sensors

MDPI

Article

An IG-LSTM-Based Tool Wear Prediction Method for Cross-Operating-Condition Forecasting

Bo Sun *, Jian Zhang, Xiangqin Wu, Lixin Zhang, Lanbo Wang, Chaoran Liu and Yuntao Wei

School of Mechanical and Vehicle Engineering, Changchun University, Changchun 130022, China; 240101002@mails.ccu.edu.cn (J.Z.); 250102027@mails.ccu.edu.cn (X.W.); 250101003@mails.ccu.edu.cn (L.Z.); 260102027@mails.ccu.edu.cn (L.W.); 260102052@mails.ccu.edu.cn (C.L.); 240102027@mails.ccu.edu.cn (Y.W.)
* Correspondence: sunb@ccu.edu.cn

Highlights

What are the main findings?

- IG-LSTM adaptively learns nonlinear tool degradation trends and integrates them with inverse Gaussian stochastic degradation modeling.
- IG-LSTM outperforms conventional IG and LSTM models in both single-tool and cross-tool tool wear prediction.

What are the implications of the main findings?

- The proposed hybrid framework improves prediction accuracy while maintaining the physical plausibility of stochastic tool wear evolution.
- Early-data-based parameter calibration enhances cross-tool generalization, enabling accurate wear prediction with limited target-tool data.

Abstract

Tool wear directly affects the accuracy and reliability of machined workpieces; accurately predicting wear trends is crucial for developing reasonable tool-change strategies and avoiding sudden tool failure. Traditional parametric Inverse Gaussian (IG) processes can effectively characterize the non-negativity and stochastic fluctuations of incremental tool wear; however, the mean functions in such models often adopt an analytically defined power-law form that is arbitrarily preset, making it difficult to adapt to the nonlinear evolution patterns of tools at different stages. Consequently, their generalization capabilities are limited in cross-tool prediction scenarios. To address this, this paper proposes an IG-LSTM tool wear prediction method designed for cross-tool forecasting. First, Variational Mode Decomposition (VMD) is employed to extract the degradation trend of wear, and a Long Short-Term Memory (LSTM) neural network is trained based on this. Next, the trend predicted by the LSTM is used as the cumulative degradation function of the IG process to construct a degradation model for wear increment; simultaneously, the model parameters of the IG process are estimated via maximum likelihood estimation based on the training data. Finally, two sets of experiments—single-tool prediction and cross-tool prediction—were designed using the PHM2010 dataset. The experimental results show that, in both scenarios, the proposed method significantly outperforms traditional parametric inverse Gaussian models and purely data-driven LSTM models, achieving an organic integration of data-driven adaptive trend extraction and physics-driven stochastic degradation modeling.

Academic Editor: Chao Liu

Received: 16 October 2026
Revised: 9 November 2026
Accepted: 9 November 2026
Published: 12 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: remaining tool life; inverse Gaussian process; variable-mode decomposition; mechanism-data dual-driven approach; long-short-term memory network

1. Introduction

As the direct agents of mechanical cutting operations, the wear condition of cutting tools directly affects the machining accuracy, surface quality, and production efficiency of parts; in severe cases, it can result in workpieces exceeding tolerance limits and being scrapped, or even cause machine tool downtime [1,2]. Relevant studies indicate that 20% of machine tool downtime is caused by tool failure, and tool monitoring systems can reduce unplanned downtime in CNC machine tools by 75% while increasing productivity by 10% to 60% [3]. Therefore, accurate prediction of tool wear trends is a prerequisite and foundation for further predicting the remaining service life of tools, and it is of great significance for formulating reasonable tool change strategies, avoiding sudden failures, and reducing production costs.

In recent years, numerous scholars have conducted extensive research on tool wear prediction. Based on their underlying principles, existing studies can be broadly categorized into two approaches: physics-based and data-driven methods [4]. The physical model approach establishes the relationship between characteristic parameters and the state of degradation in a fixed mathematical form based on the mechanisms of tool degradation. Li [5] et al. systematically reviewed theoretical analysis and numerical simulation methods for tool wear, pointing out that by coupling cutting process variables with wear rate models, it is possible to dynamically predict changes in tool geometry and tool life. Yen [6] et al. replaced empirical fitting with mechanism-based wear equations and incorporated thermo-mechanical coupling effects using the finite element method, thereby providing a basis for the physical modeling of tool wear. Physics-driven tool wear models construct analytical expressions based on cutting parameters and wear mechanisms. While these models possess clear physical interpretability, they rely heavily on idealized experimental assumptions. Their predefined structure makes it difficult to adapt to dynamic disturbances—such as sudden changes in cutting forces and abnormal wear—encountered in actual machining environments, thereby limiting their predictive accuracy under complex real-world conditions.

These limitations have prompted researchers to turn their attention to data-driven methods, which do not rely on precise physical descriptions but instead achieve adaptive tracking of wear conditions through the analysis of monitoring data. Among data-driven methods, deep learning [7,8] has garnered widespread attention due to its powerful nonlinear feature extraction capabilities. Anhua [9] et al. integrated sparse autoencoders with neural networks, using the Pearson correlation coefficient to select sensitive features and introducing noise and dropout strategies to address overfitting. Tool wear prediction has shown a clear trend toward evolving from single-network models to multi-module fusion approaches. Examples include the combination of attention mechanisms and stacked LSTMs [10], the ConvLSTM Att model [11], parallel CNNs with channel attention [12], and the combination of CNNs with bidirectional LSTMs/GRUs [13]. As research shifts from single, fixed operating conditions to variable real-world conditions, the focus has gradually shifted to generalization capabilities under variable conditions and with limited labels. Zhang [14] and others have fused multisensor data with process information as input to a multichannel residual network and fine-tuned the fully connected layers to achieve cross-condition prediction. Deep learning methods possess powerful nonlinear fitting capabilities, but such methods are black-box models that lack physical a priori constraints on the degradation process and suffer from insufficient interpretability; Most

methods only output single-point predictions, making it difficult to capture the intrinsic stochastic uncertainty in wear evolution; model performance is highly dependent on a large number of labeled samples, and overfitting is prone to occur in cross-tool transfer scenarios.

Degradation modeling methods based on stochastic processes do not rely on feature extraction from sensor signals; instead, they treat tool wear as a stochastic variable that evolves over time. By constructing stochastic process models—such as the Wiener process [15,16], the gamma process [17,18], and the inverse Gaussian process [19,20]—one can directly describe the distribution and time-varying behavior of wear increments, thereby enabling a probabilistic characterization and prediction of the wear degradation trajectory. Regarding the Wiener process, Wang [21] et al. established a multistage Wiener process model, dividing the degradation process into three stages based on typical wear curves. They estimated the parameters for each stage using the EM algorithm and combined this with Monte Carlo simulations to perform wear prediction and reliability analysis. Regarding Gamma processes, Li [22] et al. modeled tool wear degradation using a Gamma process, deriving analytical expressions for the time-varying reliability of the tool and its sensitivity under six operating conditions. Through numerical examples, they revealed the patterns of how various parameters affect reliability, providing a quantitative basis for improving tool reliability through compensation strategies and parameter adjustments. However, the Wiener process assumes that increments follow a normal distribution, which allows for negative growth in degradation—a contradiction to the strictly monotonically increasing physical nature of tool wear. Although the Gamma process satisfies the monotonicity requirement, its shape and scale parameters take the same fixed values for all individuals, making it difficult to reflect the inherent differences in degradation rates among different tool individuals. In contrast, the inverse Gaussian process combines strict monotonicity with the ability to flexibly model stochastic effects, and offers significant advantages in modeling individual heterogeneity.

Previous studies have attempted to combine physical degradation models with data-driven methods to improve the accuracy and physical plausibility of tool wear prediction. For example, Zhu [23] et al. incorporated a physical wear model into the mean function of Gaussian process regression, while Qiang [24] et al. combined a physical information loop Gaussian process with instance transfer learning to leverage existing sample information and improve the predictive performance for the target task. These studies provide valuable insights for integrating wear mechanisms with data-driven modeling; however, differences exist among the various methods in terms of how wear trends are represented, how randomness is described, and the strategies for adapting to target cutting tools. Based on this, this paper does not simply combine LSTM with the IG process; rather, it focuses on how to use LSTM to learn nonlinear degradation trends and embed them into the stochastic incremental modeling of the IG process, while incorporating early observational information of the target cutting tool for parameter calibration.

Specifically, this paper first uses VMD to extract the primary degradation trend from the wear time series, and then employs LSTM to learn and predict the cumulative degradation function t $\Lambda(t)$ of the IG process, thereby reducing the reliance of traditional IG models on preset analytical trend functions. Building on this, the predicted $\Lambda(t)$ is used as the cumulative degradation function of the IG process to model the stochastic variations in wear increments. Thus, the LSTM is used to describe the nonlinear evolution of the degradation trend, while the IG process is used to characterize the stochastic nature of the wear increments. In cross-tool prediction, considering that degradation rates may vary among different tools, this paper further utilizes a small number of early effective wear increments from the target tool to calibrate the tool-specific parameter μ, while retaining the global parameter information estimated from the training tool data. Through

https://doi.org/10.3390/s26227158

the above design, this paper combines adaptive trend prediction, stochastic increment modeling, and target tool parameter calibration to form the IG-LSTM method for cross-tool wear prediction.

The main contributions of this study are as follows:

(1) To address the limitation of traditional inverse Gaussian processes—which rely on predefined analytical forms for the mean function and struggle to capture the multi-stage, nonlinear wear evolution of cutting tools—an LSTM network is introduced to learn the nonlinear degradation trajectory of tool wear, replacing the fixed mean function of the inverse Gaussian process. While retaining the inverse Gaussian process's ability to characterize the stochastic uncertainty of wear degradation, this approach leverages deep learning's capacity to fit complex time-series data to predict tool wear.

(2) To address the issue of reduced prediction accuracy across tools caused by individual tool degradation heterogeneity, we update and calibrate the model parameters using a small amount of early monitoring data from the tool under test—without requiring a complete full-lifetime dataset—thereby improving the model's generalization performance in cross-tool scenarios.

(3) To address the issue that measured tool wear sequences are subject to random cutting fluctuations and that direct modeling is prone to noise interference, variational modal decomposition is used to decompose the wear sequence into multiple eigenmode components. Combined with the zero-crossing criterion, this method automatically filters out low-frequency degenerate main trend components, eliminating high-frequency measurement noise, thereby providing the IG-LSTM hybrid model with a smooth and reliable degraded input sequence and improving the stability and accuracy of subsequent wear predictions.

The remainder of this paper is organized as follows: Section 2 elaborates on the theoretical foundation and overall framework of the proposed IG-LSTM model. It begins by introducing parameterized inverse Gaussian processes and their limitations, followed by a description of the IG-LSTM model's overall architecture. It then sequentially introduces core modules such as data preprocessing and degradation trend extraction, LSTM trend forecasting, IG parameter estimation, and wear increment modeling; Section 3 presents the experimental validation. It begins by introducing the PHM2010 dataset and the data partitioning scheme, then progressively validates the model's effectiveness across four levels—VMD trend extraction, LSTM trend prediction, IG parameter estimation, and cumulative wear prediction—and conducts a comparative analysis with both the LSTM and traditional IG methods; Section 4 summarizes the entire paper, outlines its main contributions, and looks ahead to future research directions.

2. Theory and Methods

2.1. Parametrized Inverse Gaussian Processes

The tool wear process is physically characterized by the irreversible accumulation of damage. Due to factors such as microstructural inhomogeneity of the material, random fluctuations in cutting forces, and variations in operating conditions, the wear increment generated during each cutting operation is not a deterministic value but exhibits random fluctuations. However, from a long-term perspective, the total wear remains monotonically increasing. This implies that the wear increments are strictly non-negative and mutually independent, consistent with the model assumptions of an inverse Gaussian process [25]. If we consider tool wear as a stochastic process, where $\{Y(t),\ t \geq 0\}$, and $Y(t)$ represent the cumulative wear at time t, then this process satisfies the following conditions:

- Initial conditions: $Y(0)=0$;
- Independence of increments: For any $0=t_0<t_1<\cdots<t_n$, the increments $Y(t_i)-Y(t_{i-1})$ are independent of one another;
- Distribution of increments: For any $t<t+\Delta t$, the increments $Y(t+\Delta t)-Y(t)$ follow the inverse Gaussian distribution, i.e., $\Delta Y(t)\sim IG\left(\mu\Delta\Lambda(t),\lambda\Delta\Lambda(t)^2\right)$.

In this case, $\{Y(t)\}$ is called an inverse Gaussian process, and the probability density function of the inverse Gaussian distribution is as follows:

$$f(y\mid\mu,\lambda)=\sqrt{\frac{\lambda\Lambda(t)^2}{2\pi y^3}}\exp\left(-\frac{\lambda\left(y-\mu\Lambda(t)\right)^2}{2\mu^2 y}\right) \tag{1}$$

Accordingly, the mean and variance functions of the inverse Gaussian process are, respectively:

$$E[Y(t)]=\mu\Lambda(t) \tag{2}$$

$$\mathrm{Var}[Y(t)]=\frac{\mu^3}{\lambda}\Lambda(t) \tag{3}$$

In Equations (1)–(3), μ is the drift coefficient; λ is the scale coefficient; and $\Lambda(t)$ is the cumulative degradation function.

As can be seen from Equations (2) and (3), $\Lambda(t)$ determines the expected trajectory and variance evolution of the degradation process and is a core element in modeling IG processes. Once the functional form of $\Lambda(t)$ is determined, the variation pattern of the entire degradation process is also determined [26]. Therefore, the appropriate setting of $\Lambda(t)$ directly determines the model's ability to characterize the actual degradation process. In traditional parametric IG processes, $\Lambda(t)$ is typically assumed to follow a power-law form:

$$\Lambda(t)=\alpha t^{\beta} \tag{4}$$

In Equation (4), α is the scale parameter; β is the shape parameter.

This power-law assumption is mathematically convenient to work with; it describes accelerating decay when $\beta>1$ and decelerating decay when $\beta<1$ and can approximately characterize certain monotonically decaying processes, which is why it is widely used in engineering practice.

However, the power-law assumption imposes strong prior constraints on the wear trajectory, leading to the following limitations in practical applications. First, tool wear typically exhibits a multistage evolution: the wear rate is high initially but gradually slows down; after entering the steady-state wear phase, the rate remains approximately constant; and as failure approaches, the wear rate rises rapidly. A single power-law function struggles to simultaneously fit this segmented trend, leading to systematic biases in the model. Second, due to individual differences among tools—such as coating conditions and initial edge quality—degradation trajectories may exhibit significant shape variations. A fixed form of $\Lambda(t)$ lacks adaptive adjustment capabilities, making it difficult to ensure accuracy in cross-tool prediction scenarios.

2.2. Long-Short-Term Memory Neural Networks

Recurrent Neural Networks (RNN) are widely used in time series forecasting tasks because their ability to retain historical information allows them to effectively capture temporal dependencies. Through recurrent connections, RNN maintain a state vector within their hidden layers and are theoretically capable of capturing long-term dependencies of arbitrary length. However, in practical training, standard RNN face severe issues of gradient vanishing or gradient explosion. As the sequence length increases, the product of gradients during backpropagation may decay or grow exponentially. Consequently, RNN do not strongly capture correlations between time points that are far apart and struggle to provide accurate predictions when faced with long-term declining trends.

LSTM is a variant of RNN that effectively alleviates the aforementioned issues by introducing a gating mechanism and a cell state. The core idea behind LSTM is to design a specialized memory unit that allows information to be selectively passed between time steps, while controlling the forgetting, writing, and reading of information through three gate units. This structure enables LSTM to maintain stable gradient propagation over long time spans, thereby capturing complex temporal evolution patterns. Figure 1 illustrates the internal structure of an LSTM cell. At time step t, the cell's inputs consist of the current input x_t, the hidden state h_{t-1} from the previous time step, and the cell state c_{t-1} from the previous time step. Data flows sequentially through the three gate units according to the following steps:

Forget Gate: Read h_{t-1} and x_t, and output a value between 0 and 1. Multiply this value element-by-element with the cell state c_{t-1} from the previous time step. The closer the output value is to 0, the more the information at that position should be discarded; the closer it is to 1, the more it should be retained. The calculation is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{5}$$

In Equation (5), σ represents the sigmoid activation function; W_f and b_f are the weight matrices and bias terms of the forgetting gate, respectively.

The sigmoid function was chosen because its bounded output provides continuous adjustment coefficients for the LSTM gating mechanism. In the forget gate and input gate, this function controls the degree to which historical information is retained and the degree to which new information is written, respectively, enabling the LSTM to adaptively adjust information updates and thereby learn the time-dependent relationships in the tool wear and degradation process.

