physical interpretability. It therefore provides a methodological foundation for adaptive mechanistic modeling of nonstationary manufacturing processes.

KEYWORDS Consistency-preserving mechanistic modeling; Evolving milling processes; Process-feedback adaptation; Mechanistic prior regeneration

Received 22 Jul 2026; accepted 9 Oct 2026

E-mail: hbliu@dlut.edu.cn (Liu H.).

1 Introduction

Machining is an inherently dissipative and continuously evolving physical process. As material removal proceeds, tool wear progressively modifies the effective cutting-edge geometry, while thermal deformation and dynamic settling of the spindle may alter cutter runout [1,2]. Material heterogeneity can introduce stochastic fluctuations in the cutting response [3]. Thermally induced changes in spindle-bearing stiffness further modify the mechanical state of the machine tool [4]. These evolving factors cause the actual process state to deviate gradually from that represented during model calibration. Consequently, even an initially accurate physics-based model may lose validity as the tool, workpiece, and machine-tool states evolve. This divergence originates from the evolution of the modeled system rather than solely from calibration error and is referred to herein as process consistency degradation.

Process consistency degradation is not unique to milling; however, milling provides a highly observable and representative context in which to investigate this phenomenon. Cutting force, as a measurable response of tool-workpiece interaction, is jointly governed by tool radial runout [5], instantaneous undeformed chip thickness [6], material removal mechanisms [7], and structural vibrations [8]. As the states of the tool, workpiece, and machine tool evolve, cutting-force coefficients and other key parameters in the mechanistic model may drift, progressively reducing the validity of the initially calibrated model [9]. The intermittent nature of milling enables tooth-specific load variations to be clearly resolved, while complementary process signals provide an observational basis for characterizing the machining state and assessing model validity [10,11]. Cutting force and related process signals have therefore been widely used for process modeling, condition monitoring, and model validation [12]. Research on the physical representation, state prediction, and sustained validity of machining processes has primarily followed three approaches: mechanistic modeling, data-driven modeling, and physics–data hybrid modeling.

Mechanistic modeling describes tool–workpiece interactions through explicit physical relationships and has evolved from computationally intensive numerical simulation to efficient analytical formulations. Finite element methods implemented in platforms such as ABAQUS and ANSYS can capture cutting forces, deformation, temperature, and residual stress, but their high computational cost limits online prediction and real-time updating [13,14]. Analytical models, exemplified by the Altintas framework, express three-dimensional cutting forces as functions of force coefficients, tool geometry, engagement conditions, and instantaneous chip thickness, thereby retaining physical interpretability with substantially improved efficiency [1]. Subsequent studies have extended mechanistic modeling in several distinct directions. Wan et al. [15] decomposed the instantaneous milling force into a runout-independent nominal component and a runout-induced perturbation, enabling the force coefficients and cutter-runout parameters to be identified separately within a unified formulation for general end mills. Grossi et al. [16] investigated cutting-force coefficients over a broad range of spindle speeds, revealing their pronounced speed dependence. Li et al. [17] developed a structure-dependent cutting-force model that incorporated the continuously varying cutter entry and exit angles induced by the relative motion between the cutter and the workpiece. At the microscale, Wan et al. [18] derived a material-separation criterion from a theoretically determined dead-metal zone and formulated a force model that distinguishes the contributions of shearing and ploughing. Collectively, these studies have extended mechanistic models to account for cutter eccentricity, speed-dependent coefficients, material-specific structural engagement, and microscale separation mechanisms. However, their parameters and physical relationships are generally identified offline or predefined for specific machining conditions. As the physical conditions of the cutting process evolve, the calibrated cutting-force coefficients and associated mechanistic relationships may gradually deviate from the actual process behavior [19].

Data-driven modeling learns the relationship between process observations and target responses directly from sensing data, thereby reducing reliance on explicit descriptions of cutting mechanisms. Early studies combined conventional learning algorithms with simulation-based information. Charalampous [20] evaluated multiple regression and ensemble-learning methods for milling-force prediction using finite-element-derived data. Ni et al. [12] transferred a model pretrained on simulated samples to experimental conditions, reducing the dependence on large labeled datasets for each new operating state. With the development of deep learning, research on temporal modeling has increasingly focused on automatic feature extraction and temporal-state representation. Cheng et al. [21] combined parallel convolutional networks, attention, and bidirectional recurrent modeling for tool-wear monitoring and multistep prediction. Huang et al. [22] fused multidomain signal features within a deep convolutional architecture to identify tool-wear states without relying on handcrafted feature design. Broader reviews by Lei et al. [23] and Matsumura et al. [5] further summarized the advantages of machine learning in

nonlinear feature learning, temporal modeling, and condition recognition across mechanical diagnostics and machining monitoring. Although data-driven models can capture complex nonlinear relationships from multisource process signals, their performance depends strongly on the coverage of the training data. Distribution shifts caused by changes in operating conditions, tool states, or material properties can significantly degrade generalization [24]. Moreover, their black-box nature limits access to physically meaningful internal states. These limitations motivate physics–data hybrid modeling, which combines nonlinear representation capability with physical interpretability and constraints [25].

Physics–data hybrid modeling attempts to overcome the complementary limitations of mechanistic and data-driven approaches by embedding physical knowledge into learning-based prediction. Ni et al. [12] introduced mechanistic cutting-force information into a hybrid force-prediction framework, while Xu et al. [26] incorporated tool-workpiece engagement relationships into a neuro-physical architecture. Cheng et al. [27] embedded friction dynamics into recurrent temporal modeling. Zha et al. [28], Zhu and Li [29], and Zhu et al. [30] further introduced analytical force relations, vibration dynamics, or tool-wear knowledge through model coupling and physics-based constraints. For long-horizon prediction, Liu et al. [31] coordinated multiple neural modules with specialized temporal functions, whereas Liu et al. [32] combined non-autoregressive generation with physical-residual constraints. Ra et al. [33] and Willard et al. [34] further established the broader methodological foundations of physics-informed learning in manufacturing and engineering systems. These studies demonstrate that physical knowledge can improve prediction accuracy, generalization, and interpretability. However, the mechanistic information is generally calibrated or prescribed in advance and remains fixed during inference, leaving its progressive loss of validity under process evolution unresolved.

To overcome the aforementioned limitations, this study proposes PPAT-Net, a feedback-driven framework that integrates mechanistic knowledge with dynamic manufacturing data streams to maintain an accurate representation of the evolving milling process. Cutting-force prediction is employed as a representative validation task. By combining multisource process observations with mechanistic priors, PPAT-Net performs future-response prediction and evaluates model–process consistency online. When the mechanistic representation becomes invalid, a lightweight updater corrects key parameters and regenerates the corresponding priors. This closed-loop mechanism enables the mechanistic model to evolve with the machining process and supports performance assessment and adaptive control in high-performance manufacturing.

The main contributions of this study are summarized as follows.

(1) A feedback-driven framework is developed to maintain the validity of mechanistic representations during process evolution by coupling online consistency assessment, parameter adaptation, and prior regeneration.

(2) A lightweight incremental updating method is proposed to correct key mechanistic parameters without repeated full-model identification. Discrepancy-triggered activation and physically constrained acceptance criteria ensure stable and interpretable adaptation.

(3) The framework is validated under unseen parameter combinations, continuous operating-condition transitions, and long-duration fixed-condition machining. The results demonstrate improvements in both predictive performance and long-term mechanistic consistency, while clarifying the computational cost and current limitations for real-time implementation.

The remainder of this paper is organized as follows. Section 2 develops the runout-aware, parameter-separable mechanistic model and its two-stage calibration procedure. Section 3 presents the PPAT-Net framework, including prior-guided force prediction, online parameter adaptation, and feedback-based prior regeneration. Section 4 describes the experimental setup and evaluates mechanistic-model validity, predictive performance, closed-loop adaptation, architectural contributions, and computational feasibility. Section 5 concludes the study by summarizing the main findings, engineering implications, and current limitations.

2 Mechanistic Prior Construction and Initial Calibration

2.1 Runout-aware parameter-separable mechanistic model

A runout-aware mechanistic milling-force model is employed to generate the physical prior for PPAT-Net. Under peripheral milling conditions, the instantaneous undeformed chip thickness is governed by the feed per tooth, the angular position of each cutting edge, the tool–workpiece engagement state, and cutter runout. For a helical end mill, the angular position $\phi_j(t,\zeta)$ of the j-th tooth cutting edge varies along the axial coordinate ξ because of the helix-induced phase lag, while the engagement state $g_j(t,\zeta)$ is determined by the entry and exit angles. These geometric quantities are calculated from the cutter geometry and machining parameters following the conventional milling formulation. The corresponding tool–workpiece engagement geometry is illustrated in Fig. 1.

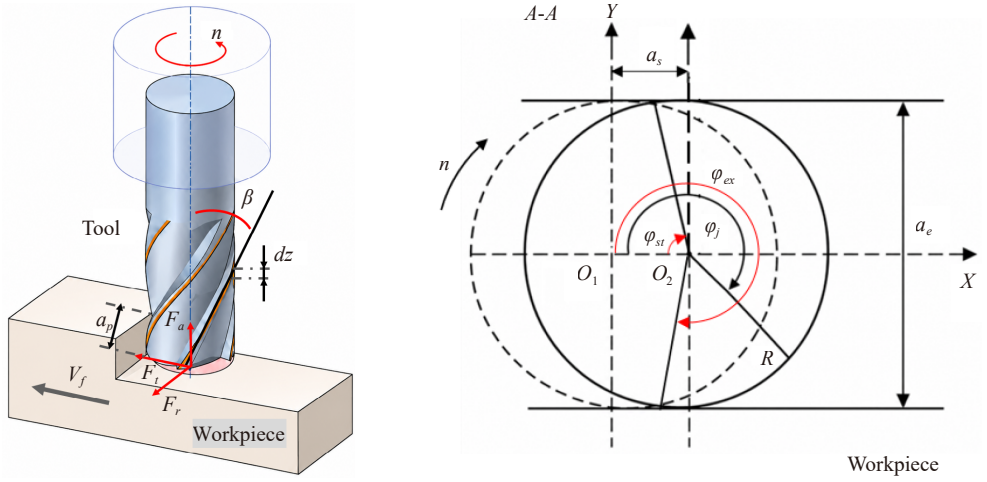


Fig. 1 Schematic diagram of end milling.

In practical milling, tool manufacturing errors, holder-clamping eccentricity, and spindle rotational errors cause the actual tool center to deviate from the nominal rotation axis. The resulting radial runout changes the effective cutting radius of each tooth and consequently alters the material removed by adjacent teeth, as shown in Fig. 2.

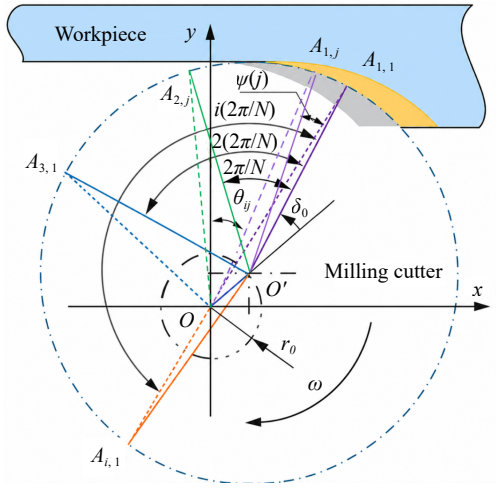


Fig. 2 Schematic diagram of a tool runout model.

Let ρ denote the runout magnitude and λ the runout orientation angle. The signed radial deviation of the j-th cutting edge at axial position ξ, together with the corresponding runout-induced chip-thickness correction, is written as:

$$\begin{aligned}\delta r_j(\zeta) &= \rho\cos\left[\lambda-\frac{\zeta\tan\beta}{R}-(j-1)\phi_p\right],\\ \Delta h_j^r(\zeta) &= \delta r_{j-1}(\zeta)-\delta r_j(\zeta)\quad \delta r_0=\delta r_{N_t}\\ &= -2\rho\sin\left(\frac{\pi}{N_t}\right)\sin\left[\lambda-\frac{\zeta\tan\beta}{R}-\frac{(2j-3)\pi}{N_t}\right].\end{aligned}\tag{1}$$

The first expression in Eq. (1) describes the tooth-dependent radial deviation caused by eccentricity. The second expression shows that the chip-thickness correction is determined by the difference between the effective radii of two consecutive teeth. Incorporating the runout-induced correction gives the unconstrained chip thickness $h_j^*(t,\zeta)$. Considering both the cutter-engagement state and the requirement for a positive chip thickness, the final runout-aware instantaneous undeformed chip thickness is defined as :

$$h_j(t,\zeta)=g_j(t,\zeta)\max\left[h_j^*(t,\zeta),0\right].\tag{2}$$

A zero value of $h_j(t,\zeta)$ indicates either that the cutting-edge element is outside the engagement region or that the chip thickness of the corresponding tooth vanishes because of radial runout. This formulation provides the geometric input to the cutting-force model.

