

Article

A Dynamic-Response-Enhanced Active Current Prediction Method for Synchronous Reluctance Motors Under Multi-Operating-Condition Switching

Fang Zhang ^{1,2} , Bo Zhao ^{3,4,*}, Longhao Li ^{3,4} and Dianlin Shen ^{3,4}

¹ Guangdong Provincial Key Laboratory of High-Speed and Energy Conservation Motor System, Zhuhai 519060, China; zhangfang1900@163.com

² Gree Electric Appliances, Inc. of Zhuhai, Zhuhai 519060, China

³ Key Discipline and Technology Research Center for Special Electric Motor Drives and Multi-Source Power Supply, Shandong University of Technology, Zibo 255000, China; lilonghao@sdut.edu.cn (L.L.)

⁴ School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China

* Correspondence: zhaobo@sdut.edu.cn

Abstract

Synchronous reluctance motors in quadruped robot joint drives are prone to active-current peaks, abrupt rate variations, and switching-neighborhood error concentration under foot-ground impacts and obstacle-induced load steps, leading to prediction lag and peak underestimation. To address these issues, this paper proposes a dynamic-response-enhanced multi-condition active-current prediction method based on TPE-VMD-BiLSTM-TRC. First, the original sequence is segmented according to operating-condition boundaries, and prediction samples are constructed within each segment to reduce cross-condition information leakage and distribution inconsistency. Second, variational mode decomposition is performed on each segmented sequence to separate multi-scale fluctuation components, and BiLSTM is used for mode-wise one-step prediction. Third, TPE jointly optimizes the VMD decomposition parameters and BiLSTM hyperparameters to improve parameter matching under different operating conditions. Furthermore, a switching-aware composite loss and a switch-gated residual correction branch are introduced to enhance dynamic tracking and compensate for structural bias in switching neighborhoods. Experiments on variable-frequency drive data show that, compared with TPE-VMD-BiLSTM, the proposed method reduces the overall RMSE, switching-neighborhood RMSE, and MAPE by approximately 29.8%, 38.8%, and 7.4%, respectively. On the additional held-out operating-condition segment, the overall RMSE is reduced by 28.8%, indicating stable prediction performance across different segments.

Keywords: synchronous reluctance motor; active current prediction; multi-condition switching; variational mode decomposition; BiLSTM; switching-aware residual correction



Academic Editor: Wen-Jer Chang

Received: 15 May 2026

Revised: 18 June 2026

Accepted: 26 June 2026

Published: 29 June 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Research Background

With the increasing application of high-performance electric drive systems in robotic joint actuators, mobile operating equipment, and high-dynamic servo systems, electric drive units are required to maintain stable operation under frequent start-stop processes, rapid load variations, and alternating forward-reverse rotation conditions [1]. Taking quadruped

robot joint actuators as an example, the transition between support and swing phases, foot-ground contact impact, and posture adjustment may induce short-term rapid variations in joint load and current response. Such operating scenarios impose higher requirements on the dynamic response, overcurrent suppression, device loss control, and thermal-load management of drive systems. Therefore, establishing a short-term active current prediction model for complex operating-condition switching processes is of practical significance for improving the state perception and safe operation capability of electric drive systems [2].

Synchronous reluctance motors, with no rotor windings or permanent magnets, have the advantages of a simple structure, high-temperature resistance, demagnetization immunity, and a wide speed regulation range. Therefore, they show promising application potential in high-dynamic servo drives, robotic joint actuators, and electric drive systems under complex operating conditions [3]. In these application scenarios, the drive system is required not only to rapidly track torque or speed commands, but also to suppress current peaks, control device losses, and satisfy temperature-rise constraints. As a result, the short-term variation of active current is of great importance for power-state evaluation, overcurrent warning, and operational safety assessment.

During multi-operating-condition processes such as frequency steps, load transients, and alternating forward-reverse rotation, the active current sequence usually exhibits short-term peaks, abrupt increases in local slope, and inter-condition distribution shifts [4]. These characteristics make the prediction task no longer a conventional fitting problem for stationary sequences. If the model is mainly optimized based on the overall average error, the prediction results are prone to peak underestimation and turning-point lag in switching neighborhoods. If the sample construction process ignores operating-condition boundaries, the distribution discrepancies among different condition segments may further weaken the credibility of model evaluation. Therefore, the key to active current prediction for synchronous reluctance motors lies not only in reducing the overall error, but also in improving the model's ability to track local dynamic variations in switching neighborhoods [5].

1.2. Related Work

Short-term active current prediction in electric motors is a complex nonlinear time-series modeling problem, as its dynamic evolution is jointly affected by electromagnetic characteristics, control structures, load disturbances, sampling errors, and operating-condition switching [6]. Existing methods can generally be divided into mechanism-driven and data-driven approaches. Mechanism-driven methods usually estimate and predict current based on motor parameter models, flux linkage equations, state observers, or model predictive frameworks. Although these methods have good physical interpretability, prediction accuracy may deteriorate under parameter perturbations, load transients, and frequent multi-condition switching due to model mismatch [7]. In contrast, data-driven methods can learn the nonlinear mapping relationship between input variables and current responses from historical operating data, and show stronger representational capability for nonstationary evolution under complex operating conditions. Therefore, they have gradually become an important research direction for short-term current signal prediction.

In data-driven time-series modeling, LSTM alleviates the gradient vanishing problem of conventional recurrent neural networks through its gating mechanism, and can extract nonlinear temporal dependencies from historical sequences [8]. Teler et al. [9] applied LSTM to stator current prediction in induction motor dynamic systems, verifying the feasibility of recurrent neural networks for one-step motor current prediction. BiLSTM further extracts temporal information within the input window through both forward and backward recurrence. Le et al. [10] combined FFT preprocessing with BiLSTM for time-series signal prediction in motor drive systems, achieving favorable prediction performance.

However, when LSTM or BiLSTM is directly applied to the original active current sequence, trend variations, local oscillations, and high-frequency disturbances remain mixed in the same input space. As a result, the model tends to favor steady-state regions with a larger sample proportion, while its ability to characterize local transients near operating-condition switching positions remains insufficient.

To address the multi-scale component mixing problem in nonstationary sequences, methods combining signal decomposition and deep prediction have attracted increasing attention. VMD can decompose a complex signal into several modal components with relatively concentrated frequency bands [11], and is often combined with network structures such as BiLSTM, GRU, and CNN for complex time-series prediction. Zhang et al. [12] proposed a VMD-SSA-BiLSTM coupled model, in which VMD was used for sequence decomposition and SSA was adopted to optimize BiLSTM parameters. Zhong et al. [13] applied VMD-SE-Attention-BiLSTM to short-term power load forecasting, further demonstrating that the combination of decomposition methods and deep sequence models can improve the prediction accuracy of nonstationary sequences. Existing studies indicate that deep modeling after decomposition can reduce the nonstationarity of the original sequence and improve overall prediction accuracy. Nevertheless, for multi-condition current sequences, modal reconstruction is still usually optimized toward global error reduction, while compensation for local peak attenuation and structural bias in switching neighborhoods remains insufficient.

In terms of model parameter optimization, intelligent optimization algorithms are commonly used to reduce manual parameter-tuning cost. Yuan et al. [14] adopted the sparrow search algorithm to optimize BiLSTM parameters, while Kaur et al. [15] improved a BiLSTM prediction model using a genetic algorithm. Unlike swarm intelligence optimization, TPE is a probabilistic-model-based sequential hyperparameter optimization method. It can adaptively select more promising candidate combinations according to historical evaluation results and has been applied to hyperparameter optimization of deep learning models [16]. For coupled models such as VMD-BiLSTM, the number of modes and penalty factor of VMD affect the frequency-band division and sequence stationarity after decomposition, while the window length, number of hidden units, and learning rate of BiLSTM determine the fitting capability of the network for each modal component. Therefore, jointly optimizing the VMD decomposition parameters and BiLSTM network hyperparameters is more consistent with the practical requirements of multi-scale prediction tasks than optimizing only one part of the model [17].

In summary, existing studies have improved the prediction accuracy of current-related signals from the perspectives of deep time-series modeling, multi-scale signal decomposition, and intelligent parameter optimization. However, several limitations still remain in active current prediction for synchronous reluctance motors under multi-condition operation. On the one hand, most studies still use overall RMSE or MAE as the main evaluation criteria, with insufficient attention paid to peak underestimation and local error concentration in switching neighborhoods [18]. On the other hand, frequency steps, load transients, and alternating forward-reverse rotation can significantly increase the current increments between adjacent sampling points. Relying only on amplitude-error constraints makes it difficult to ensure the dynamic tracking capability of the predicted curve at turning positions. In addition, different operating-condition segments exhibit discrepancies in mean value, fluctuation amplitude, and spectral structure. Without segment-level sample organization and independent validation, the stability of the model under multi-condition distributions cannot be fully demonstrated [19]. Therefore, it is necessary to construct a more targeted prediction framework around inter-segment distribution discrepancies, local variation trends, and error compensation in switching neighborhoods.

1.3. Research Gaps and Proposed Approach

As reviewed above, existing current signal prediction methods have made progress in deep time-series modeling, multi-scale decomposition, and parameter optimization. However, when applied to short-term active current prediction of synchronous reluctance motors under multi-operating-condition scenarios, they still show insufficient task-specific adaptability. Specifically, most existing methods use overall RMSE or MAE as the main optimization objective, making model training prone to being dominated by steady-state samples and insufficiently constrained for local peaks near frequency steps, load transients, and alternating forward-reverse rotation. In addition, after operating-condition switching, the active current increments between adjacent sampling points may vary significantly. Merely constraining the prediction amplitude is therefore insufficient to ensure consistency in the variation trend of the predicted curve at turning positions. Furthermore, multi-condition experimental sequences consist of multiple operating-condition segments, and different segments may exhibit discrepancies in mean value, fluctuation amplitude, and spectral structure. If sample construction, normalization, and data partitioning ignore operating-condition boundaries, cross-segment information interaction may be introduced, thereby reducing the credibility of model evaluation.