The update is performed through the collaboration of the input gate and the candidate cell state. First, the sigmoid layer outputs i_t, which determines which positions need to be updated. Next, the tanh layer generates the candidate cell state $\tilde{c}_t$, which represents the new information at the current time step. These two values are multiplied element-wise and added to the old state—which has been processed by the forget gate—to update the cell state. The calculation process is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{7}$$

$$c_t = f_t c_{t-1} + i_t \tilde{c}_t \tag{8}$$

At this point, the cell's state is updated from c_{t-1} to c_t, preserving key elements of long-term memory while incorporating relevant information from the present moment.

https://doi.org/10.3390/s26227158

The output gate o_t reads h_{t-1} and x_t and outputs a value between 0 and 1, which determines how much information from the updated cell state c_t needs to be passed to the hidden state h_t at the current time step. It is calculated as follows:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t \tanh(c_t) \tag{10}$$

After these three steps, the LSTM cell completes the computation for one time step; c_t carries the long-term memory forward to the next time step, while h_t serves as the output at the current time step and is passed to subsequent network layers. Through the aforementioned gating mechanism, the LSTM can selectively retain or forget long-term information, and gradients can be directly propagated along the temporal dimension via the cell state, effectively avoiding the vanishing gradient problem associated with standard RNNs. Consequently, LSTMs demonstrate superior performance in long-time-series regression tasks such as equipment degradation trend prediction and remaining life estimation.

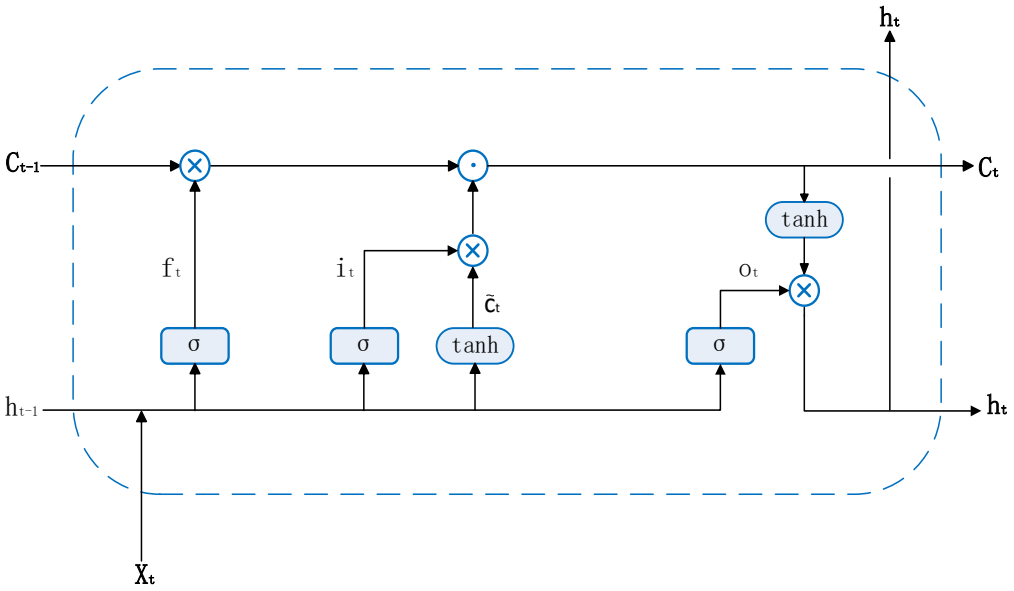


Figure 1. Schematic diagram of the internal structure of an LSTM cell. Structure of the LSTM unit. The circled "×" and "+" symbols represent element-wise multiplication and element-wise addition, respectively.

3. The IG-LSTM Prediction Method

3.1. IG-LSTM Model Framework

As discussed in Section 2.1, the limitations of traditional parametric IG processes primarily stem from the use of a preset form for the cumulative degradation function $\Lambda(t)$. If the form of $\Lambda(t)$ can be adaptively learned from the data, these limitations can be overcome. This paper proposes a neural network-based inverse Gaussian process (IG-LSTM) modeling method. By leveraging the long short-term memory (LSTM) network's ability to fit time-series data, this method learns the cumulative degradation function $\Lambda(t)$ from historical wear data without prescribing any specific functional form. This approach preserves the IG process's ability to capture the non-negativity and stochastic

fluctuations of wear increments while endowing the model with the capacity to adaptively fit complex degradation trends.

The overall framework of the proposed IG-LSTM model is shown in Figure 2; this framework can accommodate both single-tool self-prediction validation and cross-tool transfer validation scenarios. The entire process consists of three core stages: data preprocessing and trend extraction, LSTM trend prediction, and IG-LSTM incremental modeling. First, the maximum value fusion method is applied to the multi-blade wear observations to form a composite wear sequence, and VMD is used to extract the low-frequency degradation trend. Subsequently, the LSTM network is pretrained using this trend as the training target, enabling it to learn the temporal patterns of degradation evolution. After pretraining is complete, the predicted trend output by the LSTM is used as the cumulative degradation function $\Lambda(t)$ in the IG process to construct the incremental model $\Delta Y(t) \sim IG\left(\mu\Delta\Lambda(t), \lambda\Delta\Lambda(t)^2\right)$, and model parameters are determined via maximum likelihood estimation.

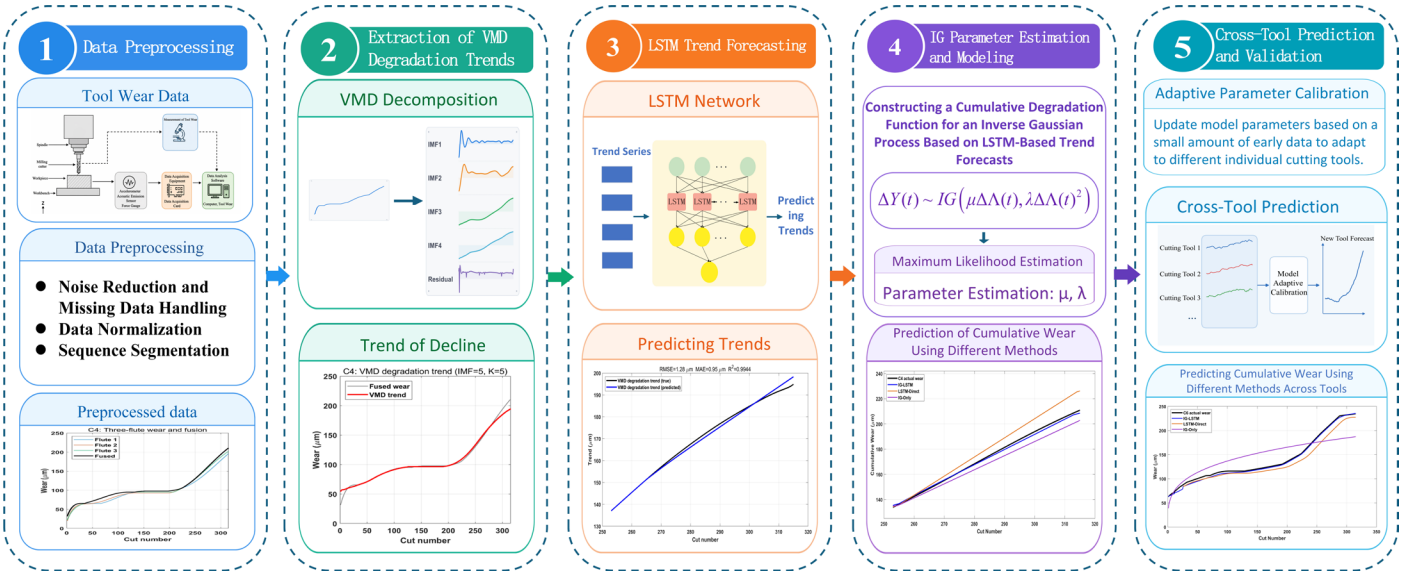


Figure 2. Architecture Diagram of the IG-LSTM-Based Tool Wear Prediction Method.

The only difference in implementation between the two scenarios lies in the composition of the training data and the parameter estimation strategy. In the single-tool self-prediction and validation scenario, historical data from the same tool is used as the training set; both LSTM pre-training and the estimation of parameters μ and λ are performed based on this dataset. Subsequently, the model directly predicts the tool's future wear trend to evaluate its fitting and prediction capabilities for a single degradation trajectory. In cross-tool prediction, the LSTM model is first trained using data from the source tool, and the global parameters μ and λ of the inverse Gaussian (IG) process are estimated based on this. For the target tool, its first 20 effective wear increments are used to calibrate the tool-specific drift parameter μ, while the parameter λ is retained as the globally estimated value. Independent prediction begins after the calibration phase is complete. During the prediction process, the model makes predictions based on the trained LSTM and the calibrated parameter μ, without using subsequent actual wear observations from the target tool to update the model or parameters. Subsequent observations are used solely to evaluate prediction performance. Therefore, RMSE, MAE, and R^2 are calculated only within the evaluation interval following the calibration phase; the

first 20 effective wear increments used for calibration are not included in the final performance evaluation.

The overall framework seamlessly integrates data-driven trend extraction with stochastic degradation modeling through $\Lambda(t)$, forming a complementary predictive architecture. The following sections will elaborate on the mathematical formulations and implementation details of each module.

3.2. Data Preprocessing and Extraction of Degradation Trends

Tool wear data is typically recorded as independent measurements for each cutting edge. During actual cutting operations, wear on the individual cutting edges does not occur entirely in sync; a particular edge may wear faster due to uneven force distribution or localized defects. If the average value or the data from a single cutting edge is directly used as an indicator of wear, this either smooths out information on extreme wear or obscures the overall wear trend of the remaining cutting edges. Therefore, this paper employs a maximum-value fusion strategy, selecting the maximum wear measurement from each edge during each cutting operation as the composite wear value at that moment:

$$\overline{y}(t) = \max\left\{y_1(t), y_2(t), \ldots, y_m(t)\right\} \tag{11}$$

In Equation (11), t represents the number of cutting operations. The rationale behind this approach is that the overall health of the cutting tool is limited by its most severe localized damage. At the same time, maximum value fusion preserves the most significant change signals during the degradation process, thereby preventing the features of late-stage rapid degradation from being obscured by mean smoothing. This approach is consistent with the rules for determining the number of safe cutting passes in the PHM2010 data competition, which defines the number of safe cutting passes based on the maximum wear value of any single cutting edge rather than the average wear value across all cutting edges. Chen [27] conducted a study on tool wear prediction based on the PHM2010 dataset and compared regression modeling based on the maximum wear value with regression modeling based on the independent wear values of each cutting edge. His experimental results indicate that, under the experimental conditions he employed, modeling the wear of each cutting edge separately did not demonstrate any advantage over modeling based on the maximum wear value. These studies provide a basis for using the maximum wear value to characterize the condition of the most severely worn cutting edge.

The fused wear sequence remains a non-smooth curve recorded point by point. The essence of tool wear is the irreversible accumulation of damage; its long-term trend should be smooth and monotonically increasing, but the measured data often exhibit random fluctuations between successive cuts. These fluctuations are not measurement noise but rather a true reflection of the inherent randomness of the cutting process itself. However, to establish a degradation trend model, it is necessary to extract the main trend that determines tool life from these fluctuations, rather than fitting each local fluctuation point by point. To this end, this paper employs VMD to decompose the fused sequence in order to extract the degradation trend and filter out random fluctuations.

Compared to recursive methods such as Empirical Mode Decomposition (EMD) [28], VMD incorporates the decomposition process into a unified optimization framework, effectively suppressing modal aliasing and endpoint effects, and yielding decomposition results that are more robust to noise and sampling frequency. In tool wear signal processing, VMD can decompose the fused wear time series into multiple modes based on frequency scales, where the low-frequency components reflect the long-term degradation trend of wear, and the high-frequency components correspond to random fluctuations between successive cuts.

Variational Mode Decomposition (VMD) is a fully non-recursive adaptive signal decomposition method that decomposes the original signal $f(t)$ into a sum of band-limited intrinsic mode functions (IMFs) centered at their respective center frequencies by solving a variational problem, thereby effectively separating different frequency components [29,30]. Its core idea is to transform the decomposition problem into a constrained variational optimization problem. Under the constraint that the sum of all modes equals the original signal, the sum of the estimated bandwidths of each mode is minimized. For a given signal $f(t)$, VMD constructs the following variational model:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \quad \text{s.t.} \sum_{k=1}^{K} u_k(t) = f(t) \tag{12}$$

In Equation (12), k is the preset number of modal decompositions, $\{u_k\} = \{u_1, \ldots, u_K\}$ is the IMF component obtained from the decomposition, $\{\omega_k\} = \{\omega_1, \ldots, \omega_K\}$ is the center frequency corresponding to each component, $\delta(t)$ is the Dirac delta function, and $*$ denotes the convolution operation.

To solve the above constrained optimization problem, VMD introduces a quadratic penalty term and Lagrange multipliers to transform the constrained variational problem into an unconstrained extended Lagrangian form, and then uses the Alternating Direction Method of Multipliers (ADMM) to iteratively update each mode and center frequency. In each iteration, u_k, ω_k and λ are updated sequentially until the convergence criteria are satisfied.

In actual measurements, tool wear sequences are subject to interference from cutting vibrations and measurement errors, and are interspersed with a large number of high-frequency random fluctuations; direct modeling would result in severe fluctuations in the predicted output. VMD alone cannot automatically distinguish between the degradation trend components and the noise components; therefore, this paper introduces the zero-crossing rate criterion to achieve adaptive filtering of IMF components. For a modal component $u_k(n)$ of length N, the zero-crossing rate ZCR_k is calculated as follows:

$$ZCR_k = \frac{1}{2(N-1)} \sum_{n=1}^{N-1} \left| \operatorname{sgn}\left[u_k(n+1)\right] - \operatorname{sgn}\left[u_k(n)\right] \right| \tag{13}$$

In this equation, $\operatorname{sgn}(\cdot)$ is the sign function:

$$\operatorname{sgn}[x] = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}$$

A zero sample point inherits the sign of the preceding non-zero sample point; the first zero at the beginning of the sequence takes the sign of the first subsequent non-zero sample point. If all sample values after averaging are zero, the cross-rate is set to zero. The 1/2 in the formula corrects for the difference of 2 caused by a change in sign, while N − 1 normalizes the number of adjacent sample point pairs. Averaging is used only for calculating the cross-zero rate and does not alter the original modal characteristics. The trend of wear and degradation changes gradually, with the signal crossing zero infrequently, corresponding to a low zero-crossing rate ZCR_k value; high-frequency noise disturbances exhibit violent oscillations, resulting in significantly higher zero-crossing rate values. Calculate the zero-crossing rates for all K IMF components, Adaptive thresholds are

used for modal filtering. For a wear sequence containing N data points, the threshold is defined as:

$$\tau = \frac{2}{N}$$

The corresponding set of retained modalities is defined as:

$$\mathcal{S} = \{k | \ \mathrm{ZCR}_k < \tau, \quad 1 \le k \le K\}$$

If no mode meets the above screening criteria, the mode with the lowest zero-crossing rate is selected as a backup to ensure that at least one IMF is retained. Finally, the main trend of wear degradation is reconstructed by summing the screened original IMF components with the residual terms obtained from the VMD decomposition:

$$\hat{f}(t) = r(t) + \sum_{k \in \mathcal{S}} u_k(t) \tag{14}$$

In Equation (14), $u_k(t)$ represents the kth original IMF component, and $r(t)$ represents the residual term returned by the VMD decomposition.

After filtering, a smoothed main trend $\hat{f}(t)$ of wear degradation is obtained, which removes high-frequency random noise from the original signal. This trend is then used as the input for the subsequent IG-LSTM hybrid model, thereby reducing the impact of random disturbances on the model and improving the stability and accuracy of tool wear prediction.