Using the runout-aware chip thickness, a tooth-resolved differential force model is formulated by combining shearing and edge-force components. The former represents material shearing and plastic deformation, whereas the latter accounts for ploughing and edge-related contact effects. Let $\mathrm{d}F_{t,j}(t,\zeta)$, $\mathrm{d}F_{n,j}(t,\zeta)$, and $\mathrm{d}F_{a,j}(t,\zeta)$ denote the tangential, radial, and axial differential forces acting on the j-th cutting edge at axial coordinate ξ, respectively. The local differential cutting forces are expressed as:

$$\begin{cases}\mathrm{d}F_{t,j}(t,\zeta)=\left[K_{\mathrm{tc}}h_j(t,\zeta)+K_{\mathrm{te}}g_j(t,\zeta)\right]\mathrm{d}\zeta\\ \mathrm{d}F_{r,j}(t,\zeta)=\left[K_{\mathrm{nc}}h_j(t,\zeta)+K_{\mathrm{re}}g_j(t,\zeta)\right]\mathrm{d}\zeta\\ \mathrm{d}F_{a,j}(t,\zeta)=\left[K_{\mathrm{ac}}h_j(t,\zeta)+K_{\mathrm{ae}}g_j(t,\zeta)\right]\mathrm{d}\zeta\end{cases},\tag{3}$$

where K_{tc}, K_{nc} and K_{ac} are the tangential, radial, and axial shearing-force coefficients, respectively, and K_{te}, K_{ne} and K_{ae} are the corresponding edge-force coefficients. The shearing-force coefficients have units of N/mm^2, whereas

the edge-force coefficients have units of N/mm.

The six coefficients are collected into the equivalent cutting-force coefficient vector.

$$\boldsymbol{\theta} = [K_{\mathrm{tc}}, K_{\mathrm{nc}}, K_{\mathrm{ac}}, K_{\mathrm{te}}, K_{\mathrm{ne}}, K_{\mathrm{ae}}]^{\mathrm{T}}. \tag{4}$$

The proposed model separates the cutting-force coefficients from the geometric and kinematic terms. At each sampling instant, state variables, such as the angular position and engagement condition, are considered independently of the coefficient values. The resulting triaxial force can therefore be expressed in the compact form:

$$\begin{bmatrix} F_x^m(t) \\ F_y^m(t) \\ F_z^m(t) \end{bmatrix} = \boldsymbol{A}(t)\boldsymbol{\theta}. \tag{5}$$

The superscript m indicates a force generated by the mechanistic model. The matrix $\boldsymbol{A}(t) \in \mathbb{R}^{3\times 6}$ contains all coefficient-independent information at time t. The design matrix is constructed directly from the differential-force model:

$$\boldsymbol{A}(t) = \sum_{j=1}^{N_t} \int_0^{a_p} \boldsymbol{T}(\phi_j(t,\zeta)) \begin{bmatrix} h_j(t,\zeta) & 0 & 0 & g_j(t,\zeta) & 0 & 0 \\ 0 & h_j(t,\zeta) & 0 & 0 & g_j(t,\zeta) & 0 \\ 0 & 0 & h_j(t,\zeta) & 0 & 0 & g_j(t,\zeta) \end{bmatrix} \mathrm{d}\zeta. \tag{6}$$

The first three columns of $\boldsymbol{A}(t)$describe the contributions of the shearing-force coefficients, and the last three columns describe those of the edge-force coefficients. In this form, the effects associated with tool engagement and radial runout remain embedded in the design matrix, while the coefficients that may evolve during machining remain isolated in $\boldsymbol{\theta}$. This parameter-separable formulation preserves linearity with respect to the coefficient vector, enabling efficient coefficient identification, online updating, and mechanistic-prior regeneration without reconstructing the geometric model.

2.2 Two-stage calibration

In the runout-aware mechanistic model, the cutting-force coefficients govern the overall load level and directional distribution, whereas radial runout determines tooth-specific load variations. Joint identification from instantaneous force signals can therefore cause strong parameter coupling. A phase-synchronized two-stage calibration procedure is adopted to decouple these effects.

In the first stage, the force components of all teeth are aligned at corresponding angular positions and averaged over a complete tooth set. Let ϕ_m denote the m-th discrete phase point within one tooth-passing period. The superscript ν denotes either the experimental force ($\nu = \exp$) or the mechanistic force ($\nu = m$). The tooth-synchronous average in direction q is calculated as:

$$\overline{F}_q^{\nu}(\phi_m) = \frac{1}{N_t} \sum_{j=1}^{N_t} F_{q,j}^{\nu}\left[\phi_m - (j-1)\phi_p\right], \quad q \in \{x,y,z\}, \nu \in \{\exp, m\} \tag{7}$$

The tooth-synchronous averaging operation is not introduced merely as signal smoothing. According to the runout model in Eq. (1), the chip-thickness correction of each tooth is determined by the difference between the radial deviations of two adjacent teeth. The corrections over a complete tooth set satisfy:

$$\sum_{j=1}^{N_t} \Delta h_j^r(\zeta) = \sum_{j=1}^{N_t} [\delta r_{j-1}(\zeta) - \delta r_j(\zeta)] = 0. \tag{8}$$

The same tooth-synchronous averaging operation is applied to the coefficient-independent terms of the design matrix established in Section 2.2. The resulting phase-averaged force samples in direction q are stacked into $\overline{\boldsymbol{f}}_q^{\exp}$, and the corresponding rows of the averaged design matrix are stacked into $\overline{\boldsymbol{A}}_q$. The first-stage design matrix is constructed from the common chip-load component, without requiring a prior estimate of the runout parameters.

The initial coefficient vector is obtained by solving a constrained multiaxis least-squares problem:

$$\boldsymbol{\theta}_0 = \arg\min_{\boldsymbol{\theta}\in\Theta} \sum_{q\in\{x,y,z\}} w_q \frac{\| \overline{\boldsymbol{f}}_q^{\exp} - \overline{\boldsymbol{A}}_q\boldsymbol{\theta}\|^2}{\| \overline{\boldsymbol{f}}_q^{\exp}\|^2 + \varepsilon}. \tag{9}$$

The feasible set θ imposes physically admissible bounds on the six coefficients. The positive weights w_q balance the contributions of the three force directions, and the small constant ε prevents numerical instability when the energy of one force component is low. Normalization by the measured-force energy also prevents the force direction with the largest amplitude from dominating the calibration.

In the second stage, the identified coefficient vector $\boldsymbol{\theta}_0$ is fixed, and the original tooth-resolved force signals are used to identify radial runout. Unlike the phase-averaged signal, the instantaneous force retains the inter-tooth peak differences required to recover the runout magnitude and orientation.

The runout orientation is defined relative to the angular reference of the cutter and is therefore coupled with the global phase offset between the measured and modeled forces. The radial-runout parameter vector $\boldsymbol{p} = [\rho, \lambda]^T$ and the phase offset s are consequently identified jointly. For

a candidate offset, $\mathcal{S}_s$ circularly shifts the mechanistic-force sequence by s sampling points. The joint calibration problem is expressed as:

$$(\boldsymbol{p}_0, s_0) = \underset{\boldsymbol{p}\in\mathcal{P}, s\in[-L,L]}{\arg\min} \sum_{q\in\{x,y,z\}} w_q \frac{\| \boldsymbol{f}_q^{\exp} - \mathcal{S}_s \boldsymbol{f}_q^{m}(\boldsymbol{\theta}_0, \boldsymbol{p})\|^2}{\| \boldsymbol{f}_q^{\exp}\|^2 + \varepsilon}. \quad (10)$$

The admissible set $\mathcal{P}$ constrains the runout magnitude to a nonnegative and physically reasonable range and treats the runout orientation as a periodic angular variable. The phase-search bound L limits the candidate displacement to a prescribed neighborhood of the nominal angular reference. Joint optimization prevents the runout orientation from absorbing an unknown acquisition-phase error.

The proposed procedure yields the initial coefficient vector $\boldsymbol{\theta}_0$, radial-runout parameters $\boldsymbol{p}_0$, and global phase offset s_0 required for mechanistic-prior generation. The tooth-synchronous first stage provides a stable estimate of the common force parameters, whereas the tooth-resolved second stage retains the inter-tooth information required for runout calibration. During subsequent online operation, the calibrated runout and phase parameters are regarded as quasi-static, while the coefficient vector is updated according to the observed model–process discrepancy described in Section 3.

3 PPAT-Net Framework

3.1 Framework overview

As shown in Fig. 3, PPAT-Net establishes a closed-loop modeling framework integrating mechanistic-prior generation, multisource process representation, consistency evaluation, and online parameter adaptation. The framework operates through a rolling-window mechanism for continuous prediction and model updating. Within each prediction window, the measured cutting forces from the preceding observation window are jointly encoded with the corresponding mechanistic responses. This representation captures both the current process state and the discrepancy between the model and the actual response. Acceleration and spindle-power signals are incorporated as auxiliary observations to characterize vibration, transient impacts, and load fluctuations that are not explicitly represented by the mechanistic model. The PPAT predictor then combines these multisource observations with the future cutting-force prior generated by the current mechanistic model to perform multistep force prediction. As the process advances, the measured forces within the prediction interval are acquired online and compared with the corresponding mechanistic responses to evaluate

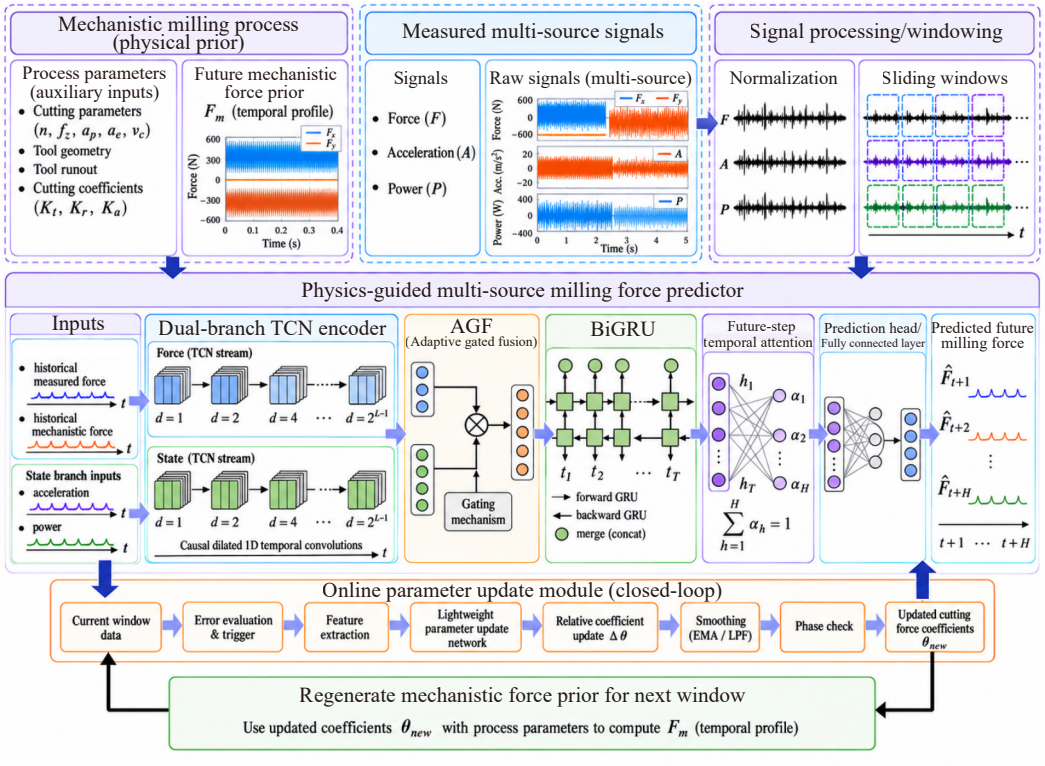


Fig. 3 Overall architecture of the proposed PPAT-Net.

model–process consistency. When the discrepancy exceeds a predefined criterion, the key mechanistic parameters are reidentified and fed back to update the model for subsequent prediction windows.

3.2 Process response prediction

As shown in Fig. 4, the Predictor comprises dual-stream temporal encoding, adaptive feature fusion, and mechanistic-prior-conditioned decoding. The force stream extracts discrepancy features from the historical measured and mechanistic forces, whereas the state stream captures complementary dynamic information from acceleration and spindle power.

Residual dilated temporal convolutions are used to extract force-related features $\boldsymbol{z}_\tau^F$ and state-related features $\boldsymbol{z}_\tau^S$ over multiple tooth-passing periods. Their contributions are adaptively weighted at each historical time step:

$$\gamma_\tau = \mathrm{softmax}(\boldsymbol{W}_g[\boldsymbol{z}_\tau^F \,\|\, \boldsymbol{z}_\tau^S] + \boldsymbol{b}_g), \boldsymbol{z}_\tau = \gamma_\tau^F \boldsymbol{z}_\tau^F + \gamma_\tau^S \boldsymbol{z}_\tau^S. \quad (11)$$

The two elements of γ_τ are nonnegative and satisfy $\gamma_\tau^F + \gamma_\tau^S = 1$. The fused sequence $\{z_\tau\}_{\tau=1}^H$ is subsequently processed by a bidirectional gated recurrent unit to construct the historical context sequence $\{h_\tau\}_{\tau=1}^H$.