To address these issues, this paper develops a TPE-VMD-BiLSTM-TRC multi-condition prediction method. The proposed method first uses the operating-condition segment as the basic organizational unit, with data partitioning and sliding-window sample construction performed according to operating-condition boundaries, ensuring that the input window and prediction target remain within the same segment. Then, VMD is applied within each operating-condition segment to decompose the active current sequence into multi-scale components, and BiLSTM is used to establish a mode-wise one-step prediction and reconstruction model. Meanwhile, TPE is introduced to jointly search the VMD decomposition parameters and BiLSTM network hyperparameters, thereby improving the matching degree between the decomposition scale and temporal modeling capability. This backbone framework is designed to reduce the nonstationary modeling difficulty of the original active current sequence and provide stable baseline predictions for subsequent error correction in switching neighborhoods.

On this basis, a switching-aware composite loss function and a lightweight residual correction branch are further introduced to enhance the dynamic response capability of the model in switching neighborhoods. The composite loss consists of reconstruction error, first-difference error, and switching-neighborhood weighted error, which, respectively, constrain overall amplitude fitting, local variation trend, and key transition samples. The residual correction branch estimates a one-step residual compensation term based on the backbone prediction, the most recent actual current, difference information, and exogenous operating-condition variables, and is used to correct the structural bias remaining in the backbone output within switching neighborhoods. In this way, segment-level data organization, multi-scale backbone prediction, joint parameter optimization, and switching-neighborhood residual compensation are integrated into a unified framework, making the method design correspond to the three key issues of local peaks, abrupt rate variations, and inter-segment distribution discrepancies.

1.4. Main Contributions

This paper focuses on the dynamic-response problem in switching neighborhoods during short-term active current prediction of synchronous reluctance motors under multiple operating conditions, and proposes a TPE-VMD-BiLSTM-TRC prediction method. The main contributions are summarized as follows.

- (1) A segment-level modeling framework for multi-condition active current prediction is constructed. The framework organizes data according to operating-condition boundaries, and performs sliding-window sample construction and VMD decomposition within the same operating-condition segment. This prevents cross-segment historical information from entering the input window, enabling model training and testing to be conducted on the basis of clearer operating-condition boundaries.
- (2) A switching-aware composite loss function is proposed. On the basis of the conventional reconstruction error, first-difference error and switching-neighborhood weighted error are introduced, so that the training objective simultaneously constrains the prediction amplitude, local variation trend, and key transition samples. This enhances the model's ability to track turning positions and rapid local variations.
- (3) A lightweight residual correction branch is designed to form the TRC switching-aware residual correction mechanism. This branch estimates a one-step residual compensation term using the backbone prediction, the most recent actual current, difference information, and exogenous operating-condition variables. It is used to compensate for local residual bias and response lag remaining in the backbone model within switching neighborhoods.
- (4) An experimental validation system corresponding to the methodological objectives is established. Through overall comparison on the primary dataset, mechanism ablation, segmented evaluation of switching and steady-state regions, and additional validation on held-out operating-condition segment, the effectiveness of the proposed method is verified from three aspects, namely overall prediction accuracy, dynamic response in switching neighborhoods, and stability across different operating-condition segments.

2. Prediction Method for Dynamic Response Enhancement Under Operating-Condition Switching

2.1. Problem Definition and Notation

Active current prediction for synchronous reluctance motors under multiple operating conditions needs to simultaneously consider sample boundaries, prediction targets, and the definition of switching neighborhoods. To avoid simply treating multiple operating-condition segments as a continuous stationary sequence, this paper takes the operating-condition segment as the basic organizational unit, and uniformly defines the input variables, prediction labels, effective sample set, and switching neighborhoods. This definition provides a consistent mathematical basis for subsequent segment-wise sample construction, VMD decomposition, switching-aware loss design, and region-wise evaluation [20].

Let the multi-condition time series obtained from the variable-frequency drive experiment of a synchronous reluctance motor be denoted as

$$\mathcal{D} = \{(\mathbf{x}_t, y_t)\}_{t=1}^N. \quad (1)$$

Here, $y_t = I_{\text{active}}(t)$ denotes the active current at time t , and $\mathbf{x}_t$ denotes the corresponding exogenous input vector. According to the structure of the experimental dataset, the basic input variables can be written as

$$\mathbf{x}_t = [U_t, F_t, |\text{spd}_t|, |\text{Torq}_t|, \text{dir}_t]^T. \quad (2)$$

where U_t is the average voltage, F_t is the frequency, $|\text{spd}_t|$ and $|\text{Torq}_t|$ are the magnitudes of speed and torque, respectively, and $\text{dir}_t = \text{sign}(\text{spd}_t)$ is the rotational direction feature. When $\text{spd}_t = 0$, dir_t is set to 1.

Since the experimental sequence consists of multiple operating-condition segments connected in series, the original sequence is first divided into S operating-condition segments according to the switching boundaries,

$$\mathcal{D} = \bigcup_{s=1}^S \mathcal{D}^{(s)}, \quad (3)$$

$$\mathcal{D}^{(s)} = \left\{ \left(\mathbf{x}_i^{(s)}, y_i^{(s)} \right) \right\}_{i=1}^{N_s}.$$

where N_s denotes the length of the s -th operating-condition segment.

This study considers a one-step prediction task, namely the prediction horizon is set to $H = 1$. Let L be the sliding-window length. For the q -th prediction target in the s -th operating-condition segment, the input window is defined as

$$\mathbf{X}_q^{(s)} = \left[\mathbf{x}_{q-L}^{(s)}, \mathbf{x}_{q-L+1}^{(s)}, \dots, \mathbf{x}_{q-1}^{(s)} \right], \quad (4)$$

and the corresponding prediction target is $y_q^{(s)}$. The set of effective prediction target positions is defined as

$$\mathcal{Q}^{(s)} = \{q \mid q = L + 1, L + 2, \dots, N_s\}. \quad (5)$$

Thus, the model task can be formulated as learning the following mapping relationship:

$$f : \mathbf{X}_q^{(s)} \mapsto \hat{y}_q^{(s)}. \quad (6)$$

Equation (6) indicates that this study focuses on one-step active current prediction based on a historical window within the same operating-condition segment. Unlike modeling tasks that only pursue the minimization of overall prediction error, the desired prediction results should not only maintain a small deviation over global samples, but also preserve good switching-neighborhood error suppression and variation-trend tracking capability for effective prediction samples after operating-condition switching. Therefore, after defining the effective prediction target set, the switching neighborhood and the approximately steady-state region are further distinguished. This partition enables local peaks and abrupt increases in the rate of change within switching neighborhoods to be explicitly characterized in both the training objective and experimental evaluation.

Accordingly, on the effective prediction target set $\mathcal{Q}^{(s)}$ of each operating-condition segment, the switching neighborhood and the approximately steady-state region are further defined as

$$\Omega_{\text{tr}}^{(s)} = \{q \in \mathcal{Q}^{(s)} \mid \text{ord}_s(q) \leq r\}, \quad (7)$$

$$\Omega_{\text{st}}^{(s)} = \mathcal{Q}^{(s)} \setminus \Omega_{\text{tr}}^{(s)}. \quad (8)$$

where $\text{ord}_s(q)$ denotes the ordinal index of target point q in the effective prediction sequence of operating-condition segment s , and r is the radius of the switching neighborhood. This definition ensures that the division between the switching region and the steady-state region is based on truly predictable samples, rather than on the initial unpredictable points at the beginning of the original segment. This is important for subsequent region-wise training and region-wise evaluation.

2.2. Data Preprocessing and Segment-Wise Sample Construction

Multi-condition experimental data are usually formed by concatenating operating-condition segments under different frequencies, loads, and rotational directions. The mean

value, fluctuation amplitude, and spectral structure may differ significantly among these segments. If sliding-window samples are directly constructed over the entire sequence, or if normalization and decomposition are performed on the full dataset before splitting the training and test sets, information on both sides of the switching boundary may be mixed into the same modeling process [21]. Therefore, this paper adopts a processing procedure of data cleaning, segmentation, data partitioning, and sliding-window sample construction, so that the sample construction process remains consistent with the structure of multi-condition data [22].

First, the raw experimental data are cleaned, parsed, and transformed into a unified numerical format. Since the main dataset and the held-out operating-condition segment are collected under different operating conditions, some operating parameters have different forms of expression in the original records. Therefore, unified encoding and consistent representation are required before modeling to ensure comparability among different data sources. It should be noted that this step is only used to unify the data format and numerical representation. It does not change the original operating-condition boundaries, nor does it merge different operating-condition segments into a continuous sequence for modeling.

Second, the original sequence is segmented according to the operating-condition switching boundaries, and the operating-condition index seg_id is generated. Suppose that the cleaned data are divided into S operating-condition segments. The s -th segment can be denoted as

$$\mathcal{D}^{(s)} = \left\{ \left(\mathbf{x}_i^{(s)}, y_i^{(s)} \right) \right\}_{i=1}^{N_s}, \quad s = 1, 2, \dots, S. \quad (9)$$

where $\mathbf{x}_i^{(s)}$ denotes the exogenous input vector of the i -th sample in the s -th segment, $y_i^{(s)}$ denotes the corresponding active current label, and N_s is the length of this segment. All training, testing, and held-out segment evaluations are performed using the operating-condition segment as the basic unit, thereby ensuring that information on both sides of the switching boundary does not overlap across different datasets.

For sample partitioning, the data are divided at the operating-condition segment level into a development set and an independent test set. Within the development set, several complete segments are further retained as the inner validation subset for hyperparameter selection, loss-weight determination, and early stopping. The independent test set is used only for final performance evaluation and is not involved in model or parameter selection. Let the development segment set and the independent test segment set be $\mathcal{S}_{\text{dev}}$ and $\mathcal{S}_{\text{test}}$, respectively. They satisfy

$$\mathcal{S}_{\text{dev}} \cap \mathcal{S}_{\text{test}} = \emptyset, \quad \mathcal{S}_{\text{dev}} \cup \mathcal{S}_{\text{test}} = \{1, 2, \dots, S\}. \quad (10)$$

The development set is further split into an inner-training subset $\mathcal{S}_{\text{tr}}$ and an inner-validation subset $\mathcal{S}_{\text{val}}$, with $\mathcal{S}_{\text{tr}} \cap \mathcal{S}_{\text{val}} = \emptyset$. This inner validation subset is used only during model selection and does not overlap with the independent test segments.