3.3. LSTM Trend Forecasting

After extracting the low-frequency decay trend, this section uses a Long Short-Term Memory (LSTM) network to predict its future evolution. Due to its advantages in modeling temporal dependencies, the LSTM can effectively capture the nonlinear evolution patterns in trend sequences. Given the non-stationary nature of the original trend sequence, using absolute values directly as the prediction target can easily lead to error accumulation in recursive forecasting. Therefore, this paper adopts a differential space modeling strategy: first, the trend sequence is first-order differenced to transform the prediction target into the incremental change between adjacent time steps, thereby reducing the sequence's non-stationarity. The differenced sequence is then fed into the LSTM network for training and forecasting. Finally, the predicted increments are summed term by term to reconstruct the trend forecast at the original scale. This strategy suppresses sequence drift through the difference transformation while preserving the LSTM's ability to fit complex nonlinear relationships, thereby balancing forecast stability and accuracy. Let $\Lambda(1), \Lambda(2), \ldots, \Lambda(N)$ denote the extracted degraded trend; its first-order difference is given by:

$$\Delta\Lambda(t) = \Lambda(t) - \Lambda(t-1), \quad t = 2, 3, \ldots, N \tag{15}$$

The values of the differenced sequence $\Delta\Lambda(t)$ are typically one order of magnitude smaller than those of $\Lambda(t)$ and fluctuate around a certain mean, making them statistically more stationary. The rate at which LSTM accumulates extrapolation error on stationary sequences is much lower than on non-stationary sequences; this is the fundamental rationale for performing predictions in the differenced space. To convert the differenced sequence into a supervised learning format, set the window length to w; the input for the t-th sample is the differenced values from the past w steps, and the output is the differenced values for the next h steps:

$$x_t = \left[\Delta\Lambda(t-w+1), \Delta\Lambda(t-w+2), \ldots, \Delta\Lambda(t)\right] \tag{16}$$

$$y_t = \left[\Delta\Lambda(t+1), \Delta\Lambda(t+2), \ldots, \Delta\Lambda(t+h)\right] \tag{17}$$

The LSTM network consists of three stacked layers of LSTM cells. The three-layer structure is designed such that the first layer captures local temporal dependencies, the second layer combines these local patterns into trend segments over a broader time span, and the third layer compresses the degradation information across the entire life cycle into a global feature representation. Based on a sliding window, the network predicts the next h differential values during each forward pass. The two experimental scenarios employ different prediction strategies: in the single-tool prediction scenario, h = 1; the model predicts one increment at a time and feeds the prediction back into the input window to continue predicting the next time step; in the cross-tool prediction scenario, h = 8; the model directly outputs the next 8 increments during each forward pass. When the prediction interval exceeds 8 steps, the output of the current prediction block is used to construct the input for the next prediction block, and this process is repeated until the entire prediction interval is covered. Finally, using the last available trend value at the start of the prediction as a reference, the predicted differential values are accumulated to reconstruct the cumulative degradation function $\Lambda(t)$, which is then used for subsequent IG process modeling.

3.4. Modeling Wear Increment and Parameter Estimation Based on the IG Process

The $\Lambda(t)$ value obtained from the LSTM prediction describes the direction of the degradation trend, but it is not, in itself, a prediction of the actual wear measurements; actual wear exhibits random fluctuations around the trend. This paper introduces an inverse Gaussian process to describe these fluctuations, treats $\Lambda(t)$ as its cumulative degradation function, and establishes a probabilistic relationship between the trend and the wear increment.

For the inverse Gaussian degradation process, let the wear increments Δy_i at each time step be independent and follow an inverse Gaussian distribution $\mathrm{IG}(\mu\Delta\Lambda_i, \lambda(\Delta\Lambda_i)^2)$, whose probability density function is:

$$f(\Delta y_i; \mu, \lambda) = \sqrt{\frac{\lambda(\Delta\Lambda_i)^2}{2\pi(\Delta y_i)^3}} \exp\left\{-\frac{\lambda\left(\Delta y_i - \mu\,\Delta\Lambda_i\right)^2}{2\mu^2 \Delta y_i}\right\} \tag{18}$$

Before constructing the likelihood function, the wear sequence is converted into a non-decreasing sequence using the cumulative maximum operation, and wear increments that result in zero after processing are excluded. The degradation trend sequence is also made monotonic, with a minor adjustment of 10^{-8} added when adjacent values are equal. Only paired samples where both the wear increment and the degradation trend increment are positive are retained for IG parameter estimation. Original negative wear increments are counted separately prior to preprocessing, and the joint likelihood function is ultimately constructed using the valid paired samples.

Assuming that the samples of wear increments at each time step are independent of one another, we construct the overall joint likelihood function for all M sets of observed samples:

$$\ln L(\mu,\lambda) = \frac{M}{2}\ln\lambda + \sum_{i=1}^{M}\ln(\Delta\Lambda_i) - \frac{1}{2}\sum_{i=1}^{M}\ln(2\pi(\Delta y_i)^3) - \frac{\lambda}{2}\sum_{i=1}^{M}\frac{(\Delta y_i - \mu\,\Delta\Lambda_i)^2}{\mu^2\Delta y_i} \quad (19)$$

Compute the first-order partial derivatives of the log-likelihood function with respect to the parameters μ and λ, respectively; set these partial derivatives equal to zero; construct the system of equations for finding the extrema: $\frac{\partial \ln L}{\partial \mu} = 0$ and $\frac{\partial \ln L}{\partial \lambda} = 0$; and solve to obtain estimates for the two parameters.

In cross-tool scenarios, the above estimation yields global parameters μ and λ, which represent the common patterns across multiple training tools [31]. However, the wear rates of test tools may exhibit systematic deviations from those of training tools; these deviations stem from individual differences such as tool material, coating condition, and initial cutting edge quality. If left uncorrected, the predicted curve may be generally too high or too low. To address this, this paper performs a tool-specific calibration of μ using early observation data from the tool under test. After calculating the local wear rate and robustly removing outliers, the median of the remaining samples is taken as the μ value for that tool.

In a single-tool self-predictive scenario, the training and test data come from the same tool, and there is no cross-tool bias in the degradation rate; therefore, no additional calibration of μ is required.

Once the above steps are completed, the parameters μ and λ of the IG process are fully determined. Next, based on these parameters and the $\Lambda(t)$ value predicted by the LSTM, a probabilistic model for tool wear increment can be constructed.

4. Experimental Validation

4.1. Experimental Dataset and Data Split

To validate the performance of the proposed IG-LSTM model, this paper uses the PHM2010 high-speed milling tool wear dataset [32]. A schematic diagram of the experimental setup is shown in Figure 3. The experiments were conducted on a Roders-Tech RFM760 high-speed CNC machine tool (Röders, Soltau, Germany), using an Inconel 718 alloy workpiece and a three-tooth ball-nose carbide milling cutter under dry milling conditions. The main experimental configuration is shown in Table 1. A Kistler force transducer and accelerometer (Kistler Instrumente AG, Winterthur, Switzerland) were used to acquire a total of 7 channels of data—including cutting forces in the X, Y, and Z directions, vibrations in all three directions, and acoustic emission signals—at a sampling frequency of 50 kHz. After each pass, the width of the rake face wear on the three cutting edges was measured using a LEICA M12Z microscope (Leica Microsystems, Wetzlar, Germany). To characterize the overall wear condition of the tool, this paper employs a maximum-value fusion strategy, in which the maximum rake face wear width measured across the three cutting edges is taken as the tool wear value corresponding to that pass and is used as the target variable for subsequent degradation trend extraction and tool wear prediction. Ultimately, wear data from three cutting tools—C1, C4, and C6—were selected as the dataset, each containing sensor data from 315 pass cycles and the corresponding target tool wear values.

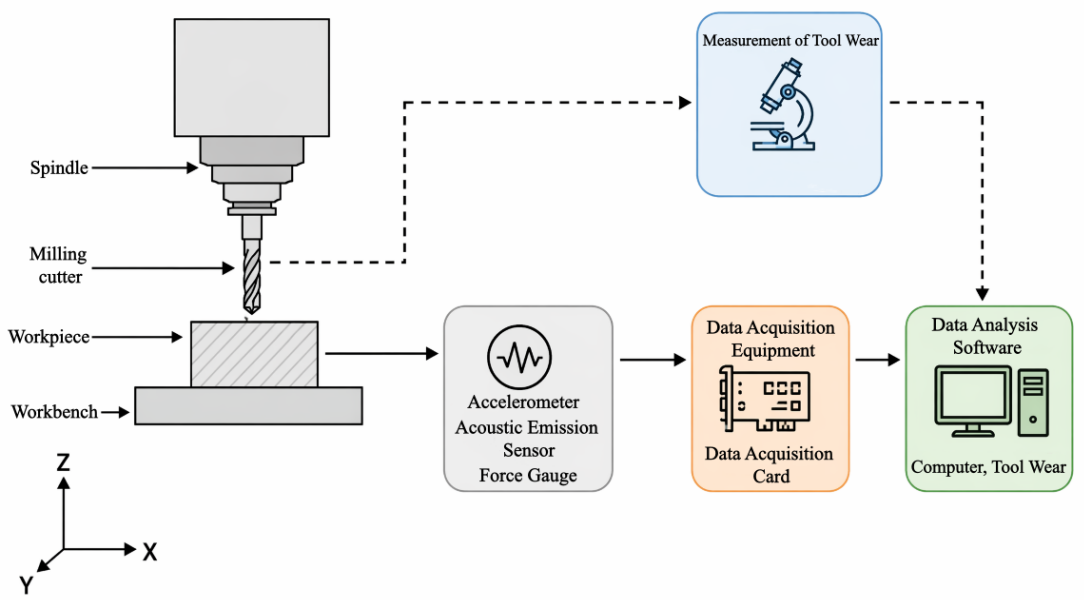


Figure 3. Schematic Diagram of the Experimental Platform.

Table 1. Experimental Platform and Cutting Parameters.

Hardware	Model and Key Specifications	Cutting Conditions	Parameters
CNC Milling Machine	High-Speed CNC Machine Tool Roders Tech RFM760	Spindle Speed (r/min)	10,400
Force Gauge	Kistler 9265B Triaxial Force Gauge	Feed Rate (mm/min)	1555
Charge Amplifier	Kistler 9019A Multichannel	Axial Cutting Depth (mm)	0.2
Materials for Milling	Inconel 718, Rectangle	Mirrored Cutting Width (mm)	0.125
Data Acquisition Card	NI DAQ Data Acquisition Card	Sampling Frequency (kHz)	50
Wear Gauge	LEICA MZ12 Microscope	Cooling Conditions	Dry-cutting

To comprehensively evaluate the model's performance across different application scenarios, this paper designed two sets of experiments corresponding to two data partitioning strategies. The first set consists of single-tool self-prediction experiments. Taking tool C4 as the subject, its 315 data sets were divided chronologically into a training set and a test set: the first 80% was used for model training, and the remaining 20% was used for performance testing. This experiment was conducted to validate the model's basic predictive capability when both the training and test data originate from the same tool. The second set consists of cross-tool transfer prediction experiments. The training set comprises all data from tools C1 and C4, while the test set consists of all data from tool C6. Since the training and test sets originate from different tools, this experiment is used to verify the model's generalization ability across different individual tools. Table 2 summarizes the data partitioning schemes for the two sets of experiments.

Table 2. Experimental Data Categorization Scheme.

Experimental Scenarios	Training Tools	Testing Cutting Tools
Single-Tool Prediction	C4 (top 80%)	C4 (bottom 20%)
Cross-Tool Prediction	C1 + C4	C6

To ensure that the model's generalization ability across tools is evaluated independently, this paper divides the process into distinct phases: model training, hyperparameter selection, target tool calibration, and final testing. Tools C1 and C4 were selected as source tools, and tool C6 as the target tool for final evaluation. Hyperparameters—

including VMD parameters, modal selection rules, LSTM network architecture, input window length, prediction step size, learning rate, and number of training iterations—were selected based solely on the source tool data. Before training the final model using the specified training data from C1 and C4, source-domain validation was completed, and the selected model configuration was fixed prior to the start of the target tool evaluation.

In cross-tool experiments, only the first 20 valid wear increments of C6 are used to calibrate the tool-specific drift parameter μ, while the global parameter λ remains unchanged. This calibration phase is considered a predefined step for tool-specific adaptation, rather than a hyperparameter selection phase. After calibration is complete, subsequent observations from C6 are used solely for the final prediction performance evaluation and are no longer used to select model configurations, adjust preprocessing rules, or update model parameters. Therefore, RMSE, MAE, and R^2 are calculated only within the evaluation interval following the calibration phase. Throughout the evaluation process, source-domain model selection, target tool parameter calibration, and final performance testing are always clearly distinguished.

4.2. Validation of the Effectiveness of VMD Trend Extraction

The quality of the trend extracted by VMD directly affects the performance of subsequent LSTM predictions and IG modeling. If the extracted trend contains excessive high-frequency noise, the LSTM will waste capacity on irrelevant components; if the trend is overly smoothed and loses the true degradation inflection points, the variance estimate of the IG process will be underestimated. This section verifies the effectiveness of VMD trend extraction based on decomposition results and screening performance.

The number of modes in the VMD decomposition was set to $K = 5$, and the bandwidth penalty coefficient to $\alpha = 2000$. By comparing the decomposition results obtained under different parameter settings and comprehensively evaluating both the effectiveness of modal separation and the ability to extract long-term wear degradation trends, the above parameters were ultimately determined. Figure 4 shows the VMD decomposition results for the three cutting tools (C1, C4, and C6) and the zero-crossing rates of each IMF component. It can be seen that the combined wear sequences for each cutting tool were decomposed into five IMF components and a residual term. Judging from the trends in the zero-crossing rates of the IMFs in Figure 4, the zero-crossing rates gradually decrease as the order increases. IMF1 through IMF3 exhibit higher zero-crossing rates and more intense waveform oscillations, corresponding to high-frequency random fluctuations and measurement noise during the cutting process. In contrast, the zero-crossing rates of IMF4 and IMF5 decrease significantly; their waveforms are smooth and change gradually, and together with the residual term, they reflect the long-term degradation trend of wear. By filtering based on the zero-crossing rate threshold, the high-frequency components are removed, and the low-frequency components are retained and superimposed with the residual term, thereby reconstructing the smooth degradation trend for each cutting tool. The three tools—C1, C4, and C6—exhibited consistent filtering patterns, indicating that this strategy is not dependent on specific tools and has good applicability. The above results demonstrate that VMD trend extraction can effectively separate high-frequency noise from low-frequency degradation trends, providing a reliable training target for subsequent LSTM predictions.

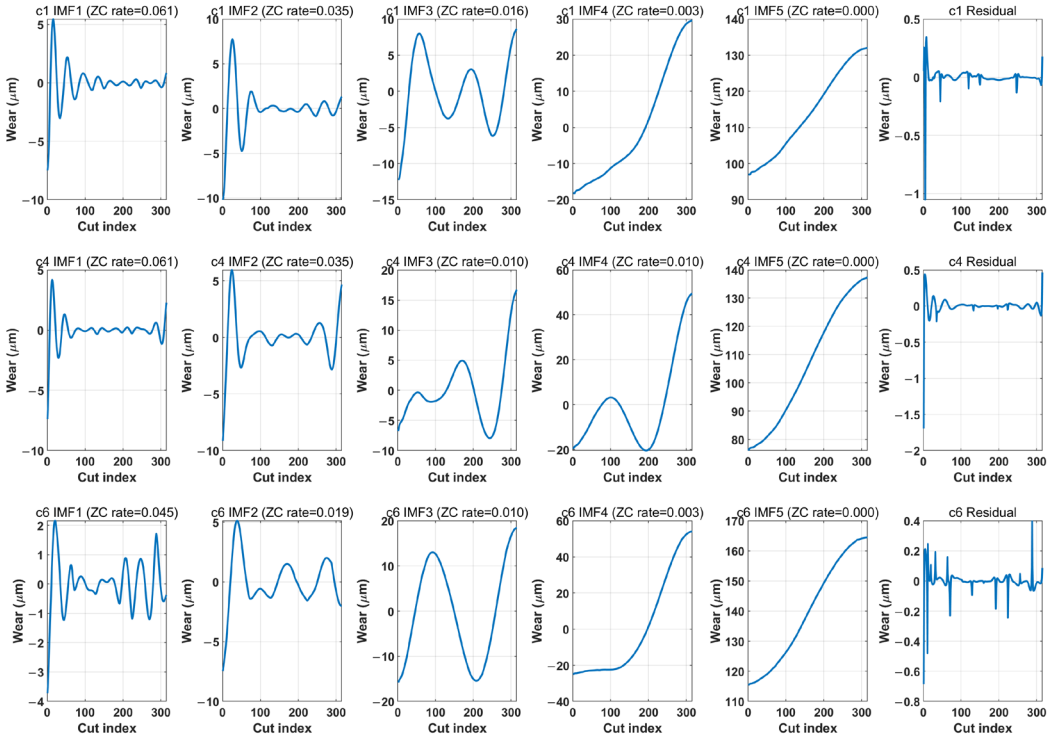
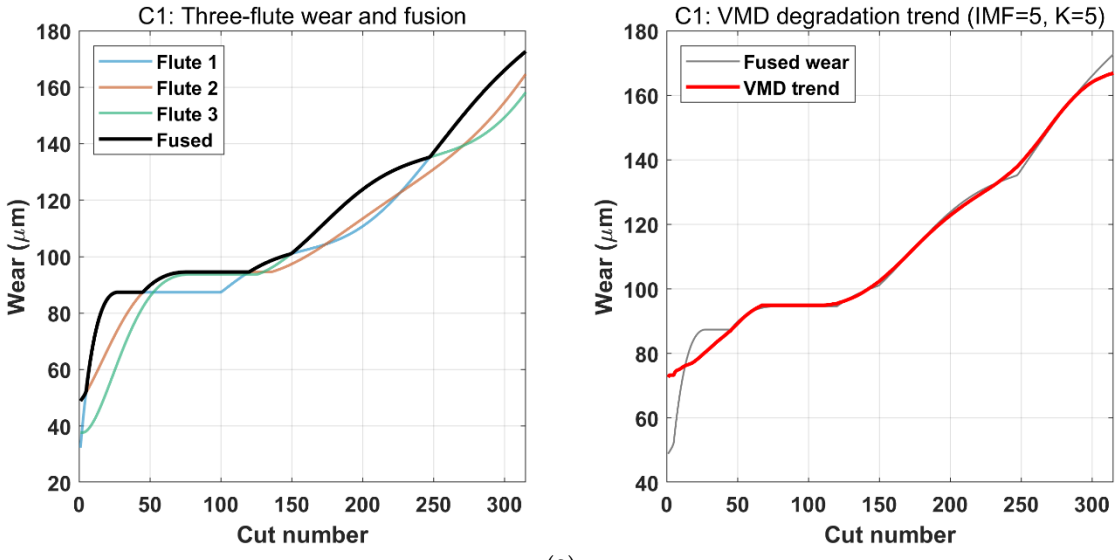


Figure 4. VMD decomposition results for the wear sequences of tools C1, C4, and C6.