For the k-th future step, the mechanistic force $\boldsymbol{F}_{t+k}^m$, the prediction-step embedding $\boldsymbol{e}_k$, and the operating-condition vector $\boldsymbol{c}_t$ are combined to form a future-step-specific query:

$$\boldsymbol{q}_k = \tanh\{\boldsymbol{W}_q[\boldsymbol{W}_m \boldsymbol{F}_{t+k}^m \,\|\, \boldsymbol{e}_k \,\|\, \boldsymbol{W}_c \boldsymbol{c}_t] + \boldsymbol{b}_q\}, k = 1, \cdots, P. \quad (12)$$

Linear projections of the historical context vectors form the key vectors $\boldsymbol{k}_\tau$ and value vectors $\boldsymbol{v}_\tau$. The query for each future step is used to retrieve the historical information most relevant to its expected mechanistic state:

$$\alpha_{k,\tau} = \frac{\exp(\boldsymbol{q}_k{}^{\mathrm{T}} \boldsymbol{k}_\tau / \sqrt{d})}{\sum_{r=1}^H \exp(\boldsymbol{q}_k{}^{\mathrm{T}} \boldsymbol{k}_r / \sqrt{d})}, \boldsymbol{r}_k = \sum_{\tau=1}^H \alpha_{k,\tau} \boldsymbol{v}_\tau. \quad (13)$$

The key dimension is denoted by d, and $\boldsymbol{r}_k$ represents the historical context retrieved for the k-th future point. Because the attention weights depend on the future

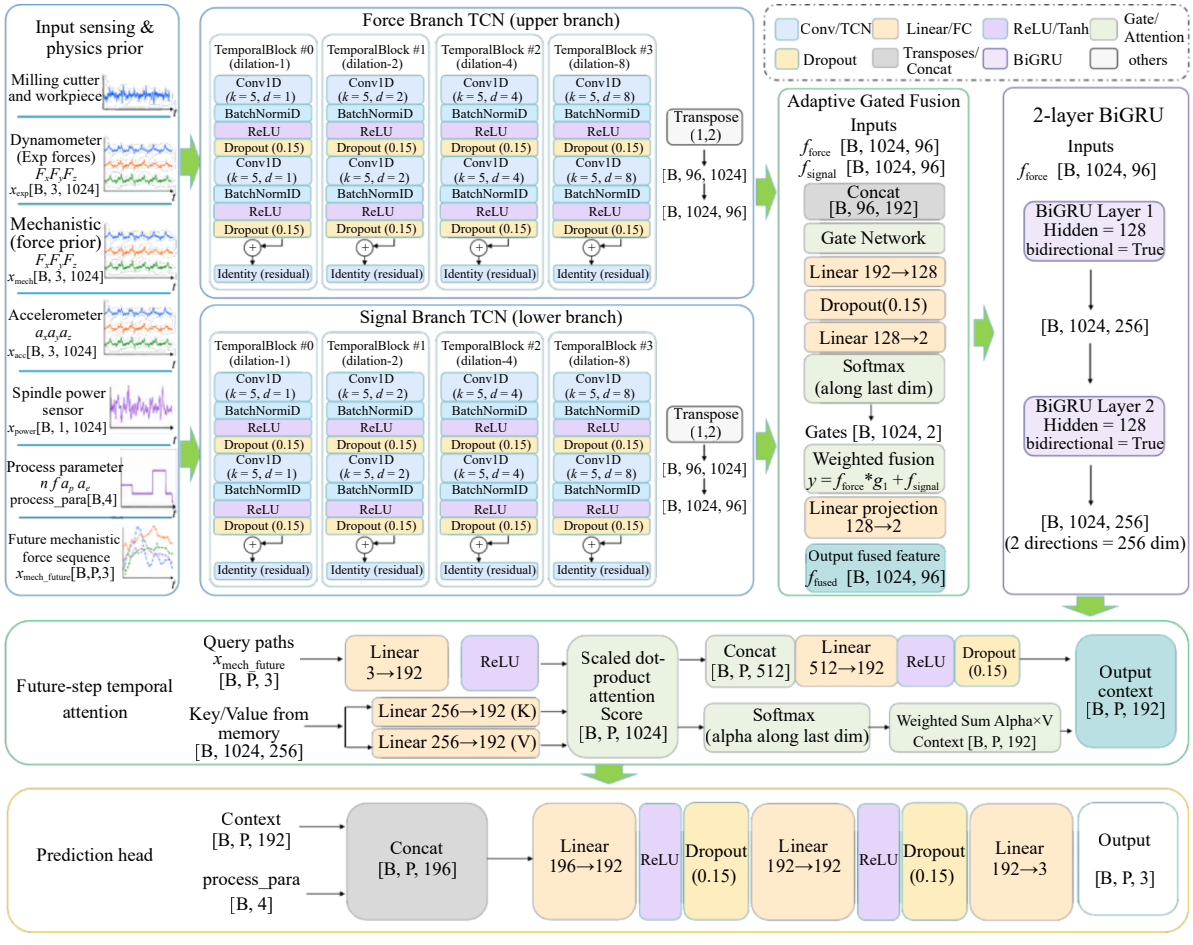


Fig. 4 Architecture of the PPAT Predictor with mechanistic-prior-conditioned future-step decoding.

mechanistic force, different future points can select different historical segments according to their expected tooth phase and nominal force level.

The retrieved context is finally combined with the mechanistic-force prior, the prediction-step embedding, and the operating-condition representation to generate the triaxial force prediction:

$$\hat{\boldsymbol{F}}_{t+k} = \mathrm{MLP}([\boldsymbol{r}_k \,\|\, \boldsymbol{W}_m \boldsymbol{F}^m_{t+k} \,\|\, \boldsymbol{e}_k \,\|\, \boldsymbol{W}_c \boldsymbol{c}_t]), k = 1, \cdots, P. \quad (14)$$

3.3 Online coefficient adaptation

For the k-th update window, the instantaneous design matrices in Eq. (9) are vertically stacked to form the window-level matrix $\boldsymbol{A}_k$, and the corresponding measured triaxial forces are stacked into $\boldsymbol{f}^{\mathrm{exp}}_k$. Owing to the parameter-separable relation in Eq. (5), reference coefficients for training the online updater can be generated offline through the following constrained regularized least-squares problem:

$$\boldsymbol{\theta}^{\mathrm{ref}}_k = \underset{\theta \in \Theta}{\mathrm{argmin}} \left[\| \boldsymbol{W}_f (\boldsymbol{f}^{\mathrm{exp}}_k - \boldsymbol{A}_k \theta) \|^2_2 + \lambda_\theta \| \boldsymbol{D}_k (\theta - \theta_{k-1}) \|^2_2 \right]. \quad (15)$$

The diagonal matrix $\boldsymbol{W}_f$ balances the contributions of the three force directions, and $\boldsymbol{D}_k$ normalizes coefficient changes according to their current magnitudes. The regularization weight λ_θ suppresses abrupt window-to-window variation, while the feasible set Θ imposes physically admissible coefficient bounds. Consequently, $\boldsymbol{\theta}^{\mathrm{ref}}_k$ provides a stable supervision target rather than an unconstrained re-identification result.

To avoid repeatedly solving the regularized problem, OnlineParamNet predicts the relative coefficient correction from a compact feature vector $\boldsymbol{x}_k$. The vector comprises measured and mechanistic force features, residual and normalized-error metrics, waveform correlation, acceleration and spindle-power statistics, operating conditions, and the current coefficient state. The candidate update is then generated and smoothed as:

$$\begin{aligned} \Delta\hat{\boldsymbol{\theta}}_k &= \mathcal{N}_\psi(\boldsymbol{x}_k), \\ \tilde{\boldsymbol{\theta}}_k &= \boldsymbol{\theta}_{k-1} \odot \left(1 + \Delta\hat{\boldsymbol{\theta}}_k\right), \\ \boldsymbol{\theta}^c_k &= \Pi_\Theta \left[(1-\eta)\boldsymbol{\theta}_{k-1} + \eta \tilde{\boldsymbol{\theta}}_k \right], \end{aligned} \quad (16)$$

where, $\mathcal{N}_\psi$ denotes OnlineParamNet, $\odot$ denotes element-wise multiplication, η is the smoothing factor, and Π_Θ projects the smoothed candidate onto the admissible coefficient set. Predicting a relative correction allows the updater to retain the information contained in the current physical state while limiting the effect of a single window.

A candidate update is generated only when the normalized model-process discrepancy $E_k = \|\boldsymbol{f}^{\mathrm{exp}}_k - \boldsymbol{f}^m_k\|_2 / (\|\boldsymbol{f}^{\mathrm{exp}}_k\|_2 + \varepsilon)$ exceeds the trigger threshold and the prescribed cooldown interval has elapsed. The candidate is then evaluated using the current buffer window. It is accepted only when it reduces the mechanistic-force error, retains sufficient waveform correlation, and does not introduce an abnormal phase shift within the prescribed search range. The accepted coefficient vector is therefore determined by:

$$\boldsymbol{\theta}_k = \begin{cases} \boldsymbol{\theta}^c_k & E_k > \tau_E, k - k_{\mathrm{last}} \geqslant C \\ & J^c_k \langle J^0_k, \rho_k \geqslant \rho_{\min}, \left| s^{\mathrm{opt}}_k \right| < L \\ \boldsymbol{\theta}_{k-1} & \text{otherwise} \end{cases}. \quad (17)$$

The quantities J^c_k and J^0_k denote the validation errors obtained with the candidate and current coefficient vectors, respectively. The maximum normalized correlation and its associated lag are denoted by ρ_k and s^{opt}_k. Once accepted, the updated coefficients are fed back into Eq. (8) to regenerate the mechanistic prior.

3.4 Training and inference

The PPAT Predictor and OnlineParamNet are trained separately because they address different sources of model–process discrepancy. OnlineParamNet corrects process-dependent drift represented by the six cutting-force coefficients, thereby realigning the mechanistic model with the current machining state. However, coefficient adaptation alone cannot capture unmodeled dynamics such as structural vibration, transient impacts, load fluctuations, and local material-response variations. The Predictor therefore uses multisource process observations together with the updated mechanistic prior to compensate for these residual dynamics and forecast the future force response. The Predictor is optimized using a composite loss that constrains force amplitude, waveform shape, curvature, and performance relative to the mechanistic prior:

$$\mathcal{L}_{\mathrm{pred}} = \mathcal{L}_{\mathrm{amp}} + \lambda_s \mathcal{L}_{\mathrm{shape}} + \lambda_c \mathcal{L}_{\mathrm{curv}} + \lambda_m \mathcal{L}_{\mathrm{prior}}. \quad (18)$$

where, $\mathcal{L}_{\mathrm{amp}}$ is a direction-weighted SmoothL1 loss. $\mathcal{L}_{\mathrm{shape}}$ and $\mathcal{L}_{\mathrm{curv}}$ constrain the first- and second-order temporal differences, respectively. $\mathcal{L}_{\mathrm{prior}}$ penalizes the Predictor only when its error exceeds that of the mechanistic prior. The weights λ_s, λ_c, and λ_m balance the auxiliary terms.

OnlineParamNet is trained to predict the relative correction between the current coefficient state and the regularized reference coefficients:

$$\mathcal{L}_{\mathrm{upd}} = \frac{1}{N_u} \sum_{k=1}^{N_u} \| \mathcal{N}_\psi(\boldsymbol{x}_k) - \frac{\boldsymbol{\theta}^{\mathrm{ref}}_k - \boldsymbol{\theta}_{k-1}}{\|\boldsymbol{\theta}_{k-1} + \varepsilon \boldsymbol{1}\|_2} \|^2_2. \quad (19)$$

The division is performed element-wise, with a small constant added for numerical stability. Because the targets are derived from Eq. (20), OnlineParamNet learns

physically bounded incremental corrections rather than unconstrained coefficient estimates.

Accordingly, the two networks are assigned distinct optimization objectives: OnlineParamNet performs physically constrained coefficient correction, whereas the Predictor compensates for residual dynamics beyond the representational capability of the mechanistic model and forecasts the future response. Separate training preserves this functional distinction and avoids the risk that joint optimization allows the Predictor to absorb parameter-induced discrepancies, which could weaken the identifiability and physical interpretability of the coefficient updates. During inference, validated coefficient corrections are fed back to regenerate the mechanistic prior, which is then combined with multisource observations by the Predictor to generate the subsequent force forecast; otherwise, the current coefficients are retained.

4 Results and Discussion

4.1 Experimental setup

The milling experiments were conducted on a VDLS850 vertical machining center equipped with a FANUC OI-MP. The workpiece material was No. 20 steel, and an 18mm-diameter, four-flute carbide end mill was used as the cutting tool. During the experiments, the cutting forces were measured by a Kistler 9129AA at a sampling frequency of 10 kHz. Meanwhile, a proprietary CNC machine tool health monitoring system was used to acquire acceleration signals and spindle power signals. The power signals were measured using an HGE99H81-50A power sensor, while the acceleration signals were acquired by TEDS Type 4535-B. The acquired force, acceleration, and spindle-power signals were temporally synchronized according to the acquisition timestamps and segmented according to the valid cutting intervals. No additional digital filtering, detrending, or normalization was applied before the signals were used as network inputs. This processing strategy was adopted to preserve the original transient fluctuations and vibration-related information contained in the measured signals, which provide complementary observations of the evolving machining state. The experimental setup is shown in Fig. 5.

To evaluate the model's predictive performance under different combinations of cutting parameters, this paper established a total of 15 sets of milling conditions. The 15 milling conditions were divided into training, validation, held-out test, and held-out-combination sets, as listed in Table 1. The validation set was used exclusively for early stopping, hyperparameter selection, and model selection, whereas the test and generalization sets were reserved for final evaluation under unseen conditions and held-out parameter combinations.

Following the framework in Section 3.1, a sliding-window strategy was adopted with a 1024 points observation window. Four independently trained prediction horizons of 50, 150, 300, and 600 points were considered, corresponding to 5, 15, 30, and 60 ms, respectively. The number of valid training samples is shown in Fig. 6.