To avoid cross-segment information mixing, samples near segment boundaries are not discarded. Instead, a strict constraint is imposed during sample construction, requiring both the input window and the prediction target to be completely located within the same seg_id . In this way, effective data within each segment are retained, while the independence between training and testing is maintained.

Considering that speed and torque may change signs under forward-reverse operating conditions, and that magnitude variation and direction variation play different roles in modeling, the relevant input variables are further decoupled into magnitude and direction

terms. Let the original speed and torque be spd_t and Torq_t , respectively. The transformed variables are constructed as

$$\text{spd}_t^* = |\text{spd}_t|, \quad \text{Torq}_t^* = |\text{Torq}_t|, \quad \text{dir}_t = \text{sign}(\text{spd}_t). \quad (11)$$

where dir_t is set to 1 when $\text{spd}_t = 0$. Accordingly, the basic input feature vector can be written as

$$\mathbf{x}_t = [U_t, F_t, \text{spd}_t^*, \text{Torq}_t^*, \text{dir}_t]^T. \quad (12)$$

where U_t denotes the average voltage, F_t denotes the frequency, spd_t^* and Torq_t^* denote the magnitudes of speed and torque, respectively, and dir_t denotes the rotational direction feature. This treatment helps reduce the interference of frequent sign changes on model stability while preserving the influence of rotational direction on the dynamic evolution of active current.

For normalization, only the statistics of the inner-training subset within the development set are used to transform continuous variables. Let the continuous input variable vector be denoted as $\mathbf{z}$. Its normalized form is

$$\tilde{\mathbf{z}} = \frac{\mathbf{z} - \boldsymbol{\mu}_{\text{tr}}}{\boldsymbol{\sigma}_{\text{tr}}}. \quad (13)$$

where $\boldsymbol{\mu}_{\text{tr}}$ and $\boldsymbol{\sigma}_{\text{tr}}$ are the mean and standard deviation of the continuous variables in the inner-training subset, respectively. The inner-validation subset, independent test set, and held-out operating-condition segment strictly reuse these statistics, without refitting normalization parameters. For the discrete direction feature dir_t , forward and reverse rotations are encoded as +1 and −1, respectively. This feature is used as a raw numerical input and is not processed by one-hot encoding or an embedding layer.

After normalization, non-cross-segment one-step prediction samples are constructed according to the definitions in Section 2.1. For each operating-condition segment, a sample is generated only when the complete input window and the corresponding prediction target are located within the same `seg_id`. Therefore, historical information before and after an operating-condition boundary is not mixed into the same input window.

The switching-neighborhood and steady-state subsets used in the composite loss and region-wise evaluation are also defined on effective prediction samples rather than on the original segment heads. This treatment ensures that the training objective and evaluation metrics are calculated on truly predictable samples. In this way, the preprocessing procedure provides consistent data boundaries for segment-wise VMD decomposition, model training, switching-aware loss calculation, and region-wise error analysis.

2.3. Segment-Wise Variational Mode Decomposition

The active current sequence in switching neighborhoods usually contains slow-varying trends, local peaks, and short-term disturbances simultaneously [23]. If the raw sequence is directly fed into a temporal network, dynamic components at different scales may be mixed in the same latent space, increasing the difficulty of learning local rapid variations. Therefore, this paper independently performs VMD within each operating-condition segment, and represents the active current sequence as several modal components with relatively concentrated frequency bands [24].

Let the normalized active current sequence in the s -th operating-condition segment be

$$\tilde{y}^{(s)}(t), \quad t = 1, 2, \dots, N_s. \quad (14)$$

where N_s denotes the length of this operating-condition segment. VMD decomposes this sequence into K modal components $u_k^{(s)}(t)$, satisfying the reconstruction constraint

$$\tilde{y}^{(s)}(t) = \sum_{k=1}^K u_k^{(s)}(t). \quad (15)$$

where $u_k^{(s)}(t)$ denotes the k -th modal component in the s -th operating-condition segment. To simplify the subsequent formula derivation, the segment superscript (s) is omitted hereafter when no ambiguity is introduced.

VMD transforms the original signal decomposition problem into a constrained variational problem by constraining each mode to be as narrow-band as possible around its center frequency. The problem can be written as

$$\min_{\{u_k\}, \{\omega_k\}} \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2, \quad (16)$$

$$\text{s.t.} \quad \sum_{k=1}^K u_k(t) = \tilde{y}(t). \quad (17)$$

where ω_k denotes the center frequency of the k -th mode, $\delta(t) + \frac{j}{\pi t}$ corresponds to the Hilbert transform kernel, and $\partial_t(\cdot)$ represents the time derivative of the analytic signal after demodulation to the baseband. The above constrained variational problem indicates that, while satisfying the reconstruction constraint of the original sequence, each mode is encouraged to be concentrated around its own center frequency, thereby obtaining modal components with relatively separated frequency bands.

In practical computation, by introducing the Lagrange multiplier and quadratic penalty term, the constrained problem is transformed into an augmented Lagrangian form that can be solved iteratively. The alternating direction method of multipliers is then used to alternately update the modal components and their center frequencies in the frequency domain. At the $(n+1)$ -th iteration, the frequency-domain update of the k -th mode can be written as

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{y}(\omega) - \sum_{i \neq k} \hat{u}_i^n(\omega) + \frac{1}{2} \hat{\lambda}^n(\omega)}{1 + 2\alpha(\omega - \omega_k^n)^2}. \quad (18)$$

The corresponding center frequency is updated as

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}{\int_0^\infty \left| \hat{u}_k^{n+1}(\omega) \right|^2 d\omega}. \quad (19)$$

where α is the penalty factor, and $\hat{\lambda}^n(\omega)$ is the frequency-domain form of the Lagrange multiplier at the n -th iteration. Through the above alternating updates, a set of modes satisfying the reconstruction constraint and having relatively compact bandwidths can be gradually obtained.

In this paper, the two key parameters of VMD are the number of modes K and the penalty factor α . Specifically, K determines the granularity of scale decomposition. When K is too small, information from different frequency bands may be mixed within the same mode. When K is too large, over-decomposition may occur and redundant modes may be introduced. The penalty factor α controls the strength of the narrow-band constraint. A small α may lead to excessively wide modal bandwidths, whereas an overly large α may split components that should remain continuous. Therefore, these two parameters are not

manually fixed, but are jointly optimized with the network hyperparameters in the TPE search process.

This paper adopts segment-wise VMD rather than unified decomposition before data partitioning. This treatment is consistent with the segment-level sample organization and reduces spurious information sharing caused by spectral-structure discrepancies among operating-condition segments. For the present task, VMD is not used merely as a smoothing tool, but to decompose the active current sequence into subsequences more suitable for mode-wise prediction, thereby providing a stable multi-scale basis for BiLSTM backbone modeling and switching-neighborhood error correction.

2.4. Mode-Wise BiLSTM Backbone Prediction

After segment-wise VMD decomposition, the original active current sequence is represented by several modal components with relatively independent scales. Compared with direct end-to-end prediction on the original sequence, mode-wise modeling reduces the difficulty for a single network to simultaneously fit slow-varying trends, local oscillations, and disturbance details. In this paper, a BiLSTM predictor is constructed for each modal component, and the backbone prediction output is obtained through modal reconstruction [25].

Let the k -th modal component in the s -th operating-condition segment be $u_k^{(s)}(t)$, the sliding-window length be L , and the prediction horizon be $H = 1$. For the target position q , the historical window of the corresponding mode is written as

$$\mathbf{u}_{k,q}^{(s)} = \left[u_k^{(s)}(q-L), u_k^{(s)}(q-L+1), \dots, u_k^{(s)}(q-1) \right]^T. \quad (20)$$

The exogenous-variable window synchronized with the modal history is denoted as

$$\mathbf{x}_q^{(s)} = \left[\mathbf{x}_{q-L}^{(s)}, \mathbf{x}_{q-L+1}^{(s)}, \dots, \mathbf{x}_{q-1}^{(s)} \right]. \quad (21)$$

where $\mathbf{x}_t^{(s)}$ denotes the exogenous input vector at sampling time t in the s -th operating-condition segment. In the present setting, the common backbone input mainly consists of average voltage, frequency, speed magnitude, torque magnitude, and rotational direction, namely

$$\mathbf{x}_t = [U_t, F_t, |\text{spd}_t|, |\text{Torq}_t|, \text{dir}_t]^T. \quad (22)$$

Thus, the input sample of the k -th mode at target position q can be expressed as

$$\mathbf{Z}_{k,q}^{(s)} = \left[\mathbf{x}_q^{(s)}, \mathbf{u}_{k,q}^{(s)} \right]. \quad (23)$$

Here, $\mathbf{Z}_{k,q}^{(s)}$ contains both the historical evolution of the k -th mode and the time-aligned exogenous operating-condition information, allowing a one-step prediction mapping to be established at each modal scale.

In terms of network structure, BiLSTM extracts temporal dependencies within the sliding window through forward and backward recurrent paths [26]. Since this study adopts a one-step prediction setting, the backward recurrence of BiLSTM only operates within the known historical window and does not use any actual observation after the prediction time. Therefore, no future information is introduced. This design better exploits the dependencies among sampling points within the historical window and provides a more stable temporal representation for modal prediction in local transient regions.