Figure 5 shows a comparison between the combined wear curves of the three cutting tools (C1, C4, and C6) and the extracted degradation trends. The trends for all three cutting tools fully preserve the low-frequency evolution patterns of the original sequences while effectively eliminating sequential fluctuations. In terms of correlation coefficients, the VMD trend for C1 has a correlation coefficient of 0.94 with the fused wear sequence, 0.93 for C4, and 0.89 for C6. The correlation coefficients for all three cutting tools exceed 0.85 [33], indicating that the extracted trends maintain a high degree of consistency with the original data in the low-frequency range.



(**a**)

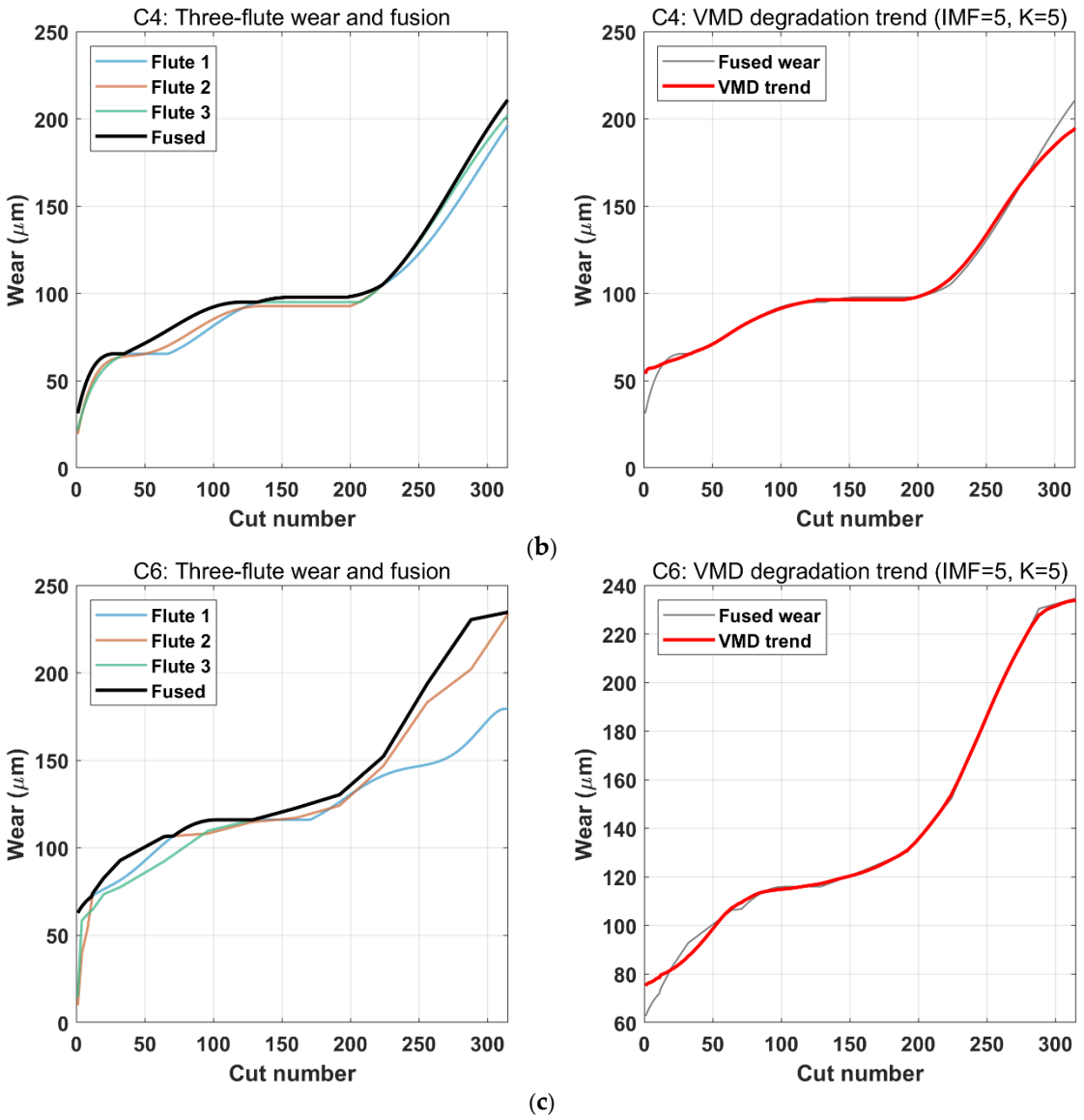


Figure 5. Comparison of the Fused Wear Curve with the VMD Degradation Trend: (**a**) comparison results for tool C1; (**b**) comparison results for tool C4; (**c**) comparison results for tool C6.

Overall, VMD trend extraction achieved the expected results in two respects: first, it effectively filtered out local fluctuations between successive cuts in the wear sequence, preserving the main trend pattern; second, the correlation coefficients for all three tools were above 0.85, indicating that the extracted trends maintained sufficient consistency with the original data. These results demonstrate that the extracted degraded trends can serve as reliable training targets for subsequent LSTM predictions.

4.3. Validation of the Effectiveness of LSTM Trend Forecasting

The main parameters of the LSTM network are shown in Table 3.

Table 3. Key Parameters of the LSTM Network.

Parameter Categories	Parameter Name	Single-Tool	Across Tools
Model Structure Parameters	Time Step	12	25
	First Layer (LSTM)	64	100
	Second Floor (LSTM)	64	64
	Third Floor (LSTM)	32	32

	Fully Connected Layer	16	16
	Output Layer	1	8
	Activation Function	tanh	tanh
Model Training Parameters	Learning Rate	0.001	0.0005
	Maximum Number of Iterations	4000	1200
	Loss Function	Root Mean Square Error	Root Mean Square Error
	Learning Rate Decay Cycle	600	300
	Learning Rate Decay Factor	0.5	0.5
	Optimizer	Adam	Adam
Input Parameters	Input Variables	Incremental Wear Trend	Incremental Wear Trend
	Time Step	12	25
Output Parameters	Output Variables	Incremental Future Wear Trends	Incremental Future Wear Trends
	Output Step Size	1	8

The hyperparameters of the LSTM model were determined based on the training and validation performance using data from the source tools (C1 and C4). The selected model configuration was fixed prior to conducting cross-tool predictions for tool C6. During this process, no adjustments were made to the hyperparameters based on the prediction results for tool C6.

To analyze the impact of key hyperparameters on the predictive performance of IG-LSTM, we employed the One-At-A-Time (OAT) method, changing one parameter at a time while keeping the others constant. We selected window length, learning rate, prediction step size, and the number of LSTM hidden units for testing; the baseline parameters from the paper were 25, 0.0005, 8, and 100, respectively. The experiments used C1 and C4 tool data as the source domain training data, and the predictive performance of different parameter configurations was evaluated on C6 tools, with the random seed fixed at 42.

Figure 6 shows the coefficient of determination (R^2) for different parameter configurations. The blue and gray curves represent the results of cumulative wear prediction and the degradation trend function $\Lambda(t)$, respectively, while the orange dots indicate the baseline parameters used in the paper.

The results show that different hyperparameters have varying degrees of impact on the final wear prediction performance. When the window length was increased from 25 to 40 or 50, the cumulative wear prediction performance declined significantly; changes in the learning rate and prediction step size also caused some fluctuations in performance; and when the number of LSTM hidden units was increased from 64 to 100, the prediction performance improved significantly. It is worth noting that the prediction accuracy of $\Lambda(t)$ does not always align with the final cumulative wear prediction accuracy, indicating that trend fitting performance does not fully represent the model's ability to predict final wear; therefore, cumulative wear prediction metrics should serve as the primary evaluation criterion. Overall, the parameter configuration adopted in this study yields satisfactory prediction results, though it is not the optimal combination of all parameters. This experiment further reveals the model's sensitivity to different hyperparameters, providing supplementary evidence for evaluating the rationality of parameter configurations.

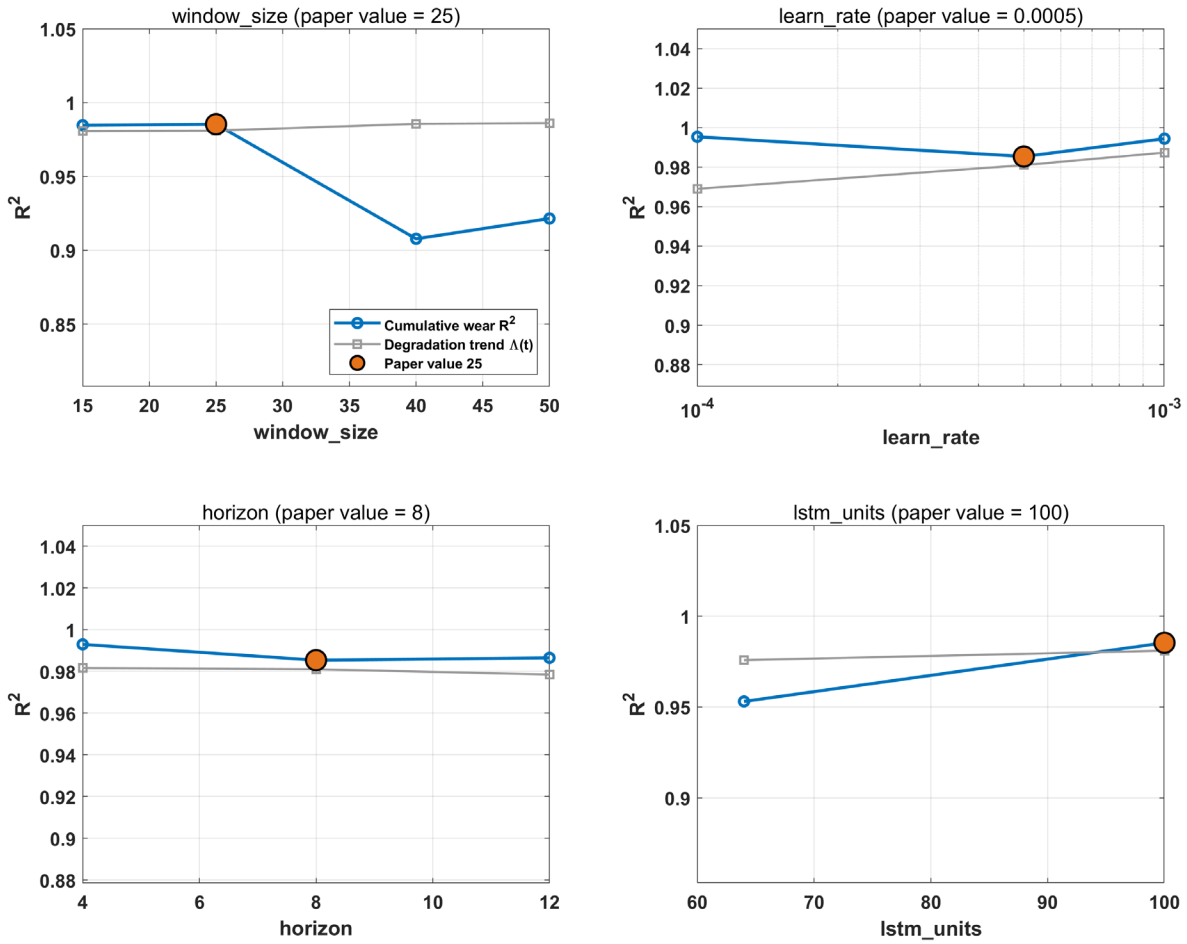


Figure 6. Hyperparameter sensitivity analysis of the IG-LSTM model.

The effectiveness of LSTM trend forecasting was validated through two sets of experiments: single-tool self-forecasting and cross-tool transfer forecasting. The former tests the LSTM's ability to learn and forecast the time-series patterns of degradation for a single tool, while the latter validates the LSTM's ability to extrapolate unknown degradation trends across different tools. Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R^2) were used as metrics to evaluate forecasting performance. These three metrics are defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{20}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{22}$$

In Equations (22)–(24), n is the total number of samples in the test set; y_i is the actual wear measurement for the i-th sample; $\hat{y}_i$ is the model's predicted wear value for the i-th sample; and $\overline{y}$ is the arithmetic mean of all actual wear values; The prediction results for the two sets of experiments are presented below.

The single-tool prediction uses the first 80% of C4 data to train an LSTM network, which then predicts the degradation trend for the remaining 20%. The first 252 tool passes serve as the training set, and the subsequent 63 tool passes serve as the test set. For the test set, the predicted RMSE is 1.28 μm, the MAE is 0.95 μm, and the R^2 is 0.92. Figure 7 shows the recursive prediction results for the final 20% of the C4 run. The predicted curve generally aligns with the actual trend across the entire test set, effectively capturing the evolution of the degradation trend and providing reliable long-term extrapolation results.

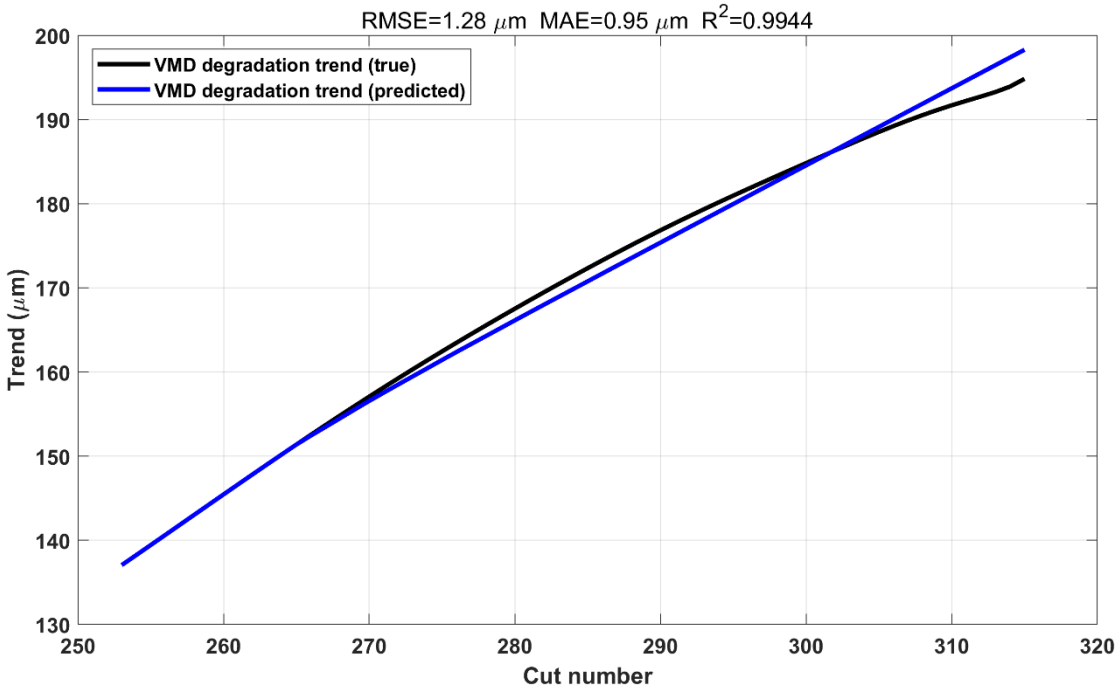


Figure 7. LSTM prediction results on the 20% test set after C4.