The model was implemented in PyTorch 2.11 with CUDA 12.8 and optimized using AdamW. The initial learning rate, weight decay, batch size, and maximum number of epochs were set to 1×10^{-4}, 5×10^{-5}, 128, and 150, respectively. Early stopping with a patience of 20 and gradient clipping with a maximum norm of 1.0 were applied. The directional loss weights for the F_x, F_y and

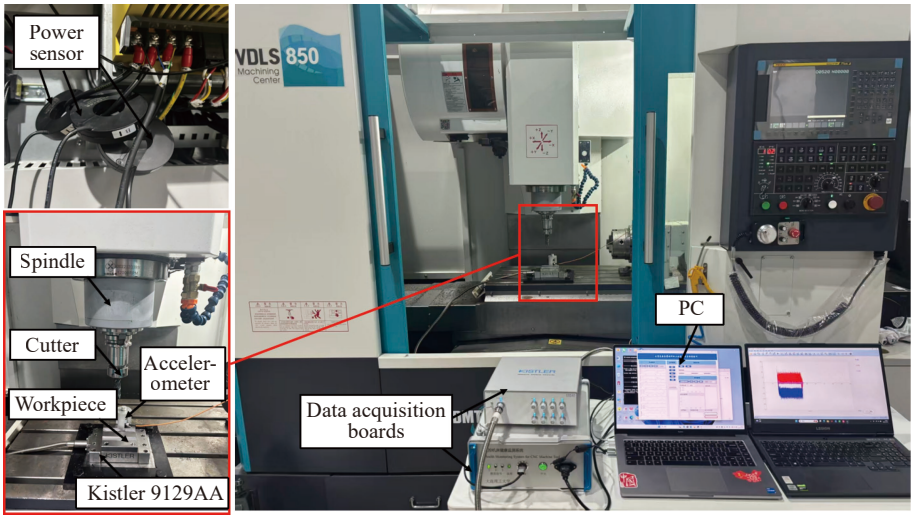


Fig. 5 Milling experimental platform and signal acquisition system.

Table 1 Milling test conditions and dataset segmentation.

Group	Spindle speed S (r/min)	Feed f (mm/min)	Depth of cut ap (mm)	Feed per tooth f_z (mm/tooth)	Dataset partition
1	1700	200	0.4	0.0294	Train
2	1700	200	0.6	0.0294	Validation
3	1700	200	1.0	0.0294	Train
4	2000	200	0.4	0.0250	Train
5	2000	280	0.4	0.0350	Test
6	2000	360	0.4	0.0450	Train
7	2000	200	0.6	0.0250	Train
8	2000	280	0.6	0.0350	Validation
9	2000	360	0.6	0.0450	Generalization
10	2000	200	1.0	0.0250	Train
11	2000	280	1.0	0.0350	Train
12	2000	360	1.0	0.0450	Train
13	2300	200	0.4	0.0217	Generalization
14	2300	200	0.6	0.0217	Train
15	2300	200	1.0	0.0217	Train

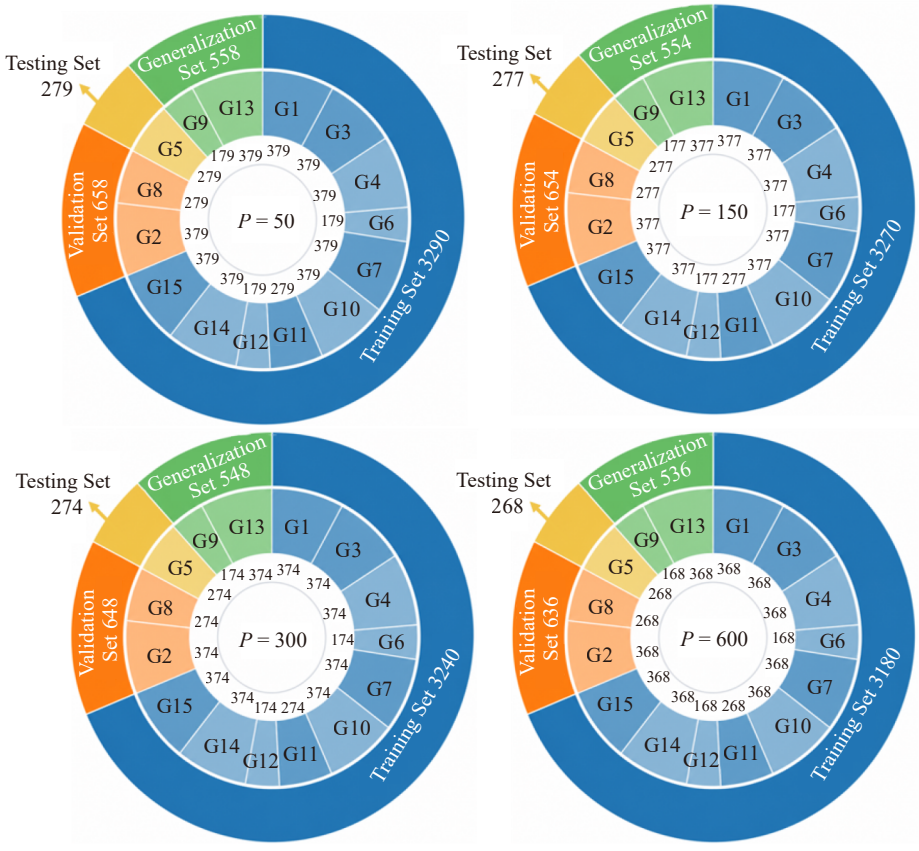


Fig. 6 Sample counts and data partitioning under four prediction windows.

F_z were set to 1.0, 2.5, and 2.5, respectively. Training and evaluation were conducted on a workstation equipped with NVIDIA GeForce RTX 5060 GPU and an Intel® Core™ i5-14600KF CPU operating at 3.50 GHz. The main architectural parameters are listed in Table 2.

Table 2 Model architecture parameters.

Parameter	Value
Network depth	4
Output channels	96
Kernel size	5
Dilation factors	1, 2, 4, 8
Dropout rate	0.15
Fusion hidden units	128
Number of recurrent layers	2
Hidden units per direction	128
Recurrent output dimension	256
Attention hidden dimension	192
Fully connected hidden units	192

All evaluation metrics were calculated after inverse normalization in the physical force domain. The root-mean-square error (RMSE) and coefficient of determination (R^2) were calculated separately for each force direction on the test set and the two unseen operating-condition combinations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i^{\mathrm{pred}} - y_i^{\mathrm{exp}}\right)^2}. \tag{20}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i^{\mathrm{exp}} - y_i^{\mathrm{pred}}\right)^2}{\sum_{i=1}^{n}\left(y_i^{\mathrm{exp}} - \bar{y}^{\mathrm{exp}}\right)^2}. \tag{21}$$

4.2 Mechanistic model evaluation

This section evaluates the matched-condition fidelity of the mechanistic model. Using the calibration procedure in Section 2, the six coefficients were identified for all 15 milling conditions to characterize their dependence on spindle speed, feed per tooth, and axial depth of cut.

Fig. 7 presents the measured coefficient distributions without interpolation because the experimental design is not a complete factorial scheme. The pronounced condition-dependent variations indicate that a coefficient vector calibrated at one operating point cannot represent the entire operating domain. Fig. 8 further shows distinct coefficient trends with axial depth of cut under different spindle speeds and feed rates, confirming that the

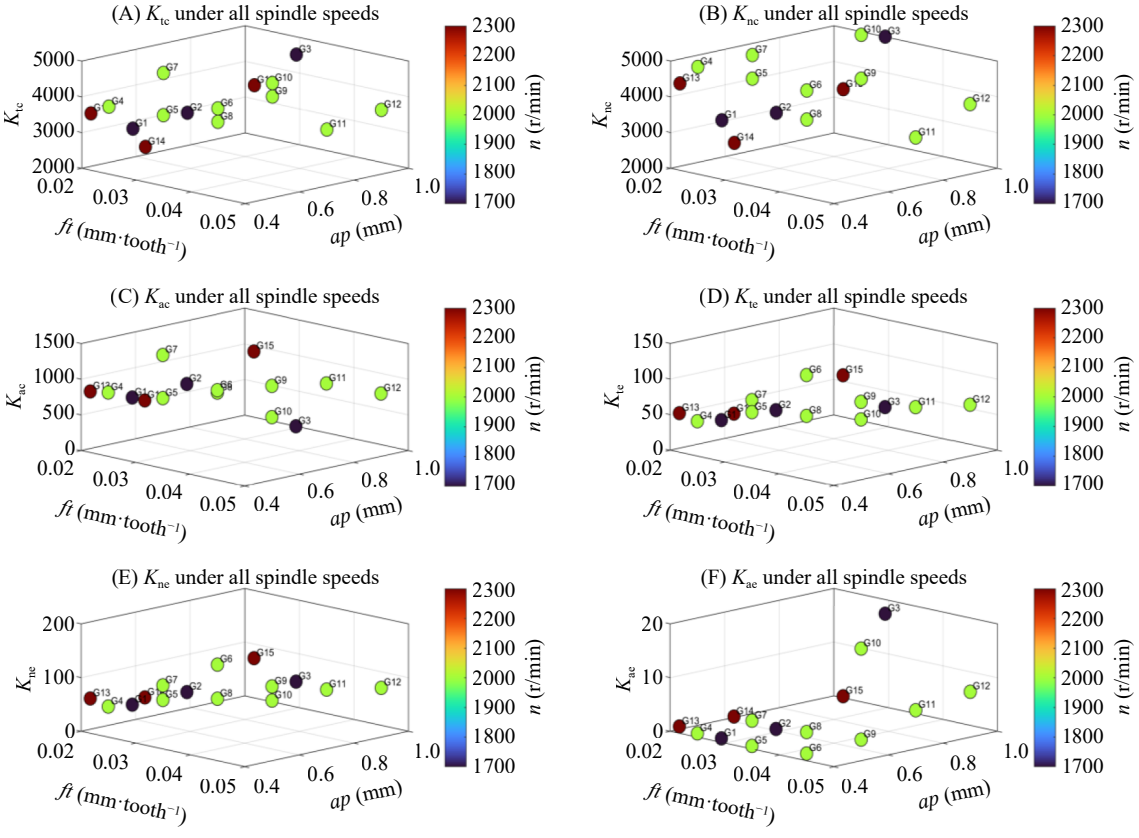


Fig. 7 Distribution of the six equivalent cutting-force coefficients across the 15 operating conditions.

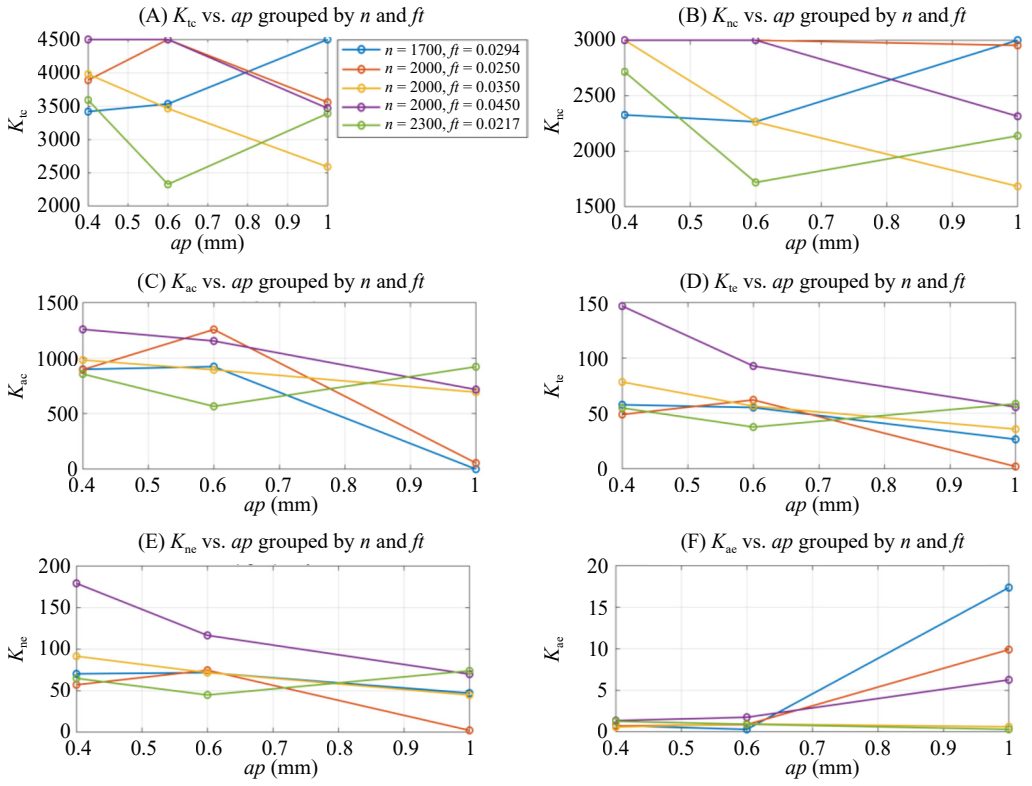


Fig. 8 Variation of equivalent cutting-force coefficients with axial depth of cut under different spindle speeds and feed per tooth.

coefficients should be treated as process-dependent states rather than universal constants.

The mechanistic model was evaluated under matched and cross-group conditions. As shown in Fig. 9A, the low RMSEs across all 15 matched conditions confirm its accuracy when calibrated for the corresponding operating state. The representative waveforms in Fig. 9B further demonstrate that the model reproduces the dominant force amplitude and periodic characteristics. However, Fig. 9C shows that applying the model calibrated on Group 4 to Group 12 substantially increases the RMSEs of F_x, F_y, and F_z. This cross-group degradation highlights the limited transferability of fixed-parameter mechanistic models and motivates online parameter adaptation.