Let the forward and backward hidden states be denoted as $\vec{\mathbf{h}}_{k,q}$ and $\overleftarrow{\mathbf{h}}_{k,q}$, respectively. The bidirectional feature representation is then given by

$$\mathbf{h}_{k,q}^{(b)} = [\vec{\mathbf{h}}_{k,q}; \overleftarrow{\mathbf{h}}_{k,q}]. \quad (24)$$

After obtaining the bidirectional hidden representation, the one-step prediction of the k -th mode at target position q is generated through a regression head,

$$\hat{u}_{k,q}^{(s)} = \phi(\mathbf{h}_{k,q}^{(b)}). \quad (25)$$

where $\phi(\cdot)$ denotes the regression mapping composed of fully connected layers. The backbone output is then obtained by linearly reconstructing the prediction results of all modes,

$$\hat{y}_{\text{base},q}^{(s)} = \sum_{k=1}^K \hat{u}_{k,q}^{(s)}. \quad (26)$$

where $\hat{y}_{\text{base},q}^{(s)}$ denotes the backbone prediction result at target position q in the s -th operating-condition segment. This process follows the basic idea of multi-scale decomposition modeling, in which sequence complexity is first reduced at the modal level and the overall evolution of the target variable is then recovered through linear reconstruction.

From the perspective of the overall method, $\hat{y}_{\text{base}}$ obtained in this subsection serves as the backbone prediction output of the complete TPE-VMD-BiLSTM-TRC framework. Segment-wise VMD reduces multi-scale component mixing, mode-wise BiLSTM establishes one-step prediction mappings for individual modes, and modal reconstruction recovers the overall evolution trend of the active current. Since the backbone output may still retain peak attenuation, response lag, or systematic residuals in switching neighborhoods, it is not used as the final output directly. Instead, it provides the basis for the subsequent composite loss constraint and residual correction branch.

2.5. Switching-Aware Composite Loss Function

The conventional mean squared error mainly constrains the amplitude deviation between predicted and actual values. However, in multi-condition data, the number of steady-state samples is usually much larger than that of switching-neighborhood samples, so the error contribution of local peaks and rapid transition samples can be easily diluted [27]. Meanwhile, constraining only the amplitude error cannot ensure the consistency between the predicted and actual sequences in terms of local slope and turning positions [28]. Therefore, this paper constructs a composite loss function consisting of reconstruction error, first-difference error, and switching-neighborhood weighted error, so that the training objective can jointly consider overall fitting, variation-trend tracking, and switching-sample enhancement.

Let the backbone output at target position q in the s -th operating-condition segment be $\hat{y}_{\text{base},q}^{(s)}$, and let the output of the residual correction branch be $r_q^{(s)}$. The final prediction of the complete framework is obtained through switch-gated residual correction:

$$\hat{y}_q^{(s)} = \hat{y}_{\text{base},q}^{(s)} + g_q^{(s)} r_q^{(s)}. \quad (27)$$

where $g_q^{(s)}$ is the switching-gate weight. For samples outside the switching neighborhood, $g_q^{(s)} = 0$, so the final output is identical to the backbone prediction. This design concentrates the correction effect on local switching regions and avoids unnecessary disturbance to steady-state samples. When the residual correction branch is not introduced, setting $r_q^{(s)} = 0$ reduces the final prediction to the backbone output.

First, the conventional reconstruction error is defined. Let $y_q^{(s)}$ denote the actual active current. The reconstruction error over all effective prediction samples is written as

$$L_{\text{rec}} = \frac{1}{N_{\text{all}}} \sum_{s \in \mathcal{S}_{\text{tr}}} \sum_{q \in Q^{(s)}} \left(\hat{y}_q^{(s)} - y_q^{(s)} \right)^2. \quad (28)$$

where N_{all} denotes the total number of effective prediction samples. This term constrains the numerical consistency between the predicted and actual values and serves as the basic component of the training objective.

To enhance the model's ability to track local variation trends, a first-difference error term is further introduced. For two adjacent effective target points q^- and q in the same operating-condition segment, where q^- denotes the effective prediction target immediately before q in $Q^{(s)}$, the actual and predicted increments are defined as

$$\Delta y_q^{(s)} = y_q^{(s)} - y_{q^-}^{(s)}, \quad (29)$$

$$\Delta \hat{y}_q^{(s)} = \hat{y}_q^{(s)} - \hat{y}_{q^-}^{(s)}. \quad (30)$$

Let $Q_{\Delta}^{(s)}$ denote the sample set excluding the first effective prediction point. The first-difference loss is defined as

$$L_{\text{diff}} = \frac{1}{N_{\Delta}} \sum_{s \in \mathcal{S}_{\text{tr}}} \sum_{q \in Q_{\Delta}^{(s)}} \left(\Delta \hat{y}_q^{(s)} - \Delta y_q^{(s)} \right)^2. \quad (31)$$

where N_{Δ} denotes the total number of effective samples for which the first difference can be calculated. Unlike the conventional reconstruction error, this term does not directly constrain the absolute amplitude deviation. Instead, it emphasizes the consistency between the predicted and actual sequences in adjacent-time increments, thereby improving the model's ability to track local slope changes and turning positions.

Considering that switching-neighborhood samples account for only a small proportion of the total samples, their error contribution may still be diluted by abundant steady-state samples without explicit enhancement. Therefore, the switching-neighborhood indicator function is defined as

$$m_q^{(s)} = \begin{cases} 1, & q \in \Omega_{\text{tr}}^{(s)}, \\ 0, & q \in \Omega_{\text{st}}^{(s)}. \end{cases} \quad (32)$$

where $\Omega_{\text{tr}}^{(s)}$ and $\Omega_{\text{st}}^{(s)}$ denote the switching neighborhood and the approximately steady-state region in the s -th operating-condition segment, respectively. Based on this indicator, the switch-weighted reconstruction loss is defined as

$$L_{\text{wrec}} = \frac{\sum_{s \in \mathcal{S}_{\text{tr}}} \sum_{q \in Q^{(s)}} \left(1 + \lambda_2 m_q^{(s)} \right) \left(\hat{y}_q^{(s)} - y_q^{(s)} \right)^2}{\sum_{s \in \mathcal{S}_{\text{tr}}} \sum_{q \in Q^{(s)}} \left(1 + \lambda_2 m_q^{(s)} \right)}. \quad (33)$$

The number of switching-neighborhood samples is

$$N_{\text{tr}} = \sum_{s \in \mathcal{S}_{\text{tr}}} \sum_{q \in Q^{(s)}} m_q^{(s)} \quad (34)$$

and is reported for region-wise evaluation. Compared with a separate local loss term, L_{wrec} keeps all samples in the reconstruction objective while increasing the contribution of switching-neighborhood samples through λ_2 .

The switching-aware auxiliary loss is then formulated as

$$L_{sw} = L_{wrec} + \lambda_3 L_{diff}. \quad (35)$$

Rather than using a simple linear summation of independent terms, the final training objective blends the base reconstruction loss and the auxiliary switching-aware loss with a warm-up/ramp coefficient:

$$L_{total} = L_{rec} + \lambda_1 a_e \Delta L_{sw}, \quad (36)$$

where a_e is an epoch-dependent coefficient and ΔL_{sw} denotes the capped positive auxiliary increment. In the implementation, an 8-epoch warm-up, an 8-epoch linear ramp, and an auxiliary-increment cap are used to prevent the local constraint from destabilizing early training. Here, λ_1 controls the overall participation of the auxiliary switching-aware loss, λ_2 controls the reconstruction weight of switching-neighborhood samples in L_{wrec} , and λ_3 controls the first-difference consistency term.

The final values are set to $\lambda_1 = 0.10$, $\lambda_2 = 0.25$, and $\lambda_3 = 0.05$. These values are selected according to the inner-validation results within the development set, and the independent test set is not used for determining the loss weights.

In terms of mechanism, L_{rec} ensures the basic numerical fitting ability of the model over all samples and avoids excessive bias toward local regions at the expense of global prediction stability. L_{diff} constrains the first-order increments between adjacent effective prediction points, making the local variation trend of the predicted sequence closer to that of the actual sequence and reducing turning-point lag caused by abrupt rate variations. L_{wrec} increases the contribution of key transition samples through the switching-neighborhood weight, preventing local peaks and short-term post-switching errors from being diluted by numerous steady-state samples. With these terms combined, the training objective changes from minimizing a single amplitude error to jointly considering overall fitting, dynamic tracking, and switching-neighborhood error suppression.

2.6. Residual Correction Branch

Segment-wise VMD and mode-wise BiLSTM can improve the overall fitting performance for multi-condition sequences. However, within several sampling points after frequency steps, load transients, or forward-reverse switching, the reconstructed backbone output may still contain structural biases such as persistent underestimation, overestimation, or peak compression. Residual learning transforms the modeling target from the complete prediction value to the compensation term between the baseline output and the actual value [29]. Based on this idea, a lightweight residual correction branch is introduced after the backbone output to learn a one-step residual compensation term and correct local biases in switching neighborhoods.

Let the backbone output at target position q in the s -th operating-condition segment be $\hat{y}_{base,q}^{(s)}$. The residual correction branch does not relearn the complete prediction value, but learns the compensation between the backbone output and the actual value. Therefore, the input vector of the residual correction branch is constructed as

$$\mathbf{z}_q^{(s)} = [\hat{y}_{base,q}^{(s)}, y_{q-1}^{(s)}, \Delta y_{q-1}^{(s)}, \mathbf{x}_{q-1}^{(s)}]^T. \quad (37)$$

where $y_{q-1}^{(s)}$ denotes the actual active current at the most recent observable time before the target position, and $\Delta y_{q-1}^{(s)}$ denotes the corresponding first-order difference, namely

$$\Delta y_{q-1}^{(s)} = y_{q-1}^{(s)} - y_{q-2}^{(s)}. \quad (38)$$

$\mathbf{x}_{q-1}^{(s)}$ denotes the exogenous input vector aligned with this time step.

The input variables in $\mathbf{z}_q^{(s)}$ are not simply stacked, but are constructed according to the possible source of the current one-step residual. Specifically, $\hat{y}_{\text{base},q}^{(s)}$ represents the preliminary estimate of the backbone model for the current target point, $y_{q-1}^{(s)}$ provides the latest actual current state, $\Delta y_{q-1}^{(s)}$ reflects the local variation trend in the switching neighborhood, and $\mathbf{x}_{q-1}^{(s)}$ supplies exogenous operating-condition information related to voltage, frequency, speed, torque, and rotational direction. With these variables, the residual branch can perform lightweight correction for peak bias and local lag that remain uncompensated in the backbone output.