An LSTM was trained using all data from tools C1 and C4 for cross-tool prediction and directly applied to predict the full-cycle degradation trend of tool C6. Figure 8 shows the LSTM recursive prediction results for tool C6. On the C6 test set, the RMSE for the recursive multi-step prediction was 3.99 μm, the MAE was 3.14 μm, and the R^2 was 0.99. This indicates that the model still maintains a strong ability to capture trend patterns in a cross-tool scenario.

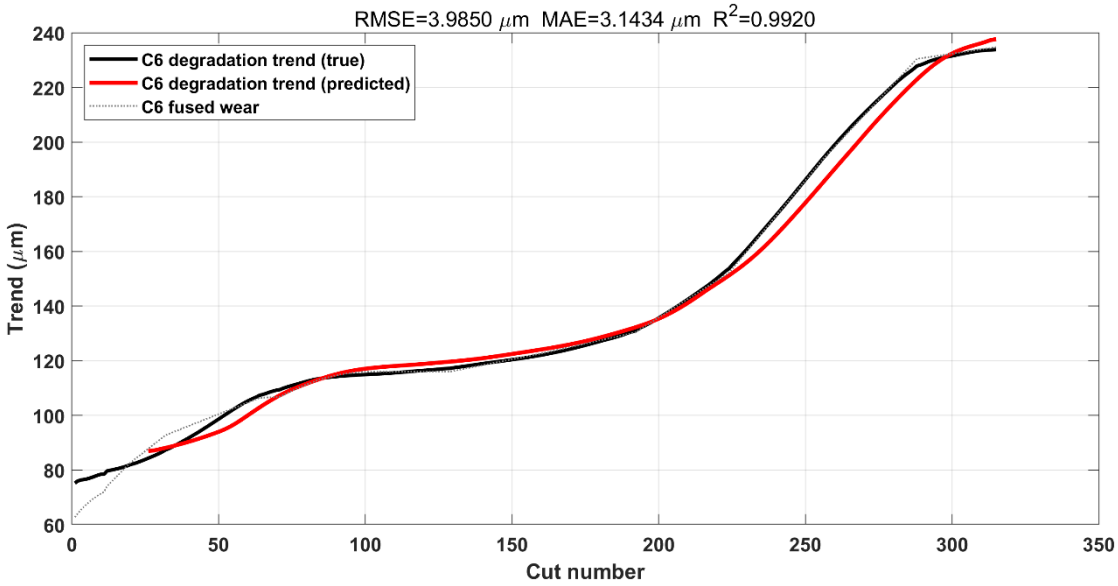


Figure 8. LSTM prediction results for C6 cross-tool data.

4.4. Results of IG Parameter Estimation

The LSTM predictions describe the direction of the degradation trend, but to convert them into a probabilistic description of the increment in wear, the parameters μ and λ of the IG process must be determined. This section presents the parameter estimation results for two types of scenarios.

In the single-tool self-prediction scenario, the first 80% of the data for tool C4 was selected, comprising a total of 252 wear observations corresponding to 251 candidate increments. Among these, 55 increments had a zero wear increment, and 63 increments had a zero degradation trend increment; the number of overlapping samples between the two categories was 36. After removing the overlapping samples, a total of 82 paired samples were excluded, leaving 169 valid paired samples that satisfied the condition of having both positive wear increments and positive degradation trend increments. Maximum likelihood estimation was performed based on these valid samples, yielding the parameters $\mu = 1.27$, $\lambda = 5.50$ Across all tooling scenarios, the complete datasets for C1 and C4 yielded a total of 628 candidate increments. After cumulative maximum preprocessing, 60 zero-wear increments were identified in C1 and 55 in C4. Following joint screening of wear increments and degradation trend increments, 513 valid paired samples were ultimately retained for parameter estimation, yielding the parameters $\mu = 1.35$, $\lambda = 6.68$ were obtained. For the test tool C6, specificity calibration was performed using its first 20 valid increment samples; after robust outlier removal, the median was calculated, resulting in $\mu_{c6} = 1.40$. The parameter estimation results for both scenarios are summarized in Table 4.

Table 4. Parameter Estimation Results.

Scene	Parameters	Estimated Value	Valid Samples	Data Sources
Single-tool	μ	1.27	169	Top 80% of C4
	λ	5.50	169	Top 80% of C4
Across Tools	μ	1.35	513	All of C1 + C4
	λ	6.68	513	All of C1 + C4
Across Tools	μ	1.40	20	Early Data on C6

4.5. Results and Analysis of Cumulative Wear Prediction Using IG-LSTM

To comprehensively evaluate the predictive performance of the IG-LSTM model, two validation scenarios were set up. The first was a single-tool self-prediction scenario, used to validate the model's basic predictive performance on in-training data. The second was a cross-tool transfer prediction scenario, used to validate the model's generalization ability across different individual tools. For comparison, both the LSTM and IG methods were included in the experiments. The former is a purely data-driven method that directly uses wear values as the prediction target without incorporating stochastic process modeling; the latter employs traditional stochastic process modeling, utilizing a cumulative degradation function with a fixed power-law form without incorporating LSTM trend prediction. Prediction performance was quantitatively evaluated using three metrics: RMSE, MAE, and R^2.

Figure 9 shows the cumulative wear prediction results for the C4 test set under a single-tool self-prediction scenario. As shown in the figure, the prediction curve of IG-LSTM largely coincides with the actual wear curve across the entire test range. In terms of quantitative evaluation, IG-LSTM achieved an RMSE of 1.42 μm, an MAE of 1.24 μm, and an R^2 of 0.9962. By comparison, the LSTM method achieved an RMSE of 8.11 μm, an MAE of 6.75 μm, and an R^2 of 0.8762; the IG method achieved an RMSE of 5.73 μm, an MAE of 5.04 μm, and an R^2 of 0.9381. In contrast, the IG-LSTM model reduced the RMSE by 82.5% and 75.2%, respectively, reduced the MAE by 81.6% and 75.4%, respectively, and increased R^2 by 12.0% and 5.8%, respectively, indicating that the proposed IG-LSTM model significantly improved prediction accuracy.

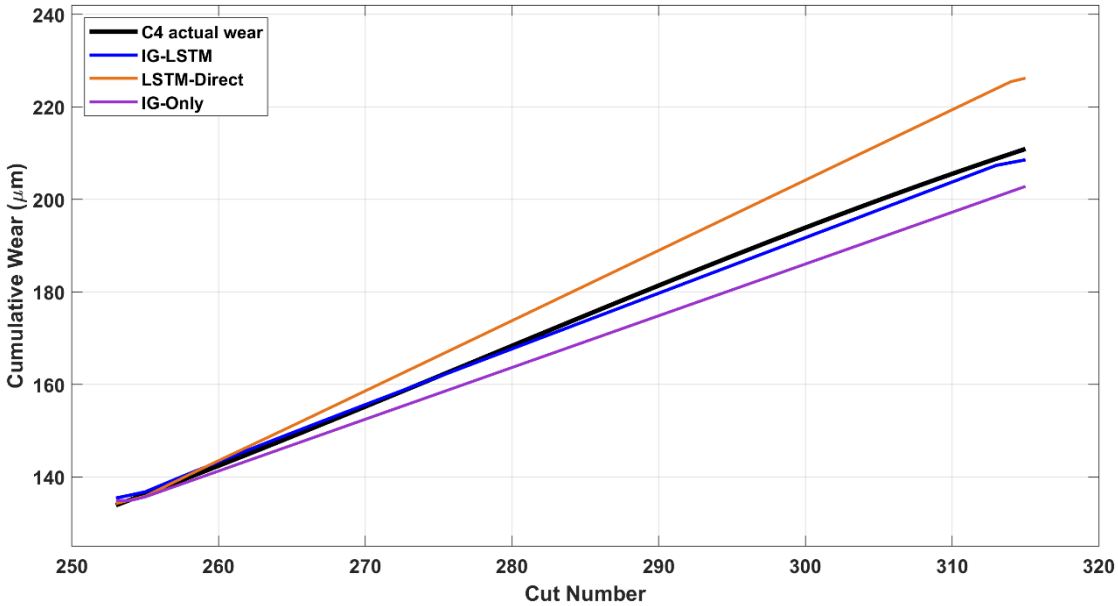


Figure 9. Cumulative wear prediction curves for different methods in a single-tool scenario (ablation comparison).

Figure 10 shows the cumulative wear prediction results for the C6 cutting tool over its entire service life across various cutting scenarios. As shown in the figure, the prediction curve of the IG-LSTM method generally aligns with the actual wear trend throughout the entire service life. In terms of quantitative evaluation, the IG-LSTM method achieved an RMSE of 3.26 μm, an MAE of 2.83 μm, and an R^2 of 0.9951. By comparison, the LSTM method had an RMSE of 10.62 μm, an MAE of 9.31 μm, and an R^2 of 0.9430; the IG method had an RMSE of 27.94 μm, an MAE of 25.25 μm, and an R^2 of 0.6049. In contrast,

the IG-LSTM method reduced the RMSE by 69.3% and 88.3%, respectively, reduced the MAE by 69.6% and 88.8%, respectively, and increased R^2 by 5.2% and 39.0%, respectively, thereby validating the generalization capability of the proposed method across different cutting tool scenarios.

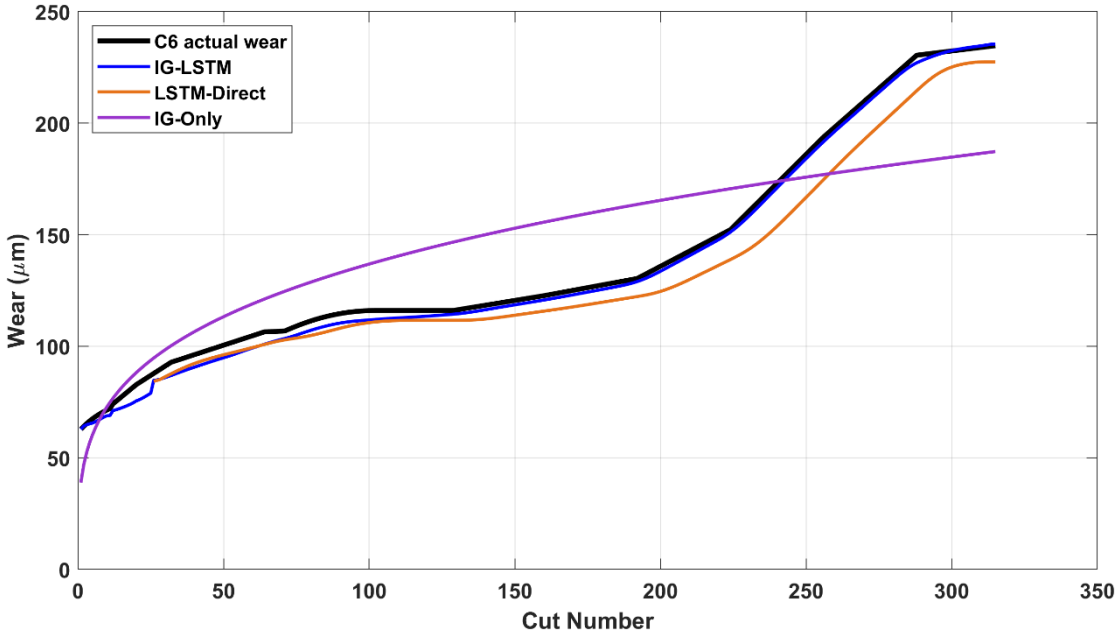


Figure 10. Cumulative wear prediction curves for different methods in a cross-tool scenario (ablation comparison).

To further evaluate the IG-LSTM model's ability to characterize prediction uncertainty, 95% prediction intervals were constructed based on a probability distribution derived from the inverse Gaussian process. Specifically, a Monte Carlo simulation method was employed to repeatedly generate random wear increments and accumulate them incrementally, yielding multiple possible cumulative wear degradation trajectories. Based on this, the 2.5th and 97.5th percentiles of the simulation results at each prediction time point were selected as the lower and upper bounds of the 95% prediction interval, respectively. The cumulative wear prediction results for the C6 cutting tool and their corresponding 95% prediction intervals are shown in Figure 11. This prediction interval reflects how random fluctuations in wear increments propagate during the cumulative prediction process, providing a probabilistic basis for evaluating the uncertainty of the model's prediction results.

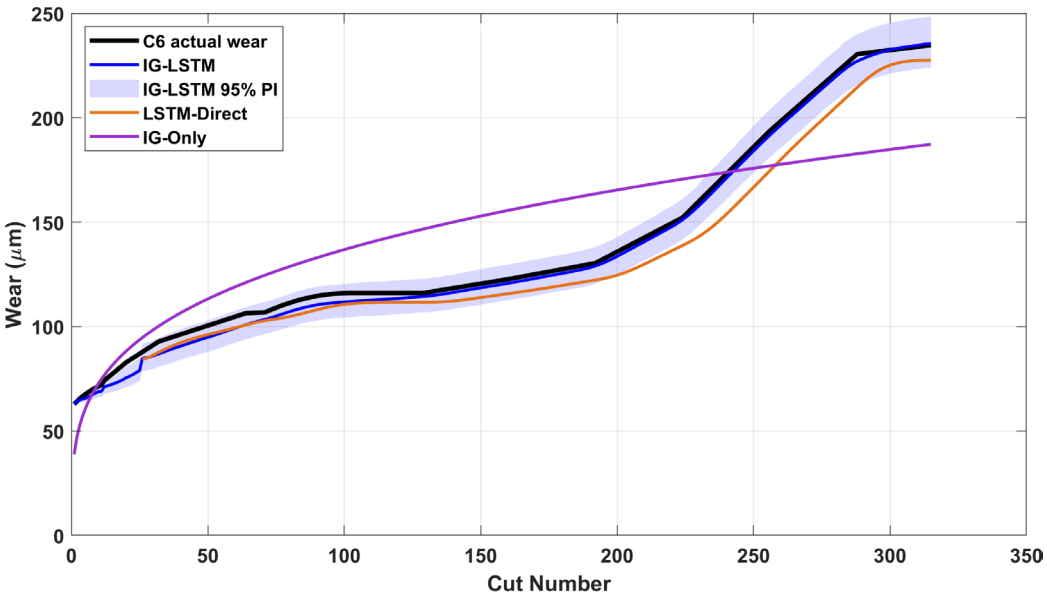


Figure 11. Cumulative wear prediction results for the IG-LSTM model and the 95% prediction interval.

Table 5 summarizes the cumulative wear prediction performance of the three methods across the two scenarios. Analysis of the data in the table shows that, regardless of whether in a single-tool or cross-tool scenario, the IG-LSTM outperforms the two comparison methods in terms of all three metrics: RMSE, MAE, and R^2. Although the LSTM method generally aligns with the trend of the actual values, the lack of physical constraints causes errors to accumulate continuously during long-term recursive predictions, resulting in a significantly higher RMSE. The performance decline of the traditional IG method is most pronounced; its fixed power-law cumulative degradation function, $\Lambda(t) = \alpha t^{\beta}$, cannot adapt to the multi-stage nonlinear degradation trend of the C6 tool. IG-LSTM combines the advantages of both approaches: on the one hand, it uses LSTM to adaptively learn the form of the cumulative degradation function from the data, overcoming the limitations of a fixed power-law; on the other hand, by converting the trend into wear predictions via the mean function of the IG process, it ensures the physical plausibility of the predicted trajectories. Figures 12 and 13 present bar charts comparing the RMSE, MAE, and R^2 values of the three methods across the two scenarios. A closer look at the bar charts allows for a more intuitive conclusion: IG-LSTM achieves significantly lower error metrics than both LSTM and traditional IG in both scenarios, with R^2 consistently remaining above 0.99. This indicates that the proposed method demonstrates excellent predictive performance and good generalization ability in both single-tool and cross-tool scenarios.

Table 5. Comparison of Wear Prediction Performance Among Different Methods.