4.3 Prediction performance

Prediction performance was evaluated on the held-out test and generalization sets across four forecast horizons. Fig. 10 compares the predicted and measured triaxial forces for the test condition at P = 50, 150, 300, and 600. Across the four horizons, the predictions remain phase-aligned with the dominant periodic waveform, while the axis-wise errors vary with the forecast length.

The corresponding quantitative results are reported in Table 3. The horizon-dependent results therefore indicate a tradeoff between axis-specific accuracy and forecast span rather than a monotonic degradation with increasing P. Fig. 11 presents representative per-sample agreement between the predicted and measured forces using the first 600 consecutive samples for each horizon; its panel-wise R^2 values are calculated on these subsets, whereas Table 3 reports the full-test-set metrics.

Generalization to unseen operating-condition combinations was evaluated using Groups 9 and 13. The corresponding quantitative results for Groups 9 and 13 are reported in Table 4. The deterioration at P = 600 for Group 13 is concentrated mainly in the F_y direction. Fig. 12 presents the corresponding scatter plots with Groups 9 and 13 overlaid in each panel. The two groups exhibit comparable dispersion around the identity line, and the overall R^2 values remain above 0.989 across all panels, confirming that the prediction consistency extends to parameter combinations unseen during training.

To validate the predictive performance of the Physics-Guided Multi-Source Milling Force Predictor, three literature-inspired baseline models were reimplemented under a unified experimental setting. (1) B1_BiGRU is a standard pure time-series baseline constructed in this study. It uses only historical measured cutting forces. (2) B2_CNNBiLSTM is adapted from the parallel CNN-BiLSTM architecture reported by Cheng et al. [21]. For a controlled comparison, it is reconfigured to use the same

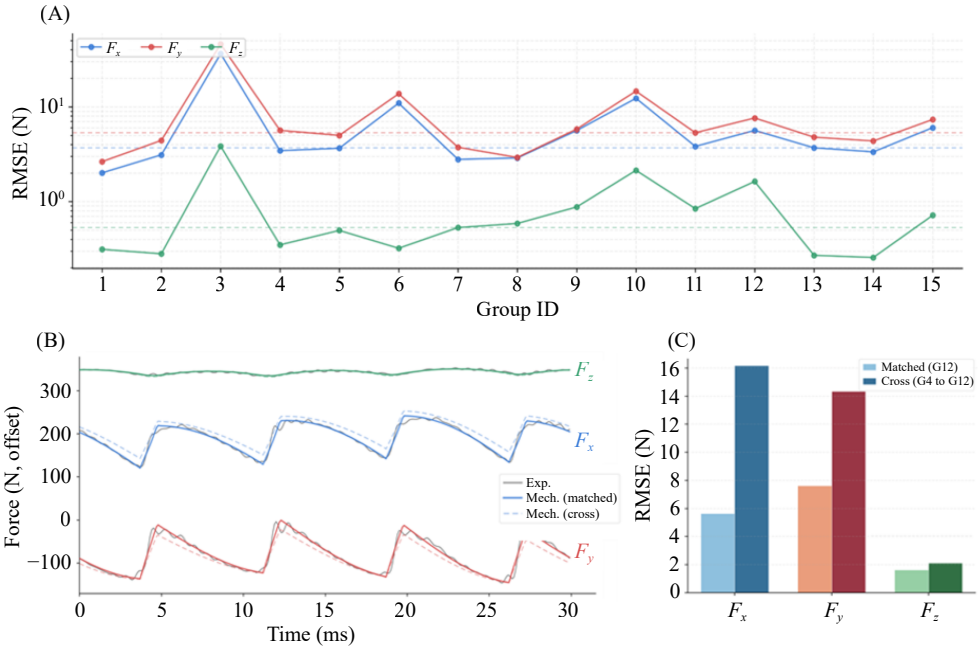


Fig. 9 Mechanistic model validation. (A) RMSE distribution across 15 independent groups. (B) Representative force waveform comparison under matched and cross-group conditions. (C) RMSE comparison under cross-group conditions.

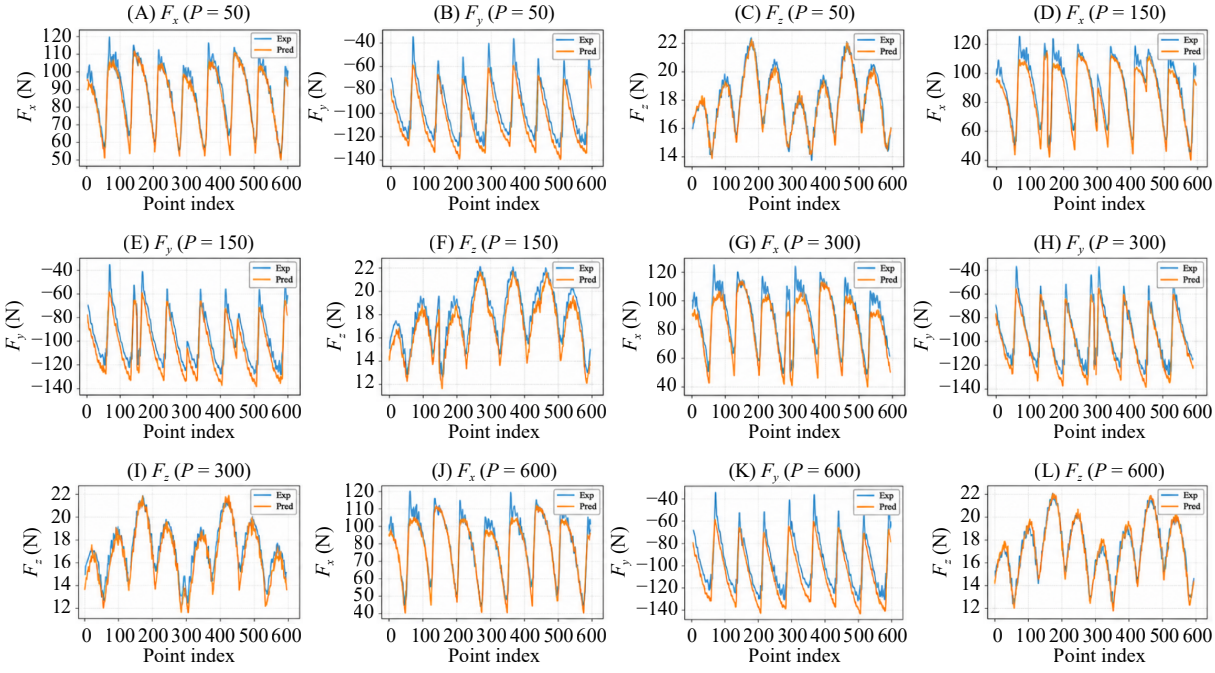


Fig. 10 Comparison of prediction results on the test set under different prediction windows.

10-channel inputs as the proposed predictor but does not incorporate the proposed mechanistic-force-guided future-step attention, thereby evaluating the contribution of the attention module. (3) B3_MechGRU is a simplified mechanism-guided GRU baseline inspired by the mechanistic-force input strategy of Ni et al. [12] and the mechanism-informed FD-GRU framework of Cheng et al. [27]. This baseline is used to evaluate the combined gains provided by the dual-branch TCN, adaptive gated fusion, and mechanistic-force-guided temporal attention modules. All models used identical data partitions, preprocessing, training settings, and prediction horizons

Table 3 Prediction accuracy under different horizons on the test set.

Pre_Window	RMSE			R^2			Overall_RMSE	Overall_R^2
	F_x	F_y	F_z	F_x	F_y	F_z		
50	5.451	10.168	0.612	0.921	0.808	0.923	6.670	0.993
150	11.506	9.689	1.684	0.650	0.826	0.418	8.739	0.987
300	8.755	14.578	1.714	0.797	0.606	0.398	9.867	0.984
600	7.245	8.280	0.644	0.861	0.873	0.915	6.363	0.993

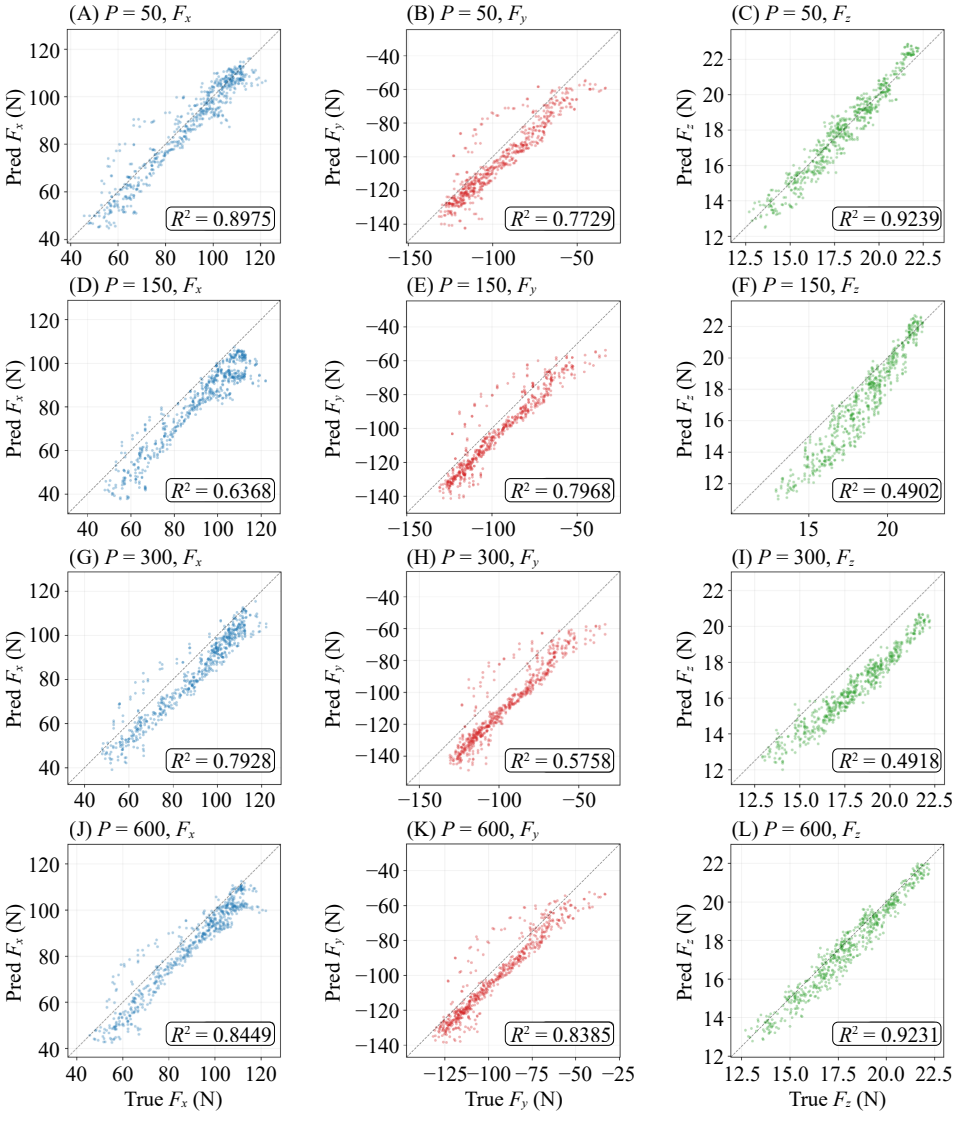


Fig. 11 Predicted versus measured cutting force on the test set (Group 5) across four prediction horizons. (A–C) $P = 50$. (D–F) $P = 150$. (G–I) $P = 300$. (J–L) $P = 600$.

of $P = 50$ and $P = 150$. Performance was evaluated using RMSE and R^2 in the physical force domain, as reported in Table 5.

The proposed Predictor consistently outperformed all baselines. On the held-out test condition, the overall RMSE was 10.70 N at $P = 50$ and 7.97 N at $P = 150$, representing reductions of 36.5%–40.5% and 55.0%–56.5%, respectively. On the unseen parameter combinations, the overall RMSEs were 4.34 and 4.57 N, with R^2 values of 0.9980 and 0.9977, corresponding to reductions of approximately 76.4%–77.7% relative to the baselines.

Table 4 Prediction accuracy on the unseen operating-condition combinations.