Let the residual correction branch be denoted as $g(\cdot)$. Its output is defined as a one-step residual compensation term:

$$r_q^{(s)} = g\left(\mathbf{z}_q^{(s)}; \theta_r\right), \quad (39)$$

where θ_r denotes the parameter set of the residual correction branch. Accordingly, the final prediction result is expressed as

$$\hat{y}_q^{(s)} = \hat{y}_{\text{base},q}^{(s)} + r_q^{(s)}. \quad (40)$$

In the complete TPE-VMD-BiLSTM-TRC framework, the residual branch does not replace the backbone predictor. Instead, it performs a secondary correction for structural errors remaining near switching neighborhoods while retaining the multi-scale prediction advantage of the backbone model.

Since this paper focuses on a one-step short-term prediction task and the available training samples are limited, the residual correction branch adopts a lightweight structure. The function $g(\cdot)$ is designed as a two-layer fully connected mapping:

$$\mathbf{h}_{r,q}^{(s)} = \sigma\left(\mathbf{W}_{r1}\mathbf{z}_q^{(s)} + \mathbf{b}_{r1}\right), \quad (41)$$

$$r_q^{(s)} = \mathbf{W}_{r2}\mathbf{h}_{r,q}^{(s)} + \mathbf{b}_{r2}. \quad (42)$$

where $\mathbf{W}_{r1}$ and $\mathbf{W}_{r2}$ are the weight matrices, $\mathbf{b}_{r1}$ and $\mathbf{b}_{r2}$ are the bias terms of the residual correction branch, and $\sigma(\cdot)$ denotes the nonlinear activation function. This structure can model local errors nonlinearly without introducing a large additional parameter burden.

When the model is reduced to a version without the residual correction branch, we set

$$r_q^{(s)} = 0. \quad (43)$$

Then,

$$\hat{y}_q^{(s)} = \hat{y}_{\text{base},q}^{(s)}. \quad (44)$$

That is, the final prediction result reduces to the backbone output. This form facilitates the independent evaluation of the residual correction branch in subsequent mechanism ablation experiments.

In summary, the residual correction branch is not a repeated modeling of the backbone predictor, but a local compensation mechanism designed for the remaining structural errors in switching neighborhoods. The composite loss strengthens switching samples and variation-trend constraints from the training-objective level, while the residual branch provides an additional compensation channel from the output-structure level. Their combined effect enables the complete model to better suppress switching-neighborhood errors, track rapid transitions, and compensate for post-switching systematic residuals without significantly increasing the complexity of the backbone model.

2.7. TPE-Based Joint Optimization Strategy

The VMD decomposition parameters and BiLSTM network parameters jointly affect the backbone prediction performance. The number of modes K and the penalty factor α determine the decomposition scale and frequency-band constraint strength, while the sliding-window length, number of hidden units, number of network layers, and learning rate affect the temporal modeling capability of the network for each modal component. If the decomposition parameters are fixed first and the network parameters are tuned separately, the model may remain in a locally matched configuration. Therefore, this paper adopts TPE to jointly search the key parameters of VMD and BiLSTM, so as to improve the coordination between multi-scale decomposition and temporal prediction [30].

Let the parameter vector to be searched be denoted as

$$\theta = \{K, \alpha, L, n_h, n_l, d, \eta, b\}. \quad (45)$$

where K and α denote the number of modes and the penalty factor of VMD, respectively; L denotes the sliding-window length; n_h denotes the number of hidden units; n_l denotes the number of network layers; d denotes the dropout coefficient; η denotes the learning rate; and b denotes the batch size. For any candidate parameter configuration θ , VMD decomposition is first performed within each operating-condition segment of the development set. The corresponding VMD-BiLSTM backbone model is trained on the inner-training subset and evaluated on the inner-validation subset.

Since final evaluation is conducted only on the independent test set, the TPE search objective is calculated from the inner-validation prediction error rather than from the test data. In this paper, the objective function is defined as the root mean square error of active current prediction over all effective prediction samples in the inner-validation segment set:

$$J(\theta) = \sqrt{\frac{1}{N_{\text{val}}} \sum_{s \in \mathcal{S}_{\text{val}}} \sum_{q \in \mathcal{Q}^{(s)}} \left(\hat{y}_{\text{base},q}^{(s)}(\theta) - y_q^{(s)} \right)^2}. \quad (46)$$

where

$$N_{\text{val}} = \sum_{s \in \mathcal{S}_{\text{val}}} |\mathcal{Q}^{(s)}|. \quad (47)$$

Here, $\hat{y}_{\text{base},q}^{(s)}(\theta)$ denotes the backbone prediction result obtained under the parameter configuration θ , and $y_q^{(s)}$ is the corresponding actual active current. This definition indicates that TPE does not directly participate in online prediction. Instead, during the training stage, it compares the inner-validation prediction errors of different candidate configurations and selects more promising parameter combinations.

Unlike directly modeling $p(J | \theta)$, TPE estimates the conditional distribution $p(\theta | J)$ in a piecewise manner. Let J^* be the threshold determined by historical evaluations. The evaluated samples are divided into good and poor parameter sets, whose conditional densities are estimated as

$$l(\theta) = p(\theta | J(\theta) \leq J^*), \quad (48)$$

$$g(\theta) = p(\theta | J(\theta) > J^*). \quad (49)$$

Here, $l(\theta)$ and $g(\theta)$ describe the parameter distributions in regions with better and poorer objective values, respectively. The next candidate configuration is obtained by maximizing the density ratio:

$$\theta_{\text{next}} = \arg \max_{\theta} \frac{l(\theta)}{g(\theta)}. \quad (50)$$

A larger density ratio indicates a higher probability of appearing in the good-sample region and a lower probability of appearing in the poor-sample region. In the proposed framework, TPE searches integer, discrete, and continuous variables covering VMD decomposition parameters, BiLSTM structural parameters, and training parameters. It is embedded as an outer-loop training strategy for the VMD-BiLSTM backbone rather than an online prediction module.

After obtaining the optimal parameter configuration, it is denoted as

$$\theta^* = \arg \min_{\theta} J(\theta). \quad (51)$$

Then, θ^* is used for final model training within the development set, and the independent test set is used only once for final evaluation. The recorded TPE search process contains 18 objective-function evaluations, and the parameter set with the minimum inner-validation RMSE is selected. The final selected parameters are window length $L = 29$, hidden size $n_h = 48$, one BiLSTM layer, dropout $d = 0$, learning rate $\eta = 0.004127$, batch size $b = 32$, VMD mode number $K = 6$, and penalty factor $\alpha = 1663.37$. Meanwhile, VMD decomposition is independently performed within the respective operating-condition segments of the development and test sets, rather than on the full sequence before data partitioning. This maintains clear boundaries among parameter search, model training, and final testing, and provides a stable backbone output for the subsequent switching-aware composite loss and TRC residual correction.

2.8. Overall Framework and Procedure

Figure 1 summarizes the overall TPE-VMD-BiLSTM-TRC procedure. Segment-level data organization and non-cross-segment window construction address inter-segment distribution discrepancies and information leakage risks; segment-wise VMD and mode-wise BiLSTM reduce the multi-scale modeling difficulty of nonstationary sequences; TPE-based joint optimization improves parameter matching; and the switching-aware composite loss together with TRC residual correction enhances the response to switching-neighborhood errors and abrupt rate variations. The specific procedure is as follows.

Step 1: Data cleaning and segment indexing. The raw data are cleaned, unified into numerical form, and divided into operating-condition segments according to switching boundaries.

Step 2: Non-cross-segment sample construction. Development and independent test sets are divided at the segment level, and one-step sliding-window samples are constructed within each segment after feature construction and inner-training-based normalization.

Step 3: Segment-wise VMD. VMD is independently performed within each operating-condition segment to obtain mode components without sharing decomposition information between training and test sets.

Step 4: Mode-wise BiLSTM prediction. Each decomposed mode is predicted by a BiLSTM predictor, and all mode-wise outputs are linearly reconstructed to obtain the backbone prediction.

Step 5: TPE-based joint optimization. TPE searches the VMD decomposition parameters and BiLSTM hyperparameters using the inner-training and inner-validation subsets within the development set.

Step 6: Switching-aware composite loss. The model is trained with a switching-aware auxiliary objective combining switch-weighted reconstruction and first-difference consistency.

Step 7: TRC residual correction. A lightweight switch-gated residual correction branch compensates for the remaining structural errors of the backbone output in switching neighborhoods.

Step 8: Testing and evaluation. The trained model is evaluated using RMSE and MAE over all samples, switching-neighborhood samples, and steady-state samples. MAPE, event-level peak-amplitude error, and switching-neighborhood first-difference RMSE are further reported as additional metrics, with additional validation on a held-out operating-condition segment.