Scene	Methods	RMSE	MAE	R^2
Single-tool	IG-LSTM	1.42	1.24	0.9962
	LSTM	8.11	6.75	0.8762
	IG	5.73	5.04	0.9381
Across Tools (Test Set)	IG-LSTM	3.26	2.83	0.9951
	LSTM	10.62	9.31	0.9430
	IG	27.94	25.25	0.6049
Across Tools (Training Set)	IG-LSTM	15.56	15.06	0.96
	LSTM	9.96	8.26	0.98
	IG	18.74	13.84	0.69

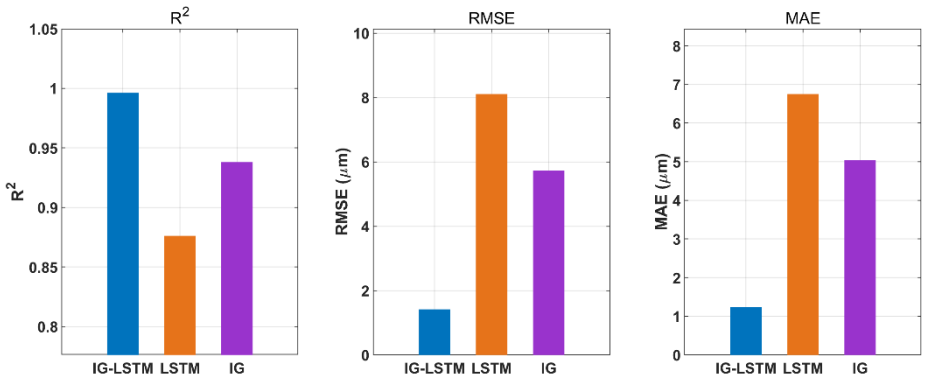


Figure 12. Comparison of the predictive performance of IG-LSTM, LSTM, and the IG method in a single-tool scenario.

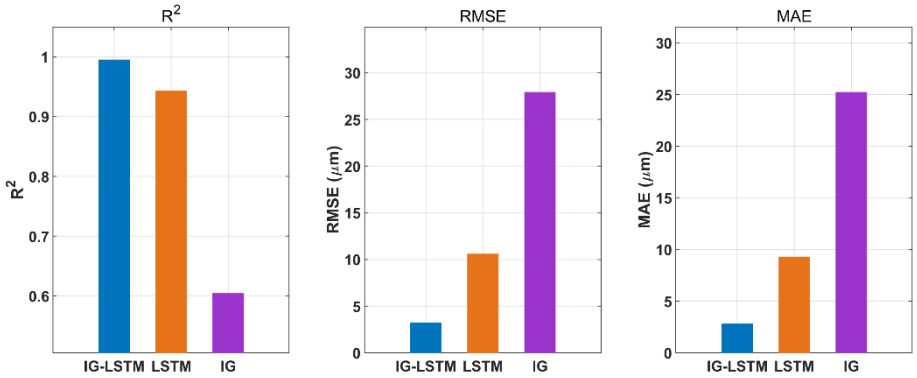


Figure 13. Comparison of the predictive performance of IG-LSTM, LSTM, and the IG method across different tooling scenarios.

4.6. Analysis of Forecasting Uncertainty and Evaluation of Model Stability

To further analyze the role of the IG random process in the proposed model, this paper compares the deterministic wear prediction results—based on LSTM-based trend forecasting—with those of the IG-LSTM model. The deterministic prediction directly calculates the expected wear using the product of the predicted trend and the drift parameter, i.e., $\hat{Y}(t) = \mu\Lambda(t)$. By comparing the two methods, this paper analyzes the role of introducing the IG random process in describing the probability distribution of wear increments and the associated stochastic uncertainty.

As shown in Figure 14, deterministic prediction calculates the expected wear solely based on the predicted degradation trend and drift parameters, whereas IG-LSTM further accounts for the randomness of wear increments and reflects the uncertainty of the wear process through a prediction interval. A comparison of the two methods demonstrates that the IG stochastic process provides a probabilistic description of the traditional deterministic trend prediction, enabling the model not only to predict wear trends but also to characterize the random fluctuations in the wear process.

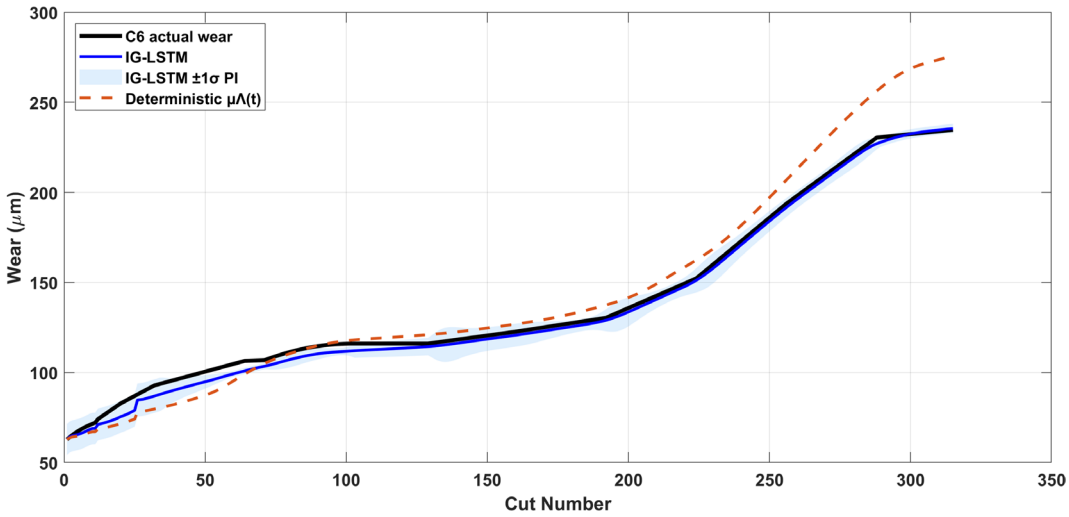


Figure 14. Comparison of Deterministic Predictions with IG-LSTM Tool Wear Prediction Results.

To further evaluate the reliability of the IG-LSTM model's uncertainty estimates, this paper conducts a quantitative analysis of the constructed 95% prediction intervals, introducing two evaluation metrics: the Prediction Interval Coverage Probability (PICP) and the Mean Prediction Interval Width (MPIW). Specifically, PICP measures the proportion of actual wear observations that fall within the prediction interval, while MPIW measures the average width of the prediction interval. Ideally, PICP should be close to the nominal 95% confidence level, while MPIW should be as small as possible while still ensuring reasonable coverage.

As shown in Figure 15, the 95% prediction interval constructed by the IG-LSTM model covers the majority of the observed tool wear values. Among the 315 test samples, 305 fell within the prediction interval, corresponding to a PICP of 96.8% and an average prediction interval width (MPIW) of 17.72 μm. Furthermore, coverage statistics show that 0% of the actual wear values were below the lower bound of the prediction interval, while 3.2% were above the upper bound. This indicates that the observations not covered by the prediction interval are all located beyond the upper bound, suggesting that the coverage bias of the prediction interval primarily manifests as an underestimation of some larger wear values.

Overall, the proposed IG-LSTM model achieved an actual coverage rate close to the nominal 95% confidence level, indicating that its prediction intervals have good empirical coverage capability. At the same time, the average prediction interval width of 17.72 μm further reflects the model's quantification of the uncertainty in tool wear predictions.

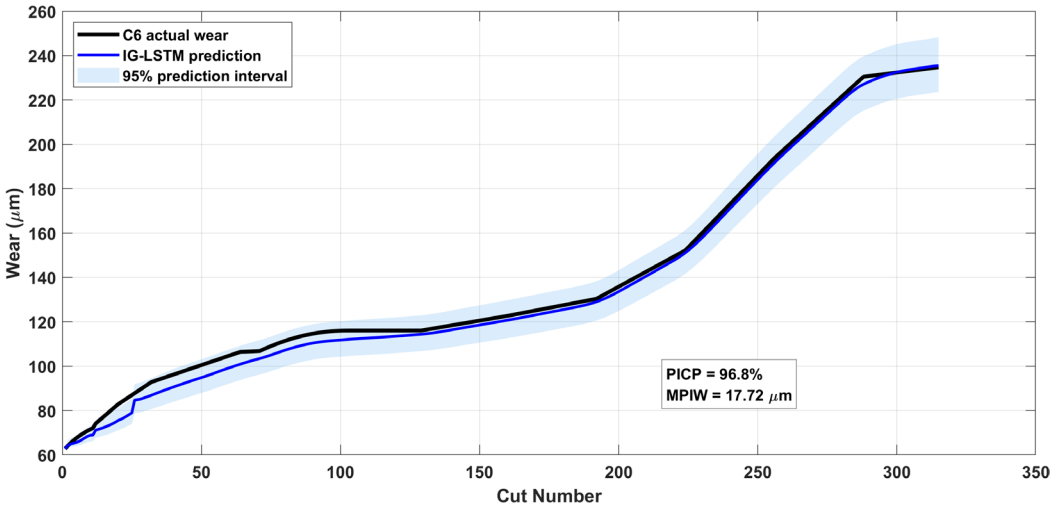


Figure 15. Evaluation of the 95% prediction interval coverage for the IG-LSTM model.

4.7. Analysis of Model Stability Under Different Random Initializations

To evaluate the impact of random initialization on the predictive performance of the IG-LSTM model, five different random seeds (10, 20, 30, 40, and 50) were selected for independent training, while keeping the dataset split, network architecture, and hyperparameters constant. For each training run, the LSTM network was reinitialized, and the resulting degradation trend predictions were fed into the inverse Gaussian process model to obtain the final tool wear predictions. RMSE, MAE, and R^2 were used to evaluate the predictive performance of each experiment; the results are shown in Table 6.

Table 6. Prediction Performance of the IG-LSTM Model Under Different Random Initializations.

Random Seed	RMSE	MAE	R^2
10	2.265	1.664	0.9976
20	3.455	3.044	0.9945
30	7.151	6.719	0.9766
40	1.974	1.663	0.9982
50	2.965	2.491	0.9960

To further analyze, in a visual manner, the impact of different random initializations on the model's prediction results, Figure 15 shows the tool wear prediction curves of the IG-LSTM model under five different random seeds. By comparing the prediction results under different random initializations with the actual wear data, we can observe how the model's prediction curves change and how well they align with the actual wear trends.

As shown in Table 6 and Figure 16, the average R^2 across the five experiments reached 0.9926, with all values exceeding 0.97, indicating that the model exhibits good fitting capability under different random initializations. Judging from the prediction curves, while there are some differences in the prediction results under different random seeds, they generally reflect the trend of tool wear. Specifically, when the random seed was 30, the RMSE and MAE reached 7.151 μm and 6.719 μm, respectively—larger errors compared to the other four experiments—yet the R^2 still reached 0.9766. This suggests that random initialization has a certain impact on the model's predictive accuracy; however, the model generally maintains good predictive performance under different initialization conditions, demonstrating a degree of stability.

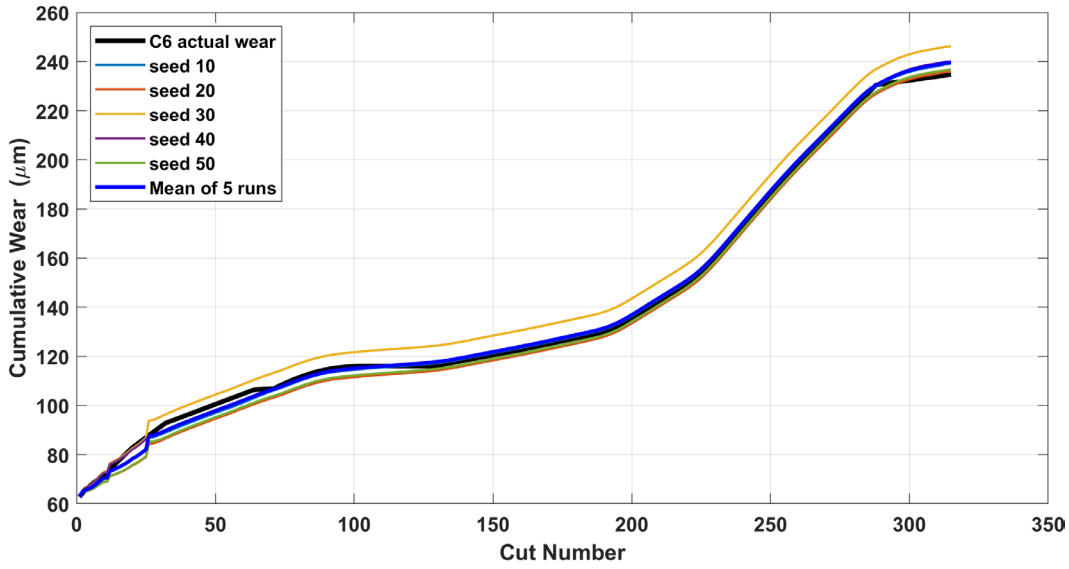


Figure 16. Comparison of prediction results from the IG-LSTM model under different random initializations.

4.8. A Comparative Analysis of Strategies for Characterizing the Wear Characteristics of Different Cutting Tools

To verify the validity of the maximum fusion strategy selected in this paper, this section uses the target test tool C6 as the subject and compares five methods for characterizing tool wear: maximum fusion (Max fusion), mean fusion (Mean fusion), and the use of wear data from the first, second, and third cutting edges alone (denoted as Flute 1, Flute 2, and Flute 3, respectively).

As can be seen from the wear evolution curve in Figure 17, due to uneven cutting forces and local defects, the wear on each cutting edge exhibits a distinct asynchronous pattern. Flute 1 shows the least wear and the slowest rate of wear progression; Flutes 2 and 3 exhibit a sharp acceleration in wear during the later stages. If the "Mean fusion" method is used, the overall wear is severely underestimated due to the "dilution" effect caused by Flute 1's lighter wear, making it difficult to reflect the tool's true degradation. On the other hand, relying solely on data from a single cutting edge (such as Flute 1) is highly one-sided and makes it very easy to overlook severe damage to other cutting edges, leading to a risk of underreporting.

Using the five sequences mentioned above as input features, we performed cross-tool predictions using the IG-LSTM model described earlier; the results are shown in Figure 18. To provide a more intuitive comparison of the various metrics, they are summarized in Figure 19.

4.9. Sensitivity Analysis of Different Data Preprocessing Strategies

To analyze the impact of different data preprocessing operations on the estimation of IG process parameters and wear prediction results, this paper designed four preprocessing schemes. Scheme A represents the complete preprocessing method; Scheme B omits the monotonicization of the wear time series; Scheme C omits the monotonicization of the degraded trend time series; and Scheme D omits the removal of invalid increments. While maintaining consistent model structures and other experimental settings, the paper compares the IG parameter estimation results and predictive performance under the different schemes.

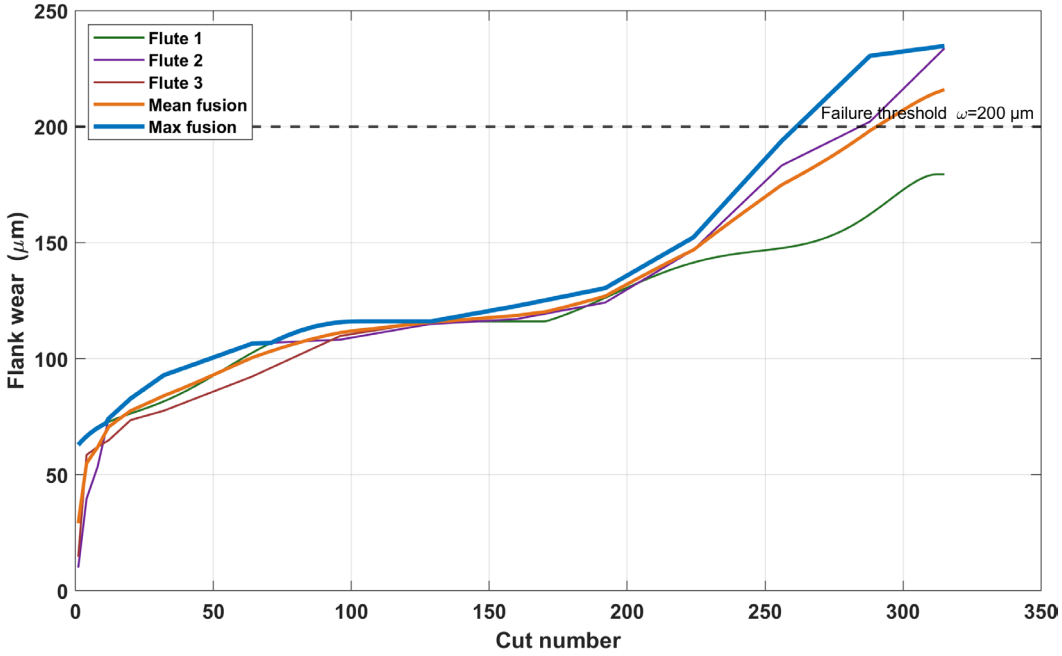


Figure 17. Comparison of Wear Curves for the C6 Tool Under Different Characterization Schemes.