Pre_Window	Group	RMSE			R^2			Overall_RMSE	Overall_R^2
		F_x	F_y	F_z	F_x	F_y	F_z		
50	9	6.129	8.83	1.149	0.964	0.948	0.935	6.241	0.998
50	13	3.139	5.168	0.403	0.948	0.901	0.909	3.499	0.995
150	9	5.117	7.701	1.294	0.975	0.960	0.917	5.390	0.999
150	13	4.300	4.594	0.914	0.903	0.922	0.531	3.671	0.994
300	9	5.394	7.106	1.618	0.972	0.966	0.871	5.235	0.999
300	13	4.347	3.762	0.915	0.901	0.948	0.530	3.361	0.995
600	9	5.495	9.006	2.239	0.971	0.946	0.752	6.226	0.998
600	13	3.326	8.391	0.626	0.942	0.740	0.780	5.224	0.989

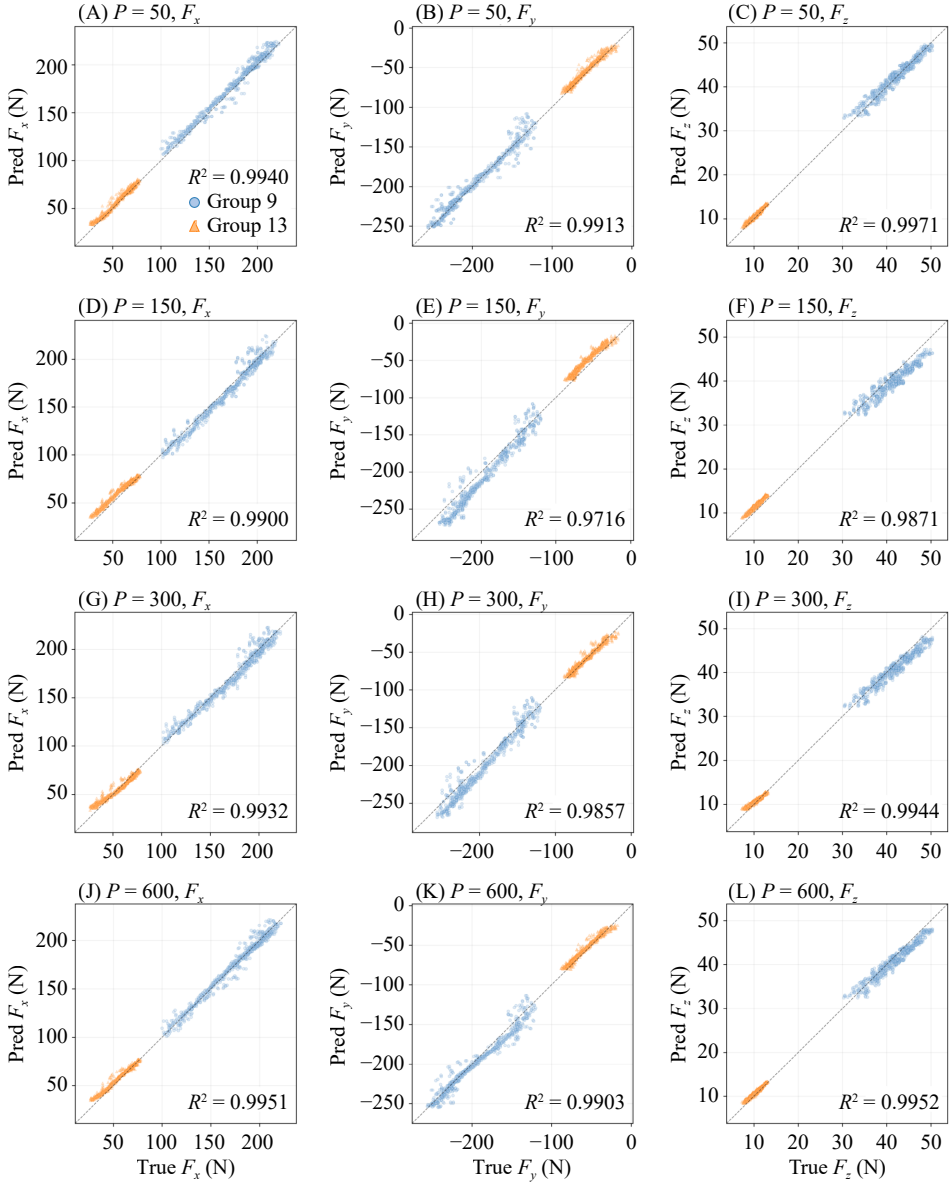


Fig. 12 Predicted versus measured cutting force on the generalization set (Groups 9 and 13) across four prediction horizons. (A–C) P = 50. (D–F) P = 150. (G–I) P = 300. (J–L) P = 600.

Table 5 Comparison with baseline models at $P = 50$ and $P = 150$.

Model	P	Set	F_x RMSE	F_y RMSE	F_z RMSE	Overall RMSE	R^2
V0_Full	50	test	5.74	17.46	2.38	10.70	0.9811
		gen	4.33	6.09	0.77	4.34	0.9980
	150	test	6.26	12.24	1.32	7.97	0.9895
		gen	5.56	5.45	1.48	4.57	0.9977
B1_BiGRU	50	test	19.03	22.03	2.03	16.85	0.9532
		gen	20.73	25.09	2.44	18.84	0.9614
	150	test	19.67	23.45	2.21	17.72	0.9483
		gen	21.61	26.86	2.71	19.96	0.9565
B2_CNNBiLSTM	50	test	20.23	23.44	3.40	17.99	0.9467
		gen	21.70	25.73	2.78	19.50	0.9586
	150	test	20.69	23.72	4.01	18.32	0.9447
		gen	21.73	25.87	3.20	19.59	0.9581
B3_MechGRU	50	test	20.10	22.05	2.01	17.27	0.9509
		gen	20.07	24.67	2.26	18.41	0.9631
	150	test	20.48	23.14	2.20	17.89	0.9473
		gen	22.02	26.24	2.76	19.84	0.9570

In addition to prediction accuracy, the computational cost of the proposed framework was evaluated. For $P = 50$, the Predictor contained approximately 1.57 million parameters and occupied 6.2 MB, whereas OnlineParamNet contained 15.3 thousand parameters and occupied less than 0.1 MB. Their measured inference latencies were approximately 98 ms and 37 μs, respectively, indicating negligible overhead for online parameter correction.

4.4 Parameter adaptation

4.4.1 Condition transitions

To evaluate OnlineParamNet under operating-condition transitions, an uninterrupted three-stage milling experiment was conducted at a constant spindle speed of 2000 r/min. The axial depth of cut and feed per tooth were sequentially varied from 0.4 mm and 0.025 mm/tooth to 0.6 mm and 0.035 mm/tooth, and finally to 1.0 mm and 0.045 mm/tooth. Triaxial cutting force, acceleration, and spindle-power signals were continuously acquired, yielding approximately 22,950 samples. For reference, a stage-wise offline-calibrated mechanistic result was constructed for each operating stage. The coefficient vector of each stage was independently identified using the complete force data of that stage and then used to calculate the corresponding mechanistic response. A 200-sample sigmoid interpolation was applied only at the stage boundaries. Because the complete stage data were used for identification, this result is non-causal and serves only as a phase-matched mechanistic reference.

In contrast, PPAT-Net generated causal predictions in a rolling-window manner using only information available before each prediction interval. As shown in Fig. 13, the predicted forces track the stage-wise load changes and dominant tooth-periodic characteristics without divergence. Here, “Exp” denotes the experimentally measured cutting force, “Pred” denotes the causal PPAT-Net prediction, and “Mech” denotes the stage-wise offline-calibrated mechanistic reference defined above. Table 6 summarizes the quantitative results. “Mech Updated RMSE” denotes the RMSE of the pure mechanistic response calculated using only the coefficient vector updated online by OnlineParamNet, without compensation from the Predictor. Therefore, the Mech metrics should be interpreted as reference measures of mechanistic fidelity rather than as the performance of a competing causal predictor, whereas Mech Updated RMSE is used to independently evaluate the online parameter-recovery capability of OnlineParamNet. “Pred Var” and “Exp Var” denote the variances of the predicted and measured cutting-force signals, respectively.

The equivalent cutting-force coefficients exhibited pronounced nonstationary variation across the three operating stages. The shearing coefficients K_{tc}, K_{nc}, and K_{ac} decreased from 3889.5, 3000.0, and 895.4 to 3468.4, 2314.9, and 716.7, corresponding to reductions of 10.8%, 22.8%, and 20.0%, respectively. Among the edge-force coefficients, K_{ae} showed the largest variation, increasing from 0.717 to 6.254, more than eight times its initial

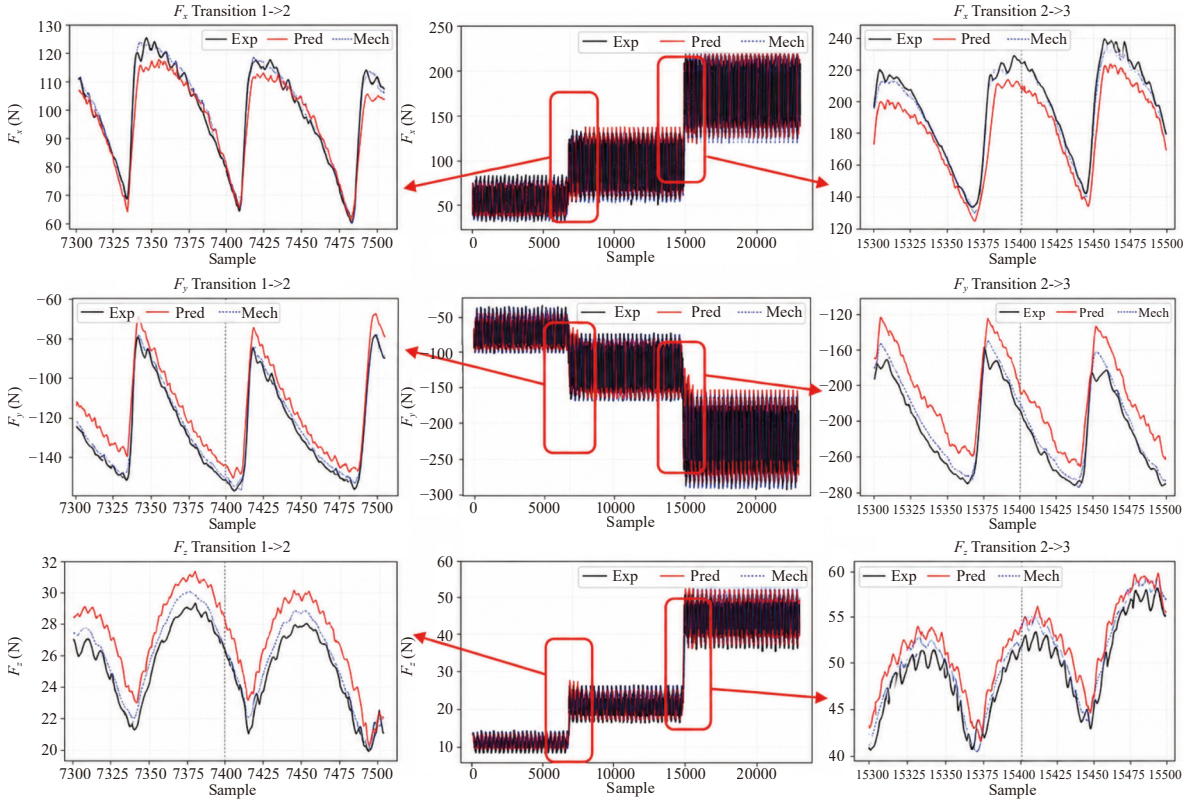


Fig. 13 Comparison of triaxial milling force prediction over the full continuous varying-condition process.

Table 6 Triaxial force accuracy across the three stages of the continuous varying-condition experiment.

Phase	Axis	Pred RMSE	Pred Var	Exp Var	Mech RMSE	Mech Updated RMSE
Phase 1 (ap = 0.4, ft = 0.025)	F_x	3.479	339.165	398.54	3.541	5.285
	F_y	4.944	434.885	543.788	6.159	7.03
	F_z	0.685	25.324	21.002	0.386	1.113
Phase 2 (ap = 0.6, ft = 0.035)	F_x	6.596	1257.193	1513.859	3.351	8.402
	F_y	11.764	1823.335	2264.211	4.074	10.155
	F_z	1.307	109.856	109.803	0.96	2.038
Phase 3 (ap = 1.0, ft = 0.045)	F_x	6.921	1046.405	1026.627	5.552	7.327
	F_y	10.199	1455.321	1464.634	7.665	7.52
	F_z	2.168	18.450	21.863	2.021	2.021

value. These stage-dependent changes demonstrate that the coefficient vector cannot be regarded as invariant during operating-condition transitions.

The accepted online corrections ranged from approximately 10% to 350% of the initial coefficient values, reflecting the substantial parameter variation encountered during the continuous experiment. In a representative Phase 1 update, nonlinear least squares reduced the normalized mechanistic-force error from 0.0843 to 0.0751 within eight iterations, providing the regularized reference correction used for updater training.

On the validation set, OnlineParamNet achieved prediction deviations of approximately ±18.8, ±18.8, ±9.4, ±0.9, ±0.9, and ±0.09 for the six coefficients, respectively, corresponding to relative errors ranging from 2.5% to 22.6%. These results demonstrate that the lightweight updater can estimate substantial coefficient variations with sufficient accuracy to support stable, physically constrained online adaptation.

4.4.2 Prolonged continuous cutting

To investigate parameter evolution during prolonged continuous cutting under an identical operating condition,

the milling process was divided into 100 consecutive stages, each corresponding to one cutting pass. Throughout the experiment, the spindle speed, feed rate, axial depth of cut, and radial width of cut were fixed at 2300 r/min, 200 mm/min, 1.0 mm, and 18 mm, respectively. Each pass generated 2 s of synchronized cutting-force, acceleration, and spindle-power data at a sampling frequency of 10 kHz, resulting in a total of 2,000,000 samples.

The cutting-force coefficients identified from the first pass were used as the fixed-parameter mechanistic baseline, while the remaining passes were processed sequentially. As shown in Fig. 14, the pass-wise RMSE between the mechanistic model and the measured forces increased progressively.

Fig. 15 compares the measured and fixed-parameter mechanistic forces at six representative passes. The dominant tooth-passing periodicity and phase alignment remain largely preserved throughout the experiment, whereas the amplitude mismatch becomes progressively more pronounced, particularly in the F_x and F_z components.

Fig. 16 compares the fixed-parameter and PPAT-Net models over 100 passes. The fixed model–process discrepancy increased from 5.82 to 16.16 N, whereas OnlineParamNet maintained the RMSE near 6.8 N.