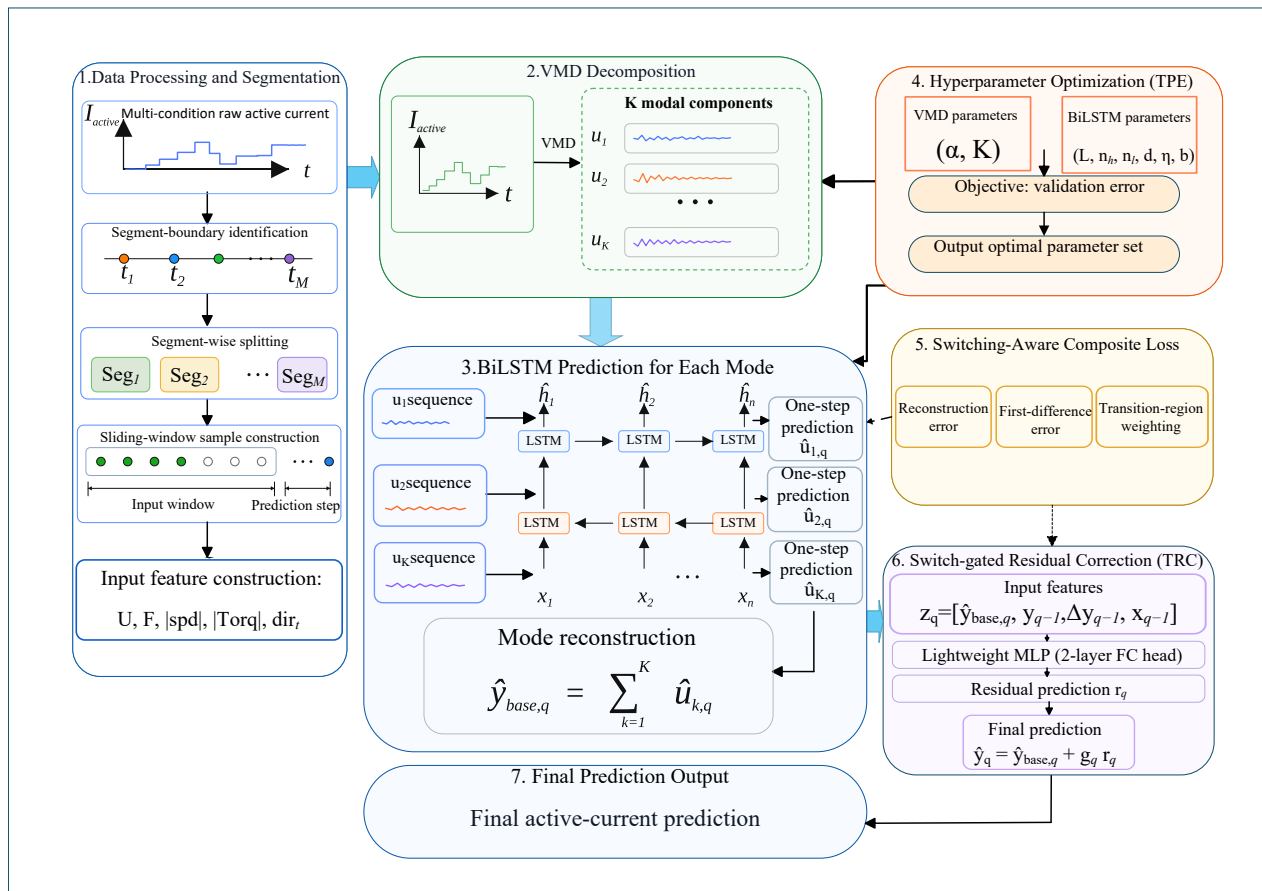


Figure 1. Framework of the proposed TPE-VMD-BiLSTM-TRC method.

In summary, the proposed TPE-VMD-BiLSTM-TRC method integrates segment-level data organization, segment-wise VMD decomposition, mode-wise BiLSTM backbone prediction, TPE-based joint optimization, switching-aware composite loss, and TRC residual correction into a unified prediction framework. Segment-level data organization and segment-wise decomposition are used to alleviate inter-segment distribution discrepancies. Multi-scale backbone prediction and TPE optimization improve the stability of baseline prediction, while the composite loss and residual correction enhance the dynamic response to switching-neighborhood errors and abrupt rate variations. This procedure establishes a clear correspondence between the method design in this section and the three core issues raised in the introduction, and also provides the methodological basis for subsequent comparative experiments, mechanism ablation, and region-wise evaluation.

3. Experimental Design and Evaluation Metrics

3.1. Data Description

The experimental data used in this study are obtained from multi-condition operating records of a synchronous reluctance motor under variable-frequency drive. During the experiment, operating conditions such as frequency steps, load switching, and alternating forward-reverse rotation were introduced, causing the system to repeatedly transition

between steady-state operation and switching transients. As a result, an active current sequence with evident nonstationary characteristics was formed. The raw data are stored in tabular form, where each row corresponds to one sampling instant and records the measurable drive-side operating variables and the corresponding active current response. The input information mainly includes voltage, frequency, speed, load torque, and rotational direction, while the prediction target is the active current I_{active} .

The data were collected from a synchronous reluctance motor test platform. The tested motor is rated at 18.5 kW and 1000 r/min, with 6 poles and a 200-frame size. During the experiment, active current and related operating variables were obtained from the drive-side recorded data. The operating conditions cover frequency variation, load switching, forward/reverse rotation, and multi-condition switching processes. Figure 2 and Table 1 summarize the confirmed platform and dataset information.



Figure 2. Experimental platform of the synchronous reluctance motor test bench.

Table 1. Experimental platform and dataset information.

Item	Description
Motor type	Synchronous reluctance motor
Rated power/speed	18.5 kW/1000 r/min
Pole number/frame size	6 poles/200 frame
Data source	Drive-side recorded operating variables
Acquired variables	Active current, voltage, frequency, speed, load torque, and direction
Operating conditions	Frequency variation, load switching, and forward/reverse rotation

In terms of data structure, the raw sequence consists of multiple consecutive operating-condition segments. The operating conditions within each segment are relatively continuous, whereas the boundaries between segments usually correspond to frequency changes, load adjustments, or rotational direction switching. Consequently, the mean value, fluctuation amplitude, and local variation pattern of active current may differ among different segments. In addition, local peaks, abrupt rate variations, and error concentration are more likely to occur within several sampling points after switching. Therefore, this study uses the operating-condition segment as the basic unit for data partitioning, sample construction,

and result statistics, ensuring that the experimental setting is consistent with the actual structure of the multi-condition sequence.

After data cleaning, field unification, and operating-condition segmentation, 3254 valid samples are retained in the primary dataset, which is divided into 18 valid operating-condition segments. All sliding-window samples are constructed within the same operating-condition segment, and the input window and prediction target are not allowed to cross different segments. For sequential data with a natural grouping structure, ignoring intra-group correlation during data partitioning may introduce implicit information leakage between the training and test sets, leading to an overly optimistic estimation of model performance [31,32]. Therefore, this paper adopts a segment-level partitioning strategy, so that training, testing, and subsequent region-wise evaluation are all conducted under consistent data boundaries.

In the experiment, the development set is used for model training, inner validation, and parameter selection, while the independent test set is used only for final performance evaluation. Four representative test segments are selected as the main objects for overall comparison, mechanism ablation, and region-wise evaluation. Another test segment that is not used in the main result presentation is retained as an additional held-out operating-condition segment. This segment is not used for training, inner validation, hyperparameter selection, or loss-weight determination. This setting is used to examine the prediction consistency of the model across different operating-condition segments, with particular attention paid to turning-point tracking, dynamic response, and local bias control in switching neighborhoods. The computational scripts were implemented in Python, and the manuscript source files and code were edited using Visual Studio Code 1.126.0 (Microsoft Corporation, Redmond, WA, USA).

3.2. Comparison Models and Experimental Settings

To verify the effectiveness of the proposed TPE-VMD-BiLSTM-TRC method in multi-condition active current prediction for synchronous reluctance motors, three types of experiments are designed in this paper: overall comparison on the primary dataset, mechanism ablation, and additional validation on a held-out operating-condition segment. All three experiments use the same data preprocessing procedure, segment-wise window construction strategy, and evaluation metrics to ensure the comparability of the results.

In the overall comparison experiment on the primary dataset, LSTM, BiLSTM, TPE-BiLSTM, VMD-BiLSTM, VMD-SSA-BiLSTM, VMD-IPOA-BiLSTM, TPE-VMD-BiLSTM, and the proposed method are selected as comparison models. Among them, LSTM and BiLSTM are used to compare the basic unidirectional and bidirectional temporal modeling capabilities. TPE-BiLSTM is used to examine the improvement brought by automatic hyperparameter optimization to the basic network. VMD-BiLSTM is used to analyze the modeling effect of multi-scale decomposition on nonstationary active current sequences. VMD-SSA-BiLSTM and VMD-IPOA-BiLSTM are used to compare the prediction performance of different intelligent optimization strategies coupled with VMD-BiLSTM. TPE-VMD-BiLSTM serves as the backbone comparison model after removing the switching-aware composite loss and residual correction branch. Through these model settings, performance differences can be analyzed from four aspects: temporal structure, multi-scale decomposition, parameter optimization, and switching enhancement.

The overall comparison experiment on the primary dataset takes the first four representative test segments as the statistical objects. First, the effective prediction samples of each operating-condition segment are aligned and concatenated, and the overall RMSE and MAE are calculated to evaluate the comprehensive prediction accuracy of each model on multi-condition sequences. Second, a fixed local zoom-in window is set for each

operating-condition segment to observe prediction details at typical switching positions, local transient positions, and plateau-holding stages. Finally, error metrics are calculated over all samples, switching-neighborhood samples, and steady-state samples, respectively, to analyze the differences among models in global fitting and local dynamic response.

The reason for using four representative test segments instead of plotting the entire test sequence as a single continuous curve is that the test data are organized by operating-condition boundaries rather than by a single stationary temporal process. Different segments correspond to different frequency, load, or rotational-direction conditions, and their boundaries contain the switching neighborhoods that are the main focus of this study. If all test segments are displayed as one uninterrupted curve, the differences among operating-condition segments and the local dynamic responses near switching positions may be visually weakened. Therefore, the four representative test segments are used for the main comparison to preserve segment boundaries and to show prediction behavior under several typical switching patterns. The quantitative metrics are still calculated on the aligned effective prediction samples of these segments, rather than on selected local windows only.

It should be noted that the overall comparison experiment and the mechanism ablation experiment have different evaluation objectives. The overall comparison experiment is used to compare the comprehensive performance of different types of models on the representative test segments, and therefore includes more comparison models. The mechanism ablation experiment is used to analyze the independent effects of the switching-aware composite loss and the TRC residual correction branch, and is therefore constructed only around the same TPE-VMD-BiLSTM backbone. To ensure that the prediction target points of all models within each experiment are completely consistent, RMSE and MAE are calculated based on the common effective prediction sample set within each experiment. Since the participating models and aligned effective target-point sets differ between experiments, the corresponding metrics are used only for comparison within their respective experiments and are not intended for item-by-item cross-experiment comparison.