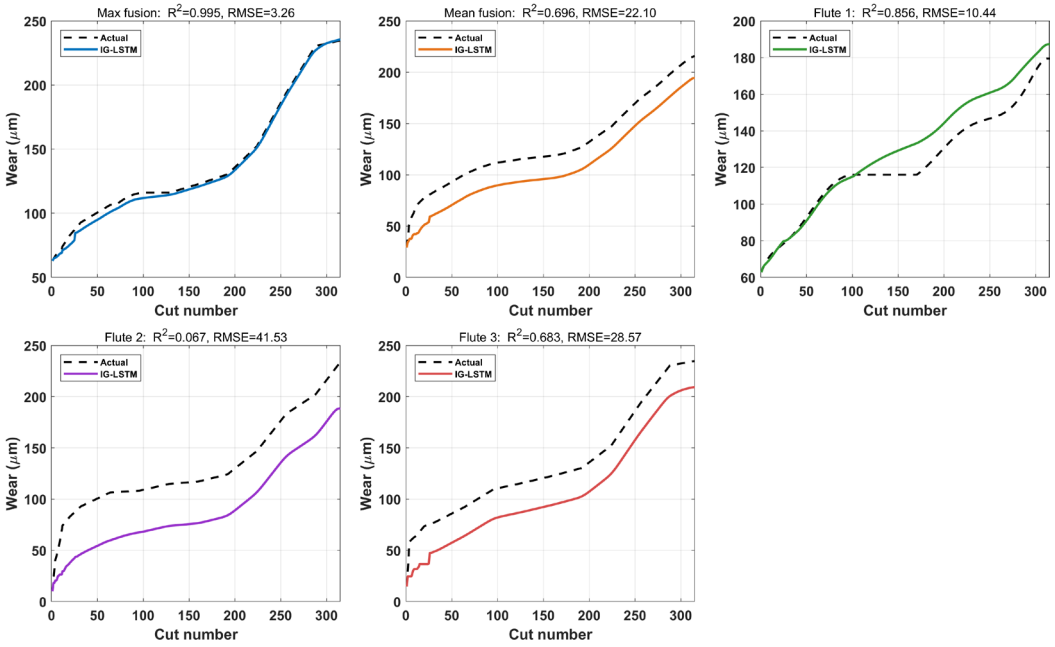


Figure 18. Comparison of IG-LSTM prediction results under different feature representation schemes.

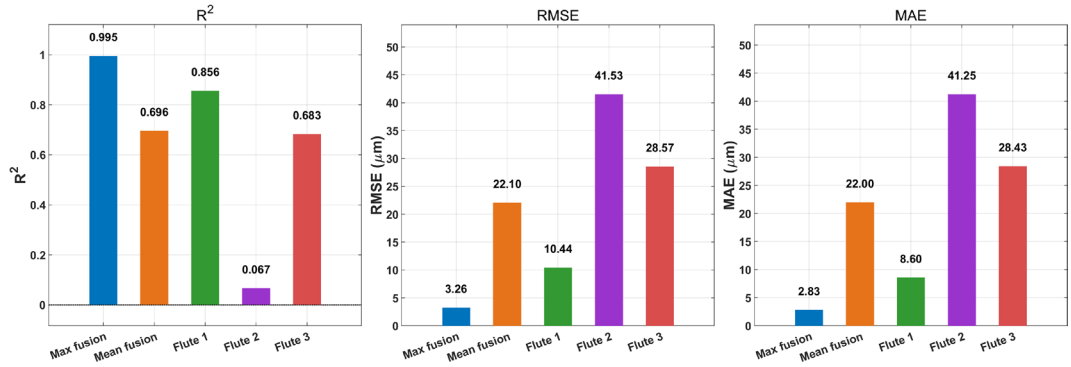


Figure 19. Comparison of Predictive Performance Metrics for Different Methods of Characterizing Tool Wear Patterns.

Figure 20 shows the estimation results for the IG process parameters under different preprocessing schemes. The parameters for Scheme A and Scheme B are identical; this is because the sequence obtained after fusing the maximum wear values of the three cutting edges already satisfies the monotonically non-decreasing property, so the cummax operation did not alter the original data. After removing the trend-monotonicity processing, μ and λ changed by −0.42% and +1.91%, respectively, indicating that this operation has a minor impact on parameter estimation. In contrast, after removing the invalid increment filtering, μ decreased by 4.62%, and the variance factor $\frac{\mu_3}{\lambda}$ decreased by 12.79%, suggesting that invalid increment filtering has a more significant impact on IG parameter estimation and the characterization of stochastic fluctuations.

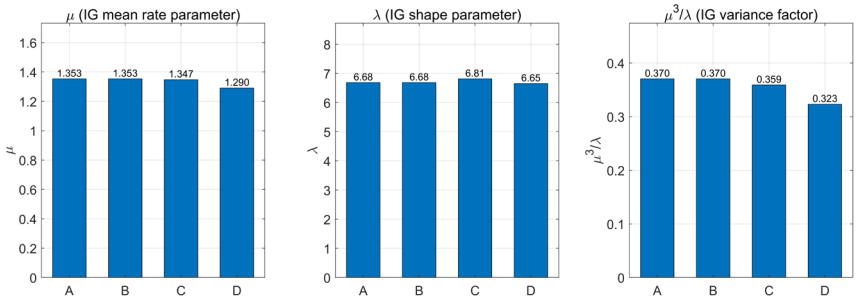


Figure 20. Effect of Different Data Preprocessing Strategies on the Estimation Results of IG Process Parameters. (A: complete preprocessing; B: without wear-series monotonicization; C: without degradation-trend monotonicization; D: without invalid-increment filtering).

To further analyze the impact of different preprocessing schemes on the model's predictive performance, a uniform baseline drift parameter of $\mu = 1.35$ was used for comparison to eliminate the influence of differences in the drift parameter on point prediction results. As shown in Figure 21, the prediction curves under the four schemes largely overlap, indicating that, under the same drift parameter conditions, different preprocessing strategies have a minimal impact on the model's point prediction results.

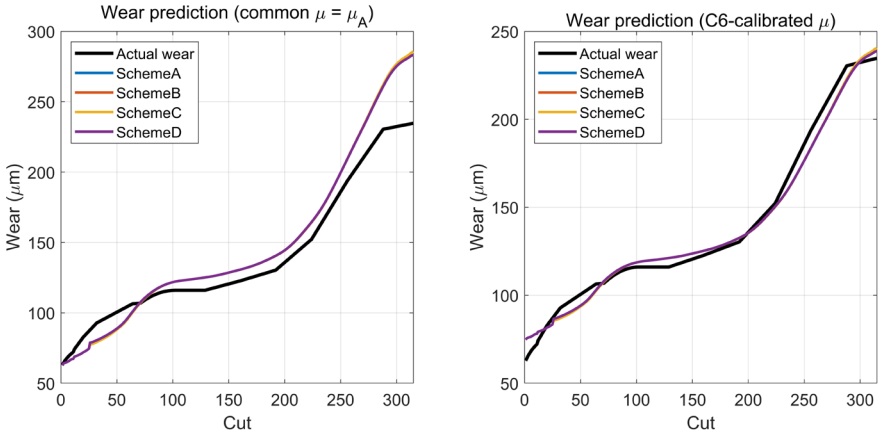


Figure 21. Comparison of Tool Wear Prediction Results Under Different Data Preprocessing Strategies.

In summary, the impact of different preprocessing schemes on the model's final point predictions is minimal, indicating that the proposed IG-LSTM model exhibits good preprocessing robustness. At the same time, the aforementioned preprocessing operations are primarily used to meet the modeling requirements of the inverse Gaussian process for positive increments. By monotonizing the wear sequence and the degradation trend sequence, we ensure that the increments are non-negative. Combined with minor positive adjustments and the removal of invalid increments, this ensures that both the wear increments and trend increments used for IG parameter estimation are strictly greater than zero, thereby guaranteeing the validity of the parameter estimates.

4.10. Ablation Studies on Differential Processing and Scale Calibration

To analyze the impact of differential processing and scale calibration on the predictive performance of the IG-LSTM, two sets of ablation experiments were designed while maintaining consistent data splitting and prediction settings, and these were compared with the full model. The first set retained VMD trend extraction and differential processing but removed scale calibration; the second set retained only VMD trend extraction

while removing both differential processing and scale calibration. The experimental results are shown in Figures 22 and 23. The full IG-LSTM model achieved R^2, RMSE, and MAE values of 0.995, 3.26 μm, and 2.83 μm, respectively, demonstrating the best prediction performance. After removing scale calibration, R^2 dropped to 0.8662, while RMSE and MAE increased to 17.09 μm and 12.67 μm, respectively, indicating that scale calibration plays a significant role in improving prediction accuracy. Upon further removal of the differencing process—leaving only the VMD—the model's predictive performance declined significantly, with R^2 dropping to −0.1427 and RMSE and MAE reaching 49.93 μm and 47.72 μm, respectively. These results indicate that the differencing process helps improve the model's predictive performance, while scale calibration further enhances predictive accuracy on this basis; both factors jointly contribute to the predictive performance of the IG-LSTM framework.

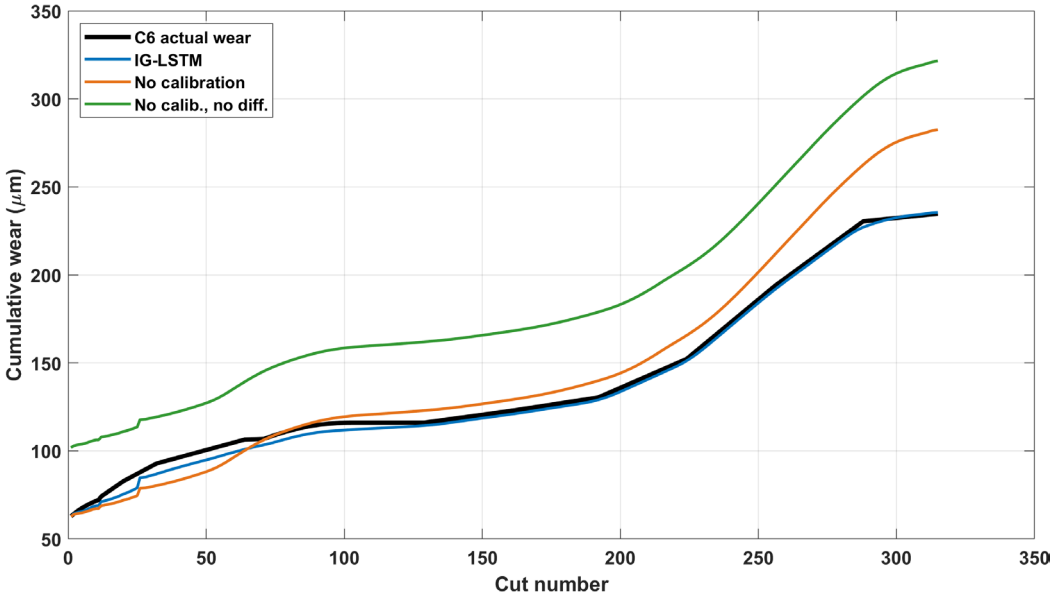


Figure 22. Predicted cumulative wear of the C6 cutting tool under different ablation configurations.

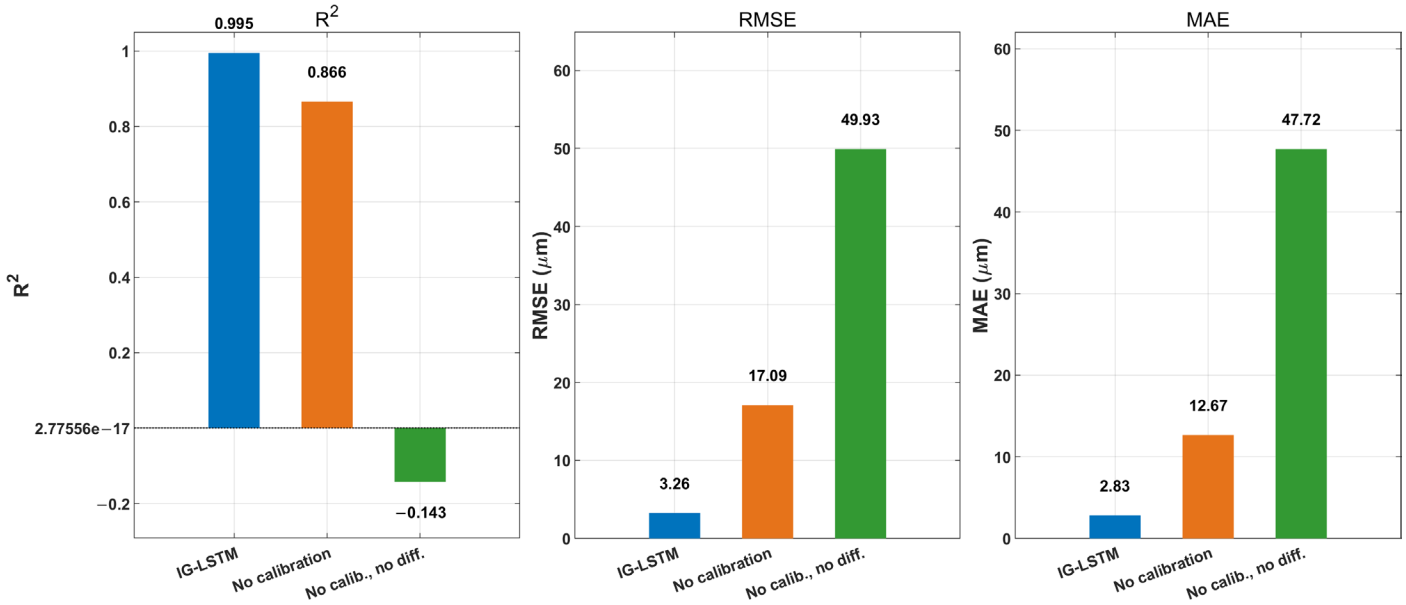


Figure 23. Comparison of Performance Evaluation Metrics for C6 Tool Wear Prediction Under Different Ablation Configurations.

4.11. Comparative Experiments of Different Types of Predictive Models

To further validate the generalization capability of the proposed method relative to other mainstream data-driven architectures, this section presents comparative experiments involving three purely data-driven methods—RNN, CNN, and CNN-LSTM—in a cross-tool scenario. Figure 24 shows the prediction curves of each method over the entire life cycle of the C6 tool, while Figure 25 presents the corresponding bar charts of performance metrics. In contrast, the prediction curve of IG-LSTM closely matches the actual wear curve throughout all stages, maintaining particularly strong predictive capability during the accelerated wear phase. Quantitative results show that IG-LSTM achieves an R^2 of 0.995, an RMSE of 3.06 μm, and an MAE of 2.70, outperforming RNN (R^2 = 0.922, RMSE = 9.41 μm, MAE = 6.34), CNN-LSTM (R^2 = 0.985, RMSE = 5.35 μm, MAE = 4.09), and CNN (R^2 = 0.9805, RMSE = 6.21 μm, MAE = 4.59). These results indicate that by modeling wear increments through the IG process, IG-LSTM effectively suppresses error drift in long-term recursive predictions—a common issue with purely data-driven models—and demonstrates stronger generalization capabilities across different tool scenarios.