Across all six coefficients, the shear coefficients K_{tc}, K_{nc}, and K_{ac} exhibit a slow upward trend of approximately 6%–8% over the full 100 passes, while the edge coefficients K_{te}, K_{ne}, and K_{ae} are more sensitive to tool condition and drift by 40%–50%. As shown in Fig. 17, the step-wise trajectory of OnlineParamNet remains close

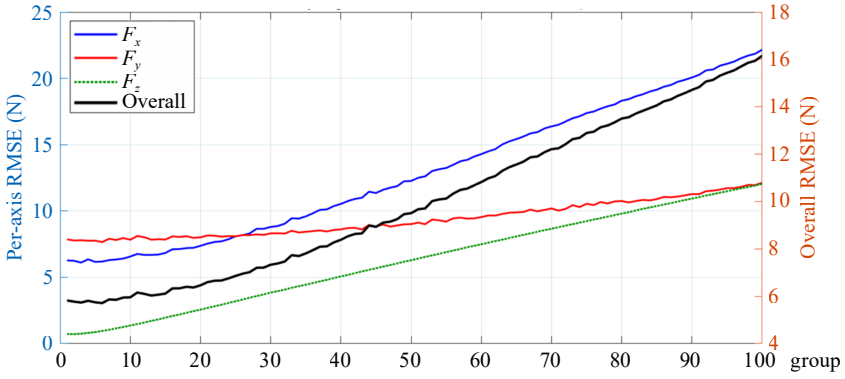


Fig. 14 Pass-wise RMSE of the fixed-parameter mechanistic model.

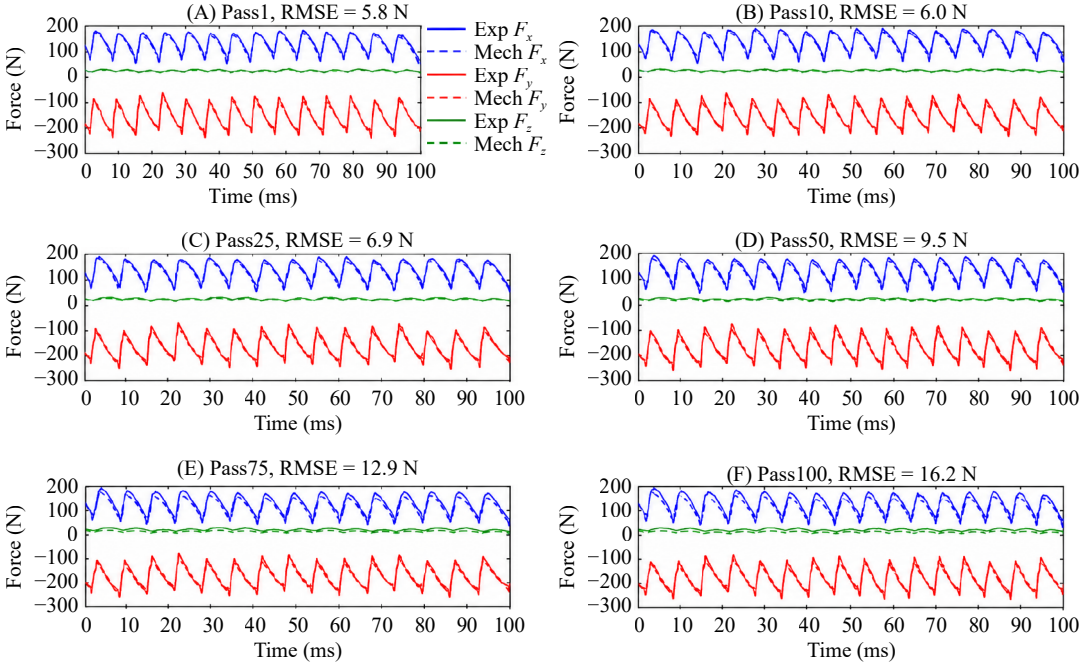


Fig. 15 Representative measured and fixed-parameter mechanistic force waveforms.

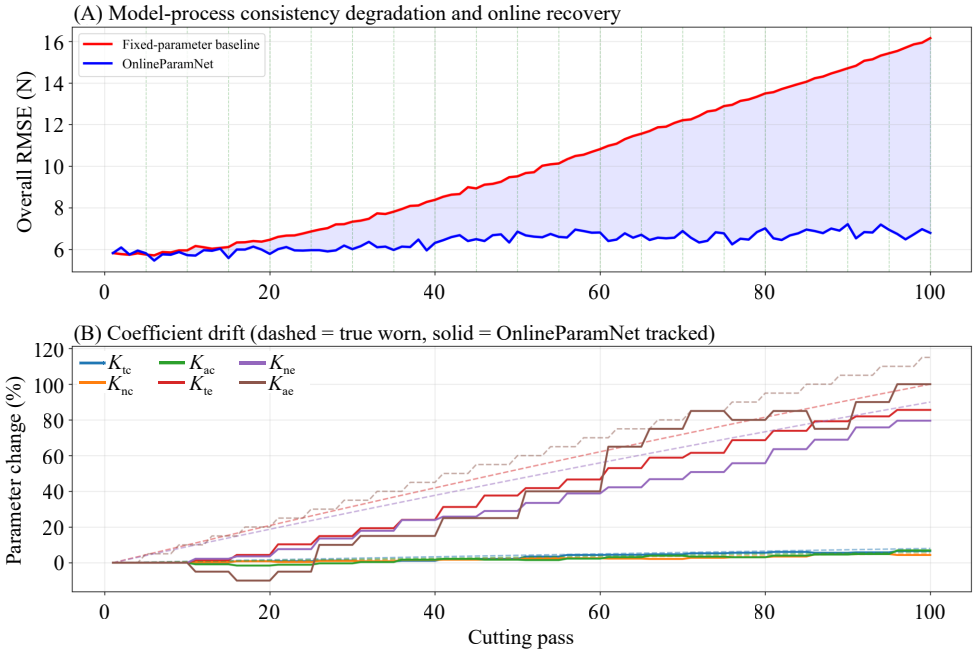


Fig. 16 Fixed-parameter versus online-updated model performance.

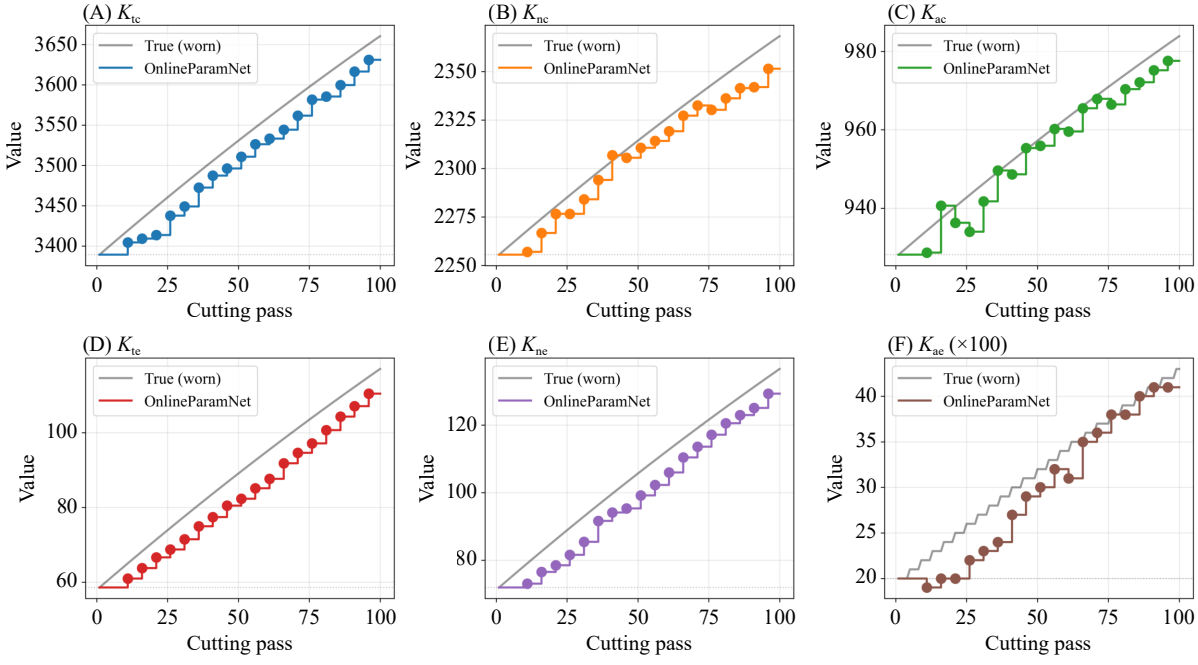


Fig. 17 OnlineParamNet coefficient tracking across 100 consecutive cutting passes.

to the true coefficient evolution throughout all passes, with no instances of divergence, overcorrection, or physically implausible parameter values.

4.5 Ablation study

The ablation study evaluates the contribution of the main forecasting components independently of the coefficient trajectory analysis in Section 4.4. Six variants were considered: the complete model (V0_Full), removal of the signal branch (V1_NoSignal), removal of the measured/mechanistic force branch (V2_NoForce), replacement of AGF with summation fusion (V3_NoAGF), removal of the BiGRU (V4_NoBiGRU),

and removal of future-step attention (V5_NoAttention). Specifically, V2_NoForce removes the entire historical measured/mechanistic force branch; therefore, no historical measured cutting-force signal is provided to the Predictor. The remaining prediction inputs include acceleration, spindle power, operating-condition information, and the future mechanistic-force prior. All variants used identical dataset partitions, prediction horizons, preprocessing, and training settings. The denormalized overall RMSE and R^2 were averaged over three independent seeds (42, 123, and 456), with Test denoting Group 5 and Gen denoting the held-out parameter combinations. The numerical conclusions are based on the RMSE values in Table 7.

Of particular relevance to the dependence of the Predictor on measured cutting force, V2_NoForce evaluates forecasting without historical measured-force input. As shown in Table 7, V2_NoForce achieved RMSEs of 7.28 and 5.07 N on the Test and Generalization sets, respectively, at P = 50, and 8.45 and 5.32 N, respectively, at P = 150. Its average RMSE across the four evaluation scenarios was 6.53 N, compared with 5.99 N for the full model. Thus, removing the historical force branch increased the average RMSE by only 0.54 N, while meaningful future-force prediction was still retained. These results indicate that historical measured force is beneficial for characterizing the current model–process discrepancy and machining state, but is not an indispensable input for future-force forecasting. In the absence of historical measured-force input, the Predictor can still exploit acceleration, spindle-power signals, operating-condition information, and the future mechanistic-force prior to estimate the subsequent cutting-force response.

Future-step attention is the dominant forecasting component. Removing it increases the average overall RMSE from 5.99 to 12.51 N. On the generalization conditions, the RMSE rises from 4.42 to 11.60 N at P = 50 and from 4.70 to 14.94 N at P = 150, demonstrating that conditioning each future step on its corresponding mechanistic force is essential for prediction across unseen operating parameter combinations.

The cross-seed distributions in Fig. 18 reinforce the attention result: V5_NoAttention remains separated from the other variants in all four scenarios, whereas the intermediate variants overlap considerably. Consequently, the relative ranking among NoSignal, NoForce, NoAGF, and NoBiGRU should be regarded as indicative rather than definitive for the present three-seed evaluation.

Beyond aggregate RMSE, the gate-weight analysis in Fig. 19 reveals that AGF provides a distinct capability unavailable to the simple concatenation variant: condition-dependent adaptive reweighting. For the unseen parameter combination G13 at P = 50, the force-branch gate weight drops to 0.333 (compared with 0.919 for the seen test condition G5), indicating that AGF autonomously shifts 58.5% of the fusion weight to the signal branch when the mechanistic-force prior becomes unreliable. For G9, where the force prior remains trustworthy, the force gate remains at 0.991. This adaptive behavior detects prior quality under each condition and redistributes trust accordingly. It cannot be replicated by a static concatenation layer and constitutes the primary architectural value of AGF beyond aggregate RMSE parity.

In summary, the ablation results demonstrate that mechanistic-force-conditioned future step attention provides the most consistent contribution to predictive performance, particularly for short-horizon forecasting and cross condition generalization. The V2_NoForce results further show that the historical force branch improves the representation of the current model–process discrepancy, although meaningful forecasting can still be achieved without historical measured-force input. The BiGRU also provides measurable benefits, whereas its effect is comparatively moderate and varies across operating conditions. By contrast, AGF does not produce a universal reduction in aggregate RMSE; its main value lies in condition-dependent assessment of prior reliability and adaptive redistribution of model reliance.

5 Conclusions

This study investigated the degradation of mechanistic-

Table 7 Comparison of overall prediction performance in ablation experiments.