In the mechanism ablation experiment, to avoid interference caused by different backbone configurations, four models are constructed around the same TPE-VMD-BiLSTM backbone: TPE-VMD-BiLSTM, TPE-VMD-BiLSTM with the switching-aware composite loss, TPE-VMD-BiLSTM with the residual correction branch, and the complete TPE-VMD-BiLSTM-TRC method. This experiment is also conducted on the first four representative test segments. The overall prediction curves, local zoom-in plots, and region-wise error metrics are jointly used to analyze the improvement effects of the composite loss and residual correction mechanism on switching-neighborhood error concentration, response lag, and post-switching bias persistence.

In the additional validation experiment on the held-out operating-condition segment, one test segment that is not included in the main result presentation is selected as the supplementary validation object. The participating models are consistent with those in the overall comparison experiment on the primary dataset, including LSTM, BiLSTM, TPE-BiLSTM, VMD-BiLSTM, VMD-SSA-BiLSTM, VMD-IPOA-BiLSTM, TPE-VMD-BiLSTM, and TPE-VMD-BiLSTM-TRC. This experiment adopts the same evaluation protocol as the representative test segments, and reports the overall prediction curve, local zoom-in plots, and RMSE and MAE over all samples, switching-neighborhood samples, and steady-state samples. This experiment further examines whether the performance ranking of the models remains consistent across different operating-condition segments, and whether the effect of the proposed method in switching neighborhoods is stable.

In summary, the experimental settings are designed around the objective of enhancing dynamic response under operating-condition switching. The overall comparison on the primary dataset evaluates the comprehensive advantage of the proposed method over basic

models, decomposition-based models, and models with different optimization strategies. The mechanism ablation analyzes the specific effects of the switching-aware composite loss and residual correction branch. The additional validation on the held-out operating-condition segment examines the prediction stability of the model across different operating-condition segments. Through the combination of these three types of experiments, the effectiveness of the proposed method is verified from the perspectives of overall prediction accuracy, local dynamic response capability, and operating-condition segment stability.

3.3. Evaluation Metrics and Region-Wise Evaluation Principle

This paper adopts root mean square error (RMSE) and mean absolute error (MAE) as the basic evaluation metrics [33]. RMSE is more sensitive to large errors and is suitable for reflecting possible local deviations near operating-condition switching. MAE directly reflects the average error level and can characterize the average prediction deviation over all samples. Let the evaluation set contain N_{eval} effective prediction samples, and let the actual and predicted values be y_i and $\hat{y}_i$, respectively. RMSE and MAE are defined as

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{eval}}} \sum_{i=1}^{N_{\text{eval}}} (y_i - \hat{y}_i)^2}. \quad (52)$$

$$\text{MAE} = \frac{1}{N_{\text{eval}}} \sum_{i=1}^{N_{\text{eval}}} |y_i - \hat{y}_i|. \quad (53)$$

To provide a more comprehensive evaluation, three additional metrics are introduced. MAPE is used to evaluate the relative prediction error,

$$\text{MAPE} = \frac{100\%}{N_{\text{eval}}} \sum_{i=1}^{N_{\text{eval}}} \left| \frac{y_i - \hat{y}_i}{|y_i| + \epsilon} \right|. \quad (54)$$

Here, ϵ is a small positive constant used to avoid numerical instability when the denominator approaches zero. For each switching event, the peak-amplitude error is calculated as the absolute difference between the predicted and actual peak amplitudes within the corresponding switching neighborhood, and the mean value over all switching events is reported. In addition, the switching-neighborhood first-difference RMSE is calculated from adjacent prediction increments and is used to evaluate the tracking ability for local rate variations. These additional metrics complement RMSE and MAE by reflecting relative accuracy, peak preservation, and dynamic tracking capability.

Since this study focuses not only on overall fitting accuracy, but also on the local dynamic response capability in switching neighborhoods, a region-wise evaluation strategy is further adopted. Within each operating-condition segment, based on the effective prediction target points obtained after non-cross-segment window construction, the first r effective prediction samples at the beginning of the segment are defined as the switching neighborhood Ω_{tr} , while the remaining samples are defined as the approximately steady-state region Ω_{st} . In this paper, r is set to 5 by default. This partitioning strategy is consistent with the definition of switching neighborhoods in Section 2, ensuring that switching-region error statistics are calculated based on effective prediction samples after non-cross-segment window construction.

On this basis, RMSE and MAE are calculated over all samples, switching-neighborhood samples, and steady-state samples, respectively. The overall metrics are used to evaluate the comprehensive prediction accuracy of the model on multi-condition test sequences. The switching-neighborhood metrics are used to assess the model's ability to track local peaks, rapid transitions, and turning positions. The steady-state metrics are used to

measure the average fitting level in relatively stable regions. This region-wise evaluation reduces the dilution of switching-region errors caused by the larger number of steady-state samples, allowing local error characteristics near operating-condition switching to be more fully reflected.

For the overall comparison experiment and the mechanism ablation experiment on the primary dataset, the first four representative test segments are used as the main statistical objects. First, the effective prediction samples of each operating-condition segment are uniformly aligned, and the overall RMSE and MAE are calculated on the concatenated sample set. Then, according to Ω_{tr} and Ω_{st} , the region-wise errors of the switching neighborhood and the steady-state region are calculated separately. Meanwhile, the overall prediction curves and fixed local zoom-in windows are used as auxiliary analyses to observe the response differences of different models at switching starting points, local peaks, and post-switching plateau stages.

For the additional validation experiment on the held-out operating-condition segment, the same evaluation protocol as that for the representative test segments is adopted. Specifically, RMSE and MAE are calculated over all samples, switching-neighborhood samples, and steady-state samples on this segment. Local zoom-in plots are also used to analyze the response details of the model in typical switching intervals. This treatment ensures a consistent evaluation criterion between the representative test segments and the held-out segment, which is beneficial for comparing the prediction stability of models across different operating-condition segments.

In addition to quantitative error metrics, this paper also analyzes the results using overall prediction curves and local zoom-in plots. The overall curves are used to observe the global tracking ability of the model for segmented variation trends, while the local zoom-in plots are used to analyze response lag, peak attenuation, and local bias persistence in switching regions. By combining numerical metrics with graphical analysis, the differences among models in overall prediction accuracy and local dynamic response capability can be evaluated more accurately.

In summary, the evaluation system consists of overall errors, region-wise errors, and local visual analysis. These metrics correspond to the three issues addressed in this paper, namely local peaks, abrupt rate variations, and inter-segment distribution discrepancies, and provide a unified basis for the subsequent overall comparison, mechanism ablation, and held-out segment validation.

3.4. Model Implementation and Parameter Settings

To ensure the comparability of experimental results, the data partitioning strategy, input variables, preprocessing rules, training procedure, and evaluation metrics are kept consistent across different models. All models use operating-condition segments as the basic units for development–test partitioning, and both the sliding-window input and the prediction target are restricted within the same segment. For models with an automatic optimization mechanism, parameter search is performed only within the development set, and the independent test set is not involved in parameter selection. TPE, SSA, and IPOA are compared under the same data partition, search space, training budget, and validation criterion; their differences lie in their candidate-generation and update mechanisms. For models without automatic optimization, parameters are configured under the same candidate ranges and training rules. Considering that different model structures may require different historical dependency lengths, this paper does not require all models to use the same fixed sliding-window length. Instead, each model is allowed to select a window length that matches its structure within a unified candidate interval.

According to the proposed method and experimental task, the main parameter ranges for segment-wise VMD decomposition, BiLSTM backbone modeling, and model training are listed in Table 2. For VMD, the number of modes K and the penalty factor α jointly determine the decomposition granularity and narrow-band constraint strength. For BiLSTM, the sliding-window length L , number of hidden units n_h , number of network layers n_l , and dropout rate d affect the temporal modeling capability and capacity of the backbone network. For the training process, the learning rate η and batch size b are directly related to training stability and convergence efficiency. In addition, the switching-neighborhood radius r is used to uniformly define switching and steady-state samples. The normalization parameters of continuous variables are fitted only on the inner-training subset and directly reused for inner validation, independent testing, and held-out operating-condition evaluation.

Table 2. Key implementation settings and parameter ranges.

Module	Parameter	Range or Setting
VMD	Number of modes K	4–8
VMD	Penalty factor α	500–3500
BiLSTM	Window length L	16–48
BiLSTM	Hidden units n_h	32–128
BiLSTM	Number of layers n_l	1–2
BiLSTM	Dropout rate d	0.0–0.4
Training	Learning rate η	$10^{-4} \sim 5 \times 10^{-3}$
Training	Batch size b	16, 32, 64
Region-wise eval.	Switching radius r	5
Data processing	Normalization	Fitted only on the inner-training subset

In terms of implementation, this paper adopts a one-step prediction setting and conducts three types of experiments under a unified data organization and evaluation protocol: overall comparison on the primary dataset, mechanism ablation, and additional validation on a held-out operating-condition segment. These experiments use consistent input variables, preprocessing rules, a segment-wise decomposition strategy, parameter search ranges, and evaluation metrics. Therefore, the performance differences among models mainly come from the model structures and improvement mechanisms themselves, rather than from differences in data processing or experimental settings.

Overall, the parameter settings in Table 2 correspond to the main modules of the proposed method. The VMD parameters control the multi-scale decomposition results, the BiLSTM parameters determine the temporal modeling capability of the backbone predictor, and the training parameters support stable optimization. The switching-neighborhood radius provides the basis for region-wise evaluation and composite loss construction, while the normalization strategy and non-cross-segment window construction maintain consistent data boundaries among the training set, test set, and held-out operating-condition segment.

4. Experimental Results and Analysis

4.1. Overall Comparison Results on the Primary Dataset

To verify the overall effectiveness of the proposed TPE-VMD-BiLSTM-TRC method for multi-condition active current prediction of synchronous reluctance motors, an overall comparison experiment is first conducted on four representative test segments of the primary dataset. The comparison models include LSTM, BiLSTM, TPE-BiLSTM, VMD-BiLSTM, VMD-SSA-BiLSTM, VMD-IPOA-BiLSTM, TPE-VMD-BiLSTM, and the proposed method. Figure 3 shows the prediction curves of the four representative test segments

and their local zoom-in windows, where the segment-wise display is used to highlight the switching-neighborhood responses under different operating-condition patterns.