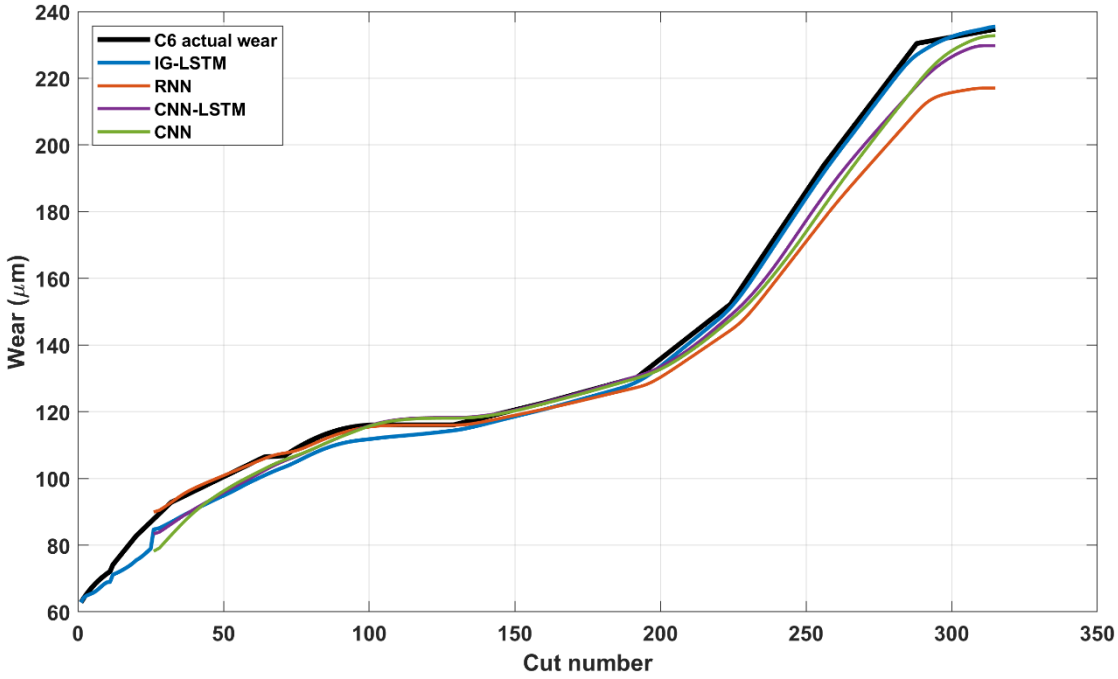


Figure 24. Comparison of wear prediction curves between IG-LSTM and purely data-driven methods in a cross-tool scenario.

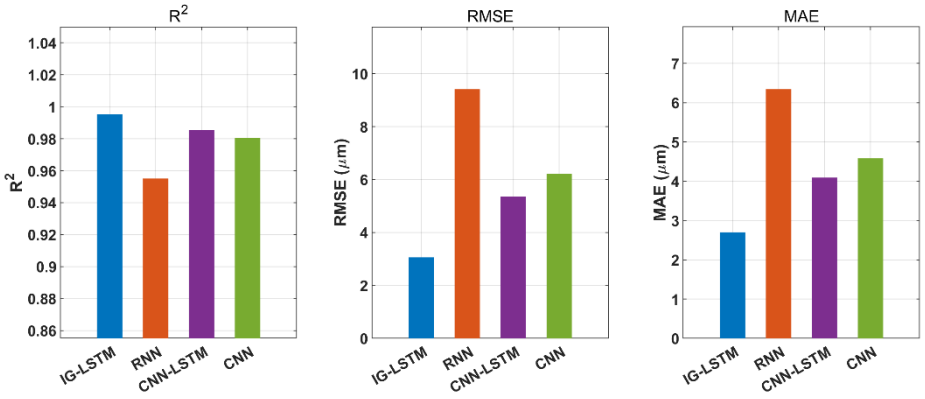


Figure 25. Comparison of the Predictive Performance of Different Methods Across Tool Scenarios.

To further evaluate the predictive performance of the proposed IG-LSTM model, comparisons were conducted with four conventional regression models, including linear regression (LR), polynomial regression (PR), support vector regression (SVR), and random forest regression (RFR). All models were evaluated using the same C6 dataset and prediction settings.

As shown in Figure 26, the proposed IG-LSTM model closely follows the actual cumulative tool wear throughout the prediction process, particularly during the later stage when the wear rate increases rapidly. In contrast, the conventional regression models exhibit larger deviations from the actual measurements and fail to adequately capture the accelerated wear growth in the later stage. Although these regression models can describe part of the overall degradation trend, their predictions do not fully reflect the nonlinear and dynamically changing characteristics of tool wear. The results demonstrate that the proposed IG-LSTM model provides better agreement with the actual wear measurements and is more effective in capturing the nonlinear degradation trend.

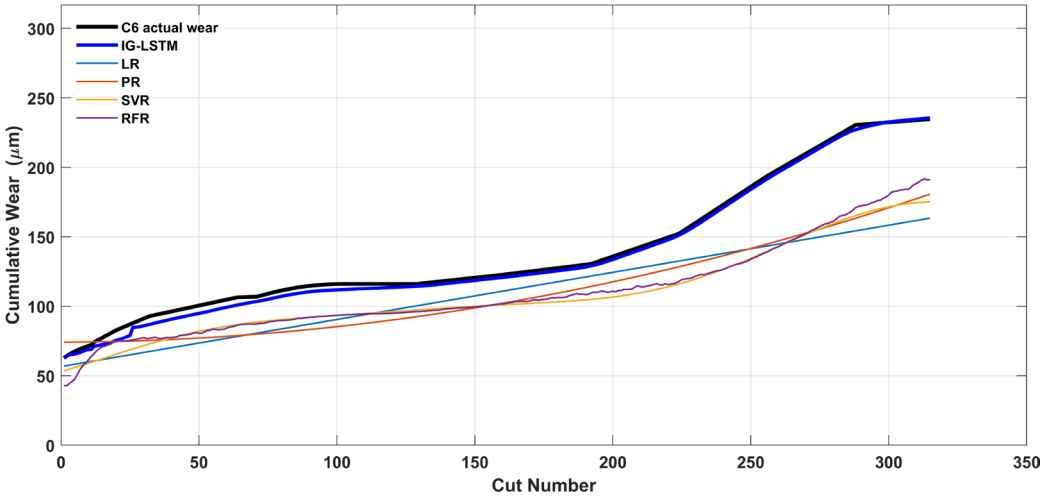


Figure 26. Comparison of the cumulative wear prediction results for the C6 cutting tool using IG-LSTM and traditional regression models.

5. Conclusions

This paper addresses the issue that traditional parametric inverse Gaussian process cumulative degradation functions have a fixed form and struggle to adapt to multi-stage nonlinear degradation under complex cutting conditions. It proposes a tool wear prediction model (IG-LSTM) that integrates LSTM with the inverse Gaussian process. This method replaces the strong assumption of a fixed power-law form with the cumulative degradation function $\Lambda(t)$ of the IG process, which is learned adaptively using an LSTM. Based on this, a probabilistic model of wear increment is constructed using the $\Lambda(t)$ predicted by the LSTM as the cumulative degradation function of the IG process, and the model parameters are estimated using maximum likelihood estimation, thereby forming a tool wear prediction framework that integrates data-driven time-series modeling with stochastic degradation processes.

Based on the PHM2010 dataset, this paper conducted wear prediction experiments for both single-tool and multi-tool scenarios. In the single-tool scenario, the IG-LSTM achieved RMSE, MAE, and R^2 values of 1.42, 1.24, and 0.9962, respectively, outperforming both the LSTM (8.11, 6.75, and 0.8762) and the IG (5.73, 5.04, and 0.9381). In the cross-tool scenario, the RMSE, MAE, and R^2 of IG-LSTM were 3.26, 2.83, and 0.9951, respectively; in contrast, the three metrics for LSTM were 10.62, 9.31, and 0.9430, while those for IG were 27.94, 25.25, and 0.6049. The above results indicate that, in both experimental

scenarios, IG-LSTM achieved superior prediction performance compared to the two baseline methods.

Although the IG-LSTM achieved good prediction results, this study still has certain limitations. The experiments were conducted solely on the PHM2010 dataset under specific experimental conditions; the model's generalization capability for other cutting tools and different operating environments still requires further validation. In the future, we will explore methods such as domain adaptation and meta-learning to improve cross-tool generalization, and consider integrating multi-source sensor signals to enhance the model's robustness under various operating conditions. The IG-LSTM framework has the potential to support online tool wear monitoring by continuously processing cutting force, vibration, and acoustic emission signals to estimate tool wear and predict its evolution. Comparing predicted wear with a predefined threshold could enable early warnings and inform tool inspection or replacement. However, real-time deployment and closed-loop control remain unvalidated. Future work will integrate the framework with online data acquisition systems and evaluate its robustness under varying machining conditions and sensor noise.

Author Contributions: Conceptualization, B.S. and J.Z.; methodology, B.S.; software, X.W.; validation, B.S., J.Z. and L.Z.; formal analysis, X.W.; investigation, L.W.; resources, B.S.; data curation, Y.W.; writing—original draft preparation, C.L.; writing—review and editing, J.Z.; supervision, B.S.; project administration, B.S.; funding acquisition, B.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Changchun University (Grant No. ZKP202201) through the project entitled "Research on Predictive Maintenance of CNC Machine Tools Based on Digital Twins.".

Data Availability Statement: The data used in this study are publicly available from the PHM2010 tool wear dataset. No new data were collected for this study.

Acknowledgments: During the drafting of this manuscript, the author refined the initial version of the introduction using the DeepSeek-V3 model.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

VMD	Variational Mode Decomposition
LSTM	Long Short-Term Memory
IG	Inverse Gaussian
EMD	Empirical Mode Decomposition
IMFs	limited intrinsic mode functions
RMSE	Root mean square error
MAE	Mean absolute error
R^2	Coefficient of determination
PHM2010	PHM2010 tool wear dataset
IG-LSTM	Inverse Gaussian process combined with a long short-term memory network

References

1. Wang, J.; Wang, P.; Gao, R.X. Enhanced particle filter for tool wear prediction. *J. Manuf. Syst.* **2015**, *36*, 35–45.
2. Li, C.; Sun, X.; Hou, X.; Zhao, X.; Wu, S. A Digital Twin-Driven Method for Online Monitoring of CNC Milling Tool Wear. *China Mech. Eng.* **2022**, *33*, 78–87.

3. Shahabi, H.H.; Ratnam, M.M. In-cycle monitoring of tool nose wear and surface roughness of turned parts using machine vision. *Int. J. Adv. Manuf. Technol.* **2009**, *40*, 1148–1157.
4. Li, X.; Liu, X.; Yue, C.; Wang, L.; Liang, S.Y. Data-model linkage prediction of tool remaining useful life based on deep feature fusion and Wiener process. *J. Manuf. Syst.* **2024**, *73*, 19–38.
5. Li, B. A review of tool wear estimation using theoretical analysis and numerical simulation technologies. *Int. J. Refract. Met. Hard Mater.* **2012**, *35*, 143–151. https://doi.org/10.1016/j.ijrmhm.2012.05.006.
6. Yen, Y.; Söhner, J.; Lilly, B.; Altan, T. Estimation of tool wear in orthogonal cutting using the finite element analysis. *J. Mater. Process. Technol.* **2004**, *146*, 82–91. https://doi.org/10.1016/S0924-0136(03)00847-1.
7. Liu, C.; Zhu, L. A two-stage approach for predicting the remaining useful life of tools using bidirectional long short-term memory. *Measurement* **2020**, *164*, 108029.
8. Ren, L.; Dong, J.; Wang, X.; Meng, Z.; Zhao, L.; Deen, M.J. A data-driven auto-CNN-LSTM prediction model for lithium-ion battery remaining useful life. *IEEE Trans. Ind. Inform.* **2020**, *17*, 3478–3487.
9. An, H.; Wang, G.; Wang, Z.; Ma, K.; Zhong, C. A Method for Tool Condition Monitoring and Remaining Life Prediction Based on Deep Learning Theory. *J. Electron. Meas. Instrum.* **2019**, *33*, 64–70. https://doi.org/10.13382/j.jemi.B1902208.
10. Cheng, J.; Sai, X.; Zhang, C.; Luo, M. Tool Wear Prediction via Multi-Sensor Information Fusion Using a Stacked LSTM Based on Attention Mechanisms. *Ind. Eng.* **2024**, *27*, 64–77+86.
11. Li, R.; Ye, X.; Yang, F.; Du, K.-L. ConvLSTM-Att: An attention-based composite deep neural network for tool wear prediction. *Machines* **2023**, *11*, 297.
12. Xu, X.; Wang, J.; Zhong, B.; Ming, W.; Chen, M. Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism. *Measurement* **2021**, *177*, 109254.
13. Ma, J.; Luo, D.; Liao, X.; Zhang, Z.; Huang, Y.; Lu, J. Tool wear mechanism and prediction in milling TC18 titanium alloy using deep learning. *Measurement* **2021**, *173*, 108554.
14. Zhang, P.; Gao, D.; Hong, D.; Lu, Y.; Wang, Z.; Liao, Z. Intelligent tool wear monitoring based on multi-channel hybrid information and deep transfer learning. *J. Manuf. Syst.* **2023**, *69*, 31–47.
15. Zhang, Z.; Si, X.; Hu, C.; Lei, Y. Degradation data analysis and remaining useful life estimation: A review on Wiener-process-based methods. *Eur. J. Oper. Res.* **2018**, *271*, 775–796.
16. Whitmore, G.A. Estimating degradation by a Wiener diffusion process subject to measurement error. *Lifetime Data Anal.* **1995**, *1*, 307–319.
17. Kang, M.; Lee, S.; Kim, J.H.; Yoo, C.S.; Jang, J.; Negatu, B.M.; Park, S. Remaining useful life estimation using accelerated degradation test, a gamma process, and the arrhenius model for nuclear power plants. *J. Mech. Sci. Technol.* **2022**, *36*, 4905–4912.
18. Lin, F.; Zhang, H.; Zhou, Y.; Zhang, Z.; Zhang, L. Tool Reliability Modeling Based on Gamma Process in Multiple Working Conditions. In Proceedings of the IEEE 2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing), Nanjing, China, 15–17 October 2021; Volume 2021, pp. 1–6.
19. Sun, B.; Li, Y.; Wang, Z.; Ren, Y.; Feng, Q.; Yang, D. An improved inverse Gaussian process with random effects and measurement errors for RUL prediction of hydraulic piston pump. *Measurement* **2021**, *173*, 108604.
20. Kong, D.; Chen, Y.; Li, N. Gaussian process regression for tool wear prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 556–574.
21. Wang, Y.; Chang, M.; Huang, X.; Li, Y.; Tang, J. Cutting tool wear prediction based on the multi-stage Wiener process. *Int. J. Adv. Manuf. Technol.* **2023**, *129*, 5319–5333.
22. Li, C.; Zhang, Y. Time-Variant Reliability Assessment and Its Sensitivity Analysis of Cutting Tool under Invariant Machining Condition Based on Gamma Process. *Math. Probl. Eng.* **2012**, *2012*, 676923.
23. Zhu, K.; Huang, C.; Li, S.; Lin, X. Physics-informed Gaussian process for tool wear prediction. *ISA Trans.* **2023**, *143*, 548–556.
24. Qiang, B.; Shi, K.; Liu, N.; Ren, J.; Shi, Y. Integrating physics-informed recurrent Gaussian process regression into instance transfer for predicting tool wear in milling process. *J. Manuf. Syst.* **2023**, *68*, 42–55.
25. Zhang, N. A Study on Data-Driven Methods for Monitoring Tool Wear and Predicting Remaining Service Life. Ph.D. Thesis, Yanshan University, Qinhuangdao, China, 2024. https://doi.org/10.27440/d.cnki.gysdu.2024.000036.
26. Sak, H.; Senior, A.; Beaufays, F. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv* **2014**, arXiv:1402.1128.
27. Chen, H. A multiple model prediction algorithm for CNC machine wear PHM. *Int. J. Progn. Health Manag.* **2011**, *2*, 129.
28. Sun, Y.; Song, H.; Gao, H.; Li, J.; Yin, S. Interpretable tool wear monitoring: Architecture with large-scale CNN and adaptive EMD. *J. Manuf. Syst.* **2025**, *78*, 294–307.

29. Yuan, J.; Liu, L.; Yang, Z.; Zhang, Y. Tool wear condition monitoring by combining variational mode decomposition and ensemble learning. *Sensors* **2020**, *20*, 6113.
30. Lu, J.; Jiang, W.; Xu, Y.; Chen, Z.; Ni, K. Measurement of aero-engine feature-hierarchy fusion degradation trend based on parameter-adaptive VMD method and improved transformer model. *Meas. Sci. Technol.* **2024**, *35*, 075005.
31. Hao, S.; Yang, J.; Berenguer, C. Degradation analysis based on an extended inverse Gaussian process model with skew-normal random effects and measurement errors. *Reliab. Eng. Syst. Saf.* **2019**, *189*, 261–270.
32. Huang, S.; Li, X.; Gan, O.P. Tool wear estimation using support vector machines in ball-nose end milling. In Proceedings of the Annual Conference of the PHM Society, Portland, OR, USA, 10–14 October 2010; Volume 2.
33. Mao, B.; Bu, Z.; Xu, B.; Gong, H.; Li, Y.; Wang, H.; Kang, J.; Jin, S.; Zhao, C. Denoising method based on VMD-PCC in φ-OTDR system. *Opt. Fiber Technol.* **2022**, *74*, 103081.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

https://doi.org/10.3390/s26227158