Model	P50 Test RMSE	P50 Gen RMSE	P150 Test RMSE	P150 Gen RMSE	Average RMSE	Relative Full
V0_Full	8.09 ± 0.14	4.42 ± 0.52	6.76 ± 1.33	4.70 ± 0.64	5.99	+0.00
V1_NoSignal	8.46 ± 0.72	4.62 ± 0.47	6.53 ± 0.06	4.84 ± 0.37	6.11	+0.12
V2_NoForce	7.28 ± 0.65	5.07 ± 0.29	8.45 ± 1.42	5.32 ± 0.60	6.53	+0.54
V3_NoAGF	6.34 ± 0.02	4.33 ± 0.17	6.61 ± 0.15	5.11 ± 0.82	5.60	−0.39
V4_NoBiGRU	7.81 ± 0.88	6.11 ± 1.21	7.06 ± 1.56	5.17 ± 0.68	6.54	+0.55
V5_NoAttention	9.44 ± 1.11	11.60 ± 0.41	14.06 ± 0.25	14.94 ± 0.53	12.51	+6.52

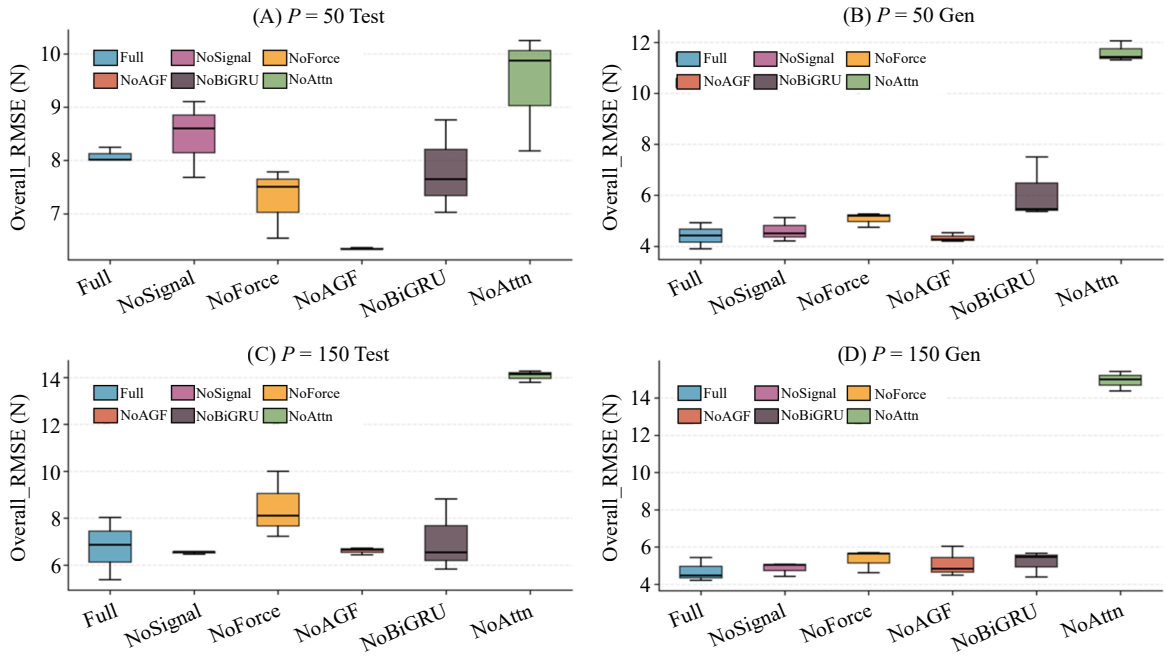


Fig. 18 Cross-seed stability of ablation results across four evaluation scenarios (seeds: 42, 123, 456).

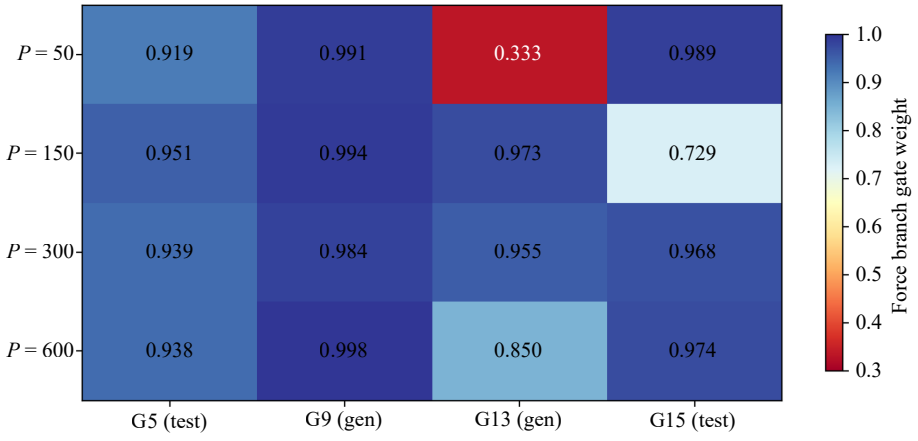


Fig. 19 AGF Force-branch gate weight.

model consistency induced by process evolution in milling and developed PPAT-Net as a closed-loop framework for sustaining the validity of the mechanistic representation. The framework integrates model–process discrepancy evaluation, physically constrained parameter adaptation, and mechanistic-prior regeneration within a rolling prediction process. Its principal contribution lies in reformulating mechanistic-model maintenance from a one-time offline calibration problem into a dynamic consistency-preservation problem, in which process-sensitive physical parameters are continuously corrected according to feedback from the evolving machining process. This formulation enables the mechanistic model to remain aligned with the current process state while preserving its physical interpretability and computational efficiency.

Experimental validation under both operating-condition transitions and prolonged continuous cutting confirmed the effectiveness of this closed-loop mechanism. During 100 consecutive cutting passes under fixed nominal conditions, the error of the initially calibrated mechanistic model increased from 5.82 to 16.16 N, whereas online adaptation maintained the model error at approximately 6.8 N. Stable model updating was also achieved during continuous three-stage condition transitions involving substantial variations in the identified cutting-force coefficients. These results indicate that a significant portion of process-induced model degradation can be

compensated through the online tracking of a compact set of physically meaningful parameters, while the data-driven Predictor accounts for transient and unmodeled dynamics beyond the representational capability of coefficient adaptation. The results therefore highlight a broader requirement for physics–data hybrid modeling: physical knowledge should not only be incorporated into the predictive architecture, but its consistency with the physical process should also be continuously evaluated and maintained during operation.

The proposed consistency-preserving mechanism has broader implications for high-performance manufacturing, in which model fidelity must be sustained under long-duration operation, changing process conditions, and coupled tool–workpiece–machine dynamics. A continuously updated mechanistic representation can provide a physically grounded basis for process-state estimation, quality prediction, adaptive parameter optimization, and closed-loop machining control. It is also directly relevant to intelligent machining digital twins, where persistent virtual–physical consistency requires the embedded process models to evolve synchronously with the physical system rather than remain fixed after initialization. From this perspective, the present framework provides a methodological basis for integrating model-consistency assessment and online model evolution into digital-twin-enabled machining systems. The current validation is restricted to milling-force modeling under a specific machine-tool, workpiece-material, and cutting-tool configuration. Future research will extend the framework toward multi-physics process representations, coordinated parameter–structure evolution, and multi-level model updating, thereby supporting persistent virtual–physical synchronization, predictive process optimization, and autonomous decision-making in high-performance manufacturing.

Data availability The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments This work was supported by the National Natural Science Foundation of China (Grant No. U2570241), the Liaoning Provincial Natural Science Foundation for Distinguished Young Scholars (Grant No. 2024JH3/50100001), the Science Fund for Distinguished Young Scholars of Dalian City (Grant No. 2023RJ004), the Fundamental Research Funds for the Central Universities (Grant No. DUT25Z2717), and the Changjiang Scholar Program of Chinese Ministry of Education (Grant No. Q2021053, TE2022037). The authors used GPT-5.5 (online version) as a language-editing tool to improve grammar, clarity, and readability. The AI tool was not used for data analysis, data generation, image processing, or scientific content creation. All outputs were carefully reviewed and verified by the authors, who take full responsibility for the final content.

Competing interests The authors declare no conflicts of interest.

References

1. Altintas Y. *Manufacturing automation: metal cutting mechanics, machine tool vibrations, and CNC design*. 2nd ed. Cambridge: Cambridge University Press, 2012
2. Miao H H, Wang C Y, Song W J, et al. Coupling modeling of thermal–dynamics–milling process for spindle system considering nonlinear characteristics. Nonlinear Dynamics, 2024, 112(8): 6061–6099
3. Liu Y, Xiong Z H, Liu Z Q. Stochastic cutting force modeling and prediction in machining. Journal of Manufacturing Science and Engineering, 2020, 142(12): 121004
4. Dong Y F, Chen F F, Lu T L, et al. Research on thermal stiffness of machine tool spindle bearing under different initial preload and speed based on FBG sensors. The International Journal of Advanced Manufacturing Technology, 2022, 119(1–2): 941–951
5. Matsumura T, Tamura S. Cutting force model in milling with cutter runout. Procedia CIRP, 2017, 58: 566–571
6. Chen K D, Zhang H, van de Wouw N, et al. An alternative approach to compute chip thickness in milling. Journal of Manufacturing Science and Engineering, 2022, 144(11): 111006
7. Bissacco G, Hansen H N, Slunsky J. Modelling the cutting edge radius size effect for force prediction in micro milling. CIRP Annals, 2008, 57(1): 113–116
8. Malekian M, Park S S, Jun M B G. Modeling of dynamic micro-milling cutting forces. International Journal of Machine Tools and Manufacture, 2009, 49(7−8): 586–598
9. Wang J C, Zou B, Liu M F, et al. Milling force prediction model based on transfer learning and neural network. Journal of Intelligent Manufacturing, 2021, 32(4): 947–956
10. Pimenov D Y, Bustillo A, Wojciechowski S, et al. Artificial intelligence systems for tool condition monitoring in machining: analysis and critical review. Journal of Intelligent Manufacturing, 2023, 34(5): 2079–2121
11. Teti R, Mourtzis D, D'Addona D M, et al. Process monitoring of machining. CIRP Annals, 2022, 71(2): 529–552
12. Ni C, Yang J X, Ding H. Mechanism and data hybrid-driven cutting forces prediction model for end milling. Journal of Intelligent Manufacturing, 2026, 37(1): 481–503
13. Malea C I, Nitu E L, Iordache M D. A brief review of numerical simulation in process machining. IOP Conference Series: Materials Science and Engineering, 2021, 1009(1): 012035
14. Soori M, Arezoo B. Virtual machining systems for CNC milling and turning machine tools: a review. International Journal of Engineering and Technology, 2020, 18(1): 56–104
15. Wan M, Zhang W H, Qin G H, et al. Efficient calibration of instantaneous cutting force coefficients and runout parameters for general end mills. International Journal of Machine Tools and Manufacture, 2007, 47(11): 1767–1776
16. Grossi N, Sallese L, Scippa A, et al. Speed-varying cutting force coefficient identification in milling. Precision Engineering, 2015, 42: 321–334
17. Li C, Duan C Z, Chang B B. Instantaneous cutting force model

considering the material structural characteristics and dynamic variations in the entry and exit angles during end milling of the aluminum honeycomb core. Mechanical Systems and Signal Processing, 2022, 181: 109456

18. Wan M, Wen D Y, Ma Y C, et al. On material separation and cutting force prediction in micro milling through involving the effect of dead metal zone. International Journal of Machine Tools and Manufacture, 2019, 146: 103452
19. Liao Z R, Abdelhafeez A, Li H N, et al. State-of-the-art of surface integrity in machining of metal matrix composites. International Journal of Machine Tools and Manufacture, 2019, 143: 63–91
20. Charalampous P. Prediction of cutting forces in milling using machine learning algorithms and finite element analysis. Journal of Materials Engineering and Performance, 2021, 30(3): 2002–2013
21. Cheng M H, Jiao L, Yan P, et al. Intelligent tool wear monitoring and multi-step prediction based on deep learning model. Journal of Manufacturing Systems, 2022, 62: 286–300
22. Huang Z W, Zhu J M, Lei J T, et al. Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations. Journal of Intelligent Manufacturing, 2020, 31(4): 953–966
23. Lei Y G, Yang B, Jiang X W, et al. Applications of machine learning to machine fault diagnosis: a review and roadmap. Mechanical Systems and Signal Processing, 2020, 138: 106587
24. Wang J J, Li Y L, Zhao R, et al. Physics guided neural network for machining tool wear prediction. Journal of Manufacturing Systems, 2020, 57: 298–310
25. Liu X, Jiang D, Tao B, et al. A systematic review of digital twin about physical entities, virtual models, twin data, and applications. Advanced Engineering Informatics, 2023, 55: 101876
26. Xu K, Li Y G, Zhang J C, et al. ForceNet: an offline cutting force prediction model based on neuro-physical learning approach. Journal of Manufacturing Systems, 2021, 61: 1–15
27. Cheng Y H, Li Y G, Zhuang Q Y, et al. Mechanism-informed friction-dynamics coupling GRU neural network for real-time cutting force prediction. Mechanical Systems and Signal Processing, 2024, 221: 111749
28. Zha Z H, Guo P, Xiong Z W, et al. Deep learning physical hybrid model-driven *in situ* cutting force monitoring of ultra-precision diamond turning. Measurement Science and Technology, 2025, 36(9): 096120
29. Zhu K P, Li J. Physics-guided deep network for milling dynamics prediction. Engineering, 2025, 55: 71–85
30. Zhu K P, Guo H, Li S, et al. Physics-informed deep learning for tool wear monitoring. IEEE Transactions on Industrial Informatics, 2024, 20(1): 524–533
31. Liu M, Xie H, He X K, et al. Long time series prediction of milling force via a hybrid multi neuro-network-based algorithm. Engineering Applications of Artificial Intelligence, 2026, 163(Pt 1): 112805
32. Liu Y H, Zuo Y, An Y D, et al. Physics-informed hybrid conditional diffusion transformer model for cutting force prediction under param-space optimization. Journal of Intelligent Manufacturing, 2026.
33. Ra D, Lee J, Lee M, et al. Physics-informed machine learning across manufacturing processes: recent advances, challenges, and directions. Journal of Manufacturing Systems, 2026, 85: 72–95
34. Willard J, Jia X W, Xu S M, et al. Integrating scientific knowledge with machine learning for engineering and environmental systems. ACM Computing Surveys, 2022, 55(4): 66