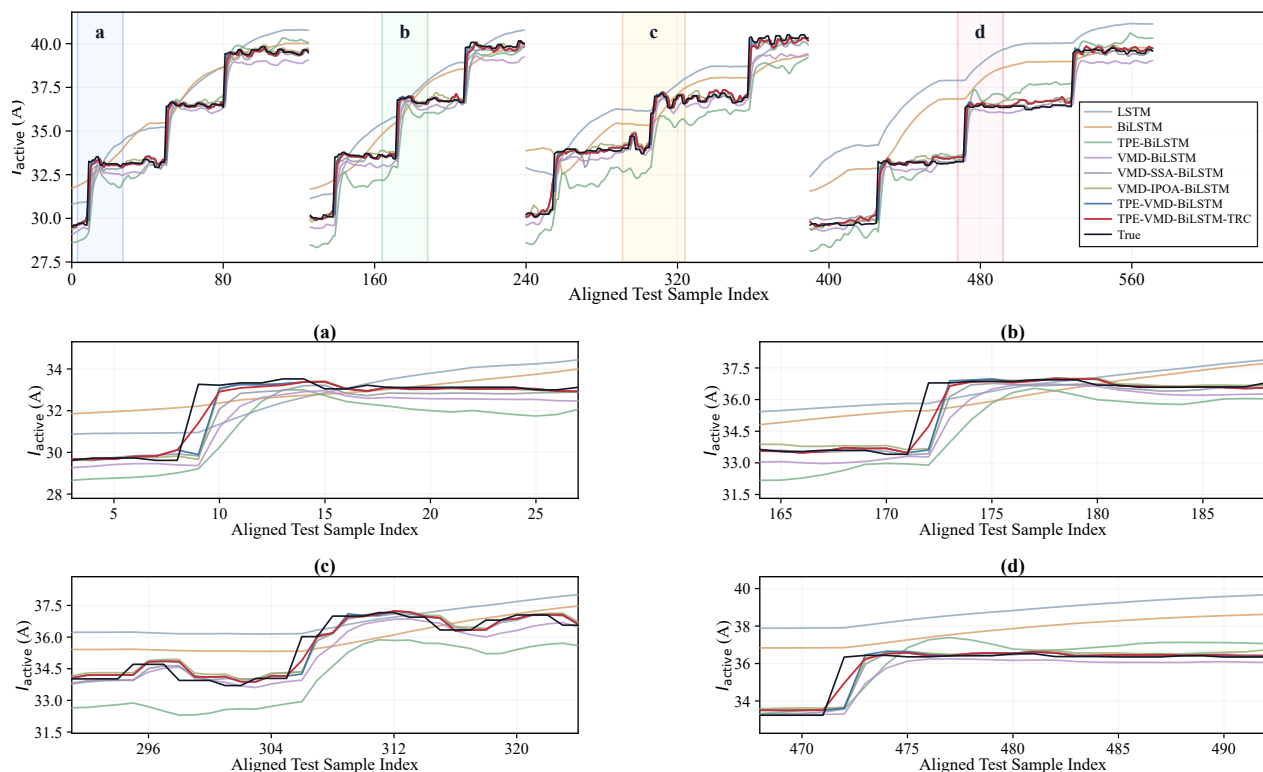


Figure 3. Overall prediction curves and local zoom-in results on the primary test segments: (a) segment 1; (b) segment 2; (c) segment 3; (d) segment 4; lower panels show the corresponding local zoom-in windows.

Table 3 reports the error statistics over all samples, switching-neighborhood samples, and steady-state samples. The overall errors evaluate the average fitting ability over multi-condition sequences, the switching-neighborhood errors reflect the response to local peaks and abrupt rate variations, and the steady-state errors indicate whether the local enhancement mechanism affects prediction stability in relatively stable regions.

Table 3. Overall comparison results on the primary dataset.

Model	Overall RMSE	Overall MAE	Switching RMSE	Switching MAE	Steady-State RMSE	Steady-State MAE
LSTM	2.3390	1.9205	1.5027	1.2739	2.4367	2.0145
BiLSTM	1.6872	1.3986	1.4577	1.2072	1.7179	1.4264
TPE-BiLSTM	1.0695	0.8301	1.9300	1.6390	0.8768	0.7126
VMD-BiLSTM	0.7127	0.4976	1.4963	1.1174	0.5064	0.4076
VMD-SSA-BiLSTM	0.4737	0.2484	1.1627	0.6758	0.2460	0.1864
VMD-IPOA-BiLSTM	0.4555	0.2258	1.1472	0.5802	0.2154	0.1743
TPE-VMD-BiLSTM	0.4228	0.1937	1.0693	0.5074	0.1965	0.1481
TPE-VMD-BiLSTM-TRC	0.2965	0.1792	0.6542	0.3934	0.1965	0.1481

As shown in Table 3, the prediction error decreases progressively from the basic recurrent models to the decomposition- and optimization-based models. LSTM gives the largest overall error, while BiLSTM reduces the overall RMSE from 2.3390 to 1.6872 by using bidirectional temporal modeling. TPE-BiLSTM further reduces the overall RMSE to 1.0695, indicating that hyperparameter configuration has a clear influence on the fitting ability of the basic temporal network. However, its switching RMSE rises to 1.9300, showing that

hyperparameter optimization alone mainly improves global and steady-state fitting but does not sufficiently constrain rapid transition samples.

After VMD is introduced, VMD-BiLSTM reduces the overall RMSE to 0.7127, indicating that multi-scale decomposition alleviates the nonstationarity of the original active current sequence. With SSA, IPOA, and TPE optimization, the VMD-BiLSTM variants further improve the overall and switching-neighborhood metrics. Among them, TPE-VMD-BiLSTM obtains the best backbone performance, with an overall RMSE of 0.4228 and a switching RMSE of 1.0693, suggesting that TPE better coordinates the VMD decomposition scale and BiLSTM hyperparameters.

The proposed TPE-VMD-BiLSTM-TRC achieves the best overall and switching-neighborhood results. Its overall RMSE and MAE are 0.2965 and 0.1792, respectively, and its switching RMSE and MAE decrease to 0.6542 and 0.3934. Compared with TPE-VMD-BiLSTM, the overall RMSE and switching RMSE are reduced by 29.87% and 38.82%, respectively, while the steady-state RMSE and MAE remain unchanged at 0.1965 and 0.1481. This indicates that the gain of the proposed method mainly comes from targeted suppression of switching-neighborhood errors, abrupt rate variations, and structural residuals, rather than from repeated optimization of steady-state samples.

Figure 3 supports these quantitative results. All models can follow the segmented stepwise trend to some extent, but LSTM and BiLSTM show clear response lag and over-smoothing near switching positions. TPE-BiLSTM improves steady-state fitting but still underestimates several transition processes. VMD-based models shorten the transition lag and improve local fitting, while TPE-VMD-BiLSTM-TRC gives the closest agreement with the actual curve after repeated step changes.

Panels (a) and (b) mainly show typical step-switching responses. The basic recurrent models transition more slowly than the actual current, whereas VMD-based models approach the new current level more rapidly. The proposed method provides a more accurate rising position and a smaller post-switching platform bias. Panels (c) and (d) contain stronger local fluctuations and higher-current switching processes. In these windows, the proposed method better follows local variation trends, transition slopes, and post-switching plateaus, demonstrating stronger local dynamic tracking under more challenging operating patterns.

Overall, TPE-VMD-BiLSTM-TRC achieves the most balanced performance in overall fitting and switching-neighborhood response. Its advantage is reflected not only in lower average error, but also in improved dynamic tracking, rapid-transition tracking, and post-switching bias suppression.

To further evaluate relative accuracy, peak preservation, and dynamic tracking, Table 4 reports three additional indicators for all comparison models in the overall experiment. The proposed method achieves the lowest MAPE and switching-neighborhood first-difference RMSE, indicating better overall relative accuracy and stronger tracking of local rate variations. Its event-level peak-amplitude error is not the lowest among all compared models, suggesting that the current TRC module mainly improves the complete transient trajectory and dynamic trend rather than strictly preserving the maximum amplitude of every switching event.

4.2. Mechanism Ablation Results

To further analyze the contribution of each key component in the proposed method, four ablation models are constructed around the same TPE-VMD-BiLSTM backbone. They are defined as A_0 = TPE-VMD-BiLSTM, A_1 = TPE-VMD-BiLSTM with switching-aware composite loss, A_2 = TPE-VMD-BiLSTM with residual correction branch, and A_3 = TPE-VMD-BiLSTM-TRC under the ablation setting. This experiment focuses on two questions.

The first is whether the first-difference error and switching-neighborhood weighted error can improve the prediction constraint on abrupt rate variations and key transition samples. The second is whether the residual correction branch can compensate for the structural bias remaining in the backbone model within switching neighborhoods. Figure 4 shows the overall prediction curves and local zoom-in results of the four ablation models, and Table 5 reports the corresponding overall and region-wise error statistics. The ablation metrics are used for within-ablation comparison and are not intended for item-by-item comparison with Table 3.

Table 4. Additional evaluation metrics on the primary dataset.

Model	MAPE (%)	Switch Diff. RMSE	Peak Amp. Error
TPE-VMD-BiLSTM-TRC	0.5060	0.8407	0.1487
TPE-VMD-BiLSTM	0.5463	1.4728	0.1038
VMD-IPOA-BiLSTM	0.6384	1.4713	0.1362
VMD-SSA-BiLSTM	0.7017	1.3655	0.2114
VMD-BiLSTM	1.3607	1.2974	0.4779
TPE-BiLSTM	2.3751	1.2660	0.8568
BiLSTM	4.1045	1.1578	1.1356
LSTM	5.6695	1.1563	1.2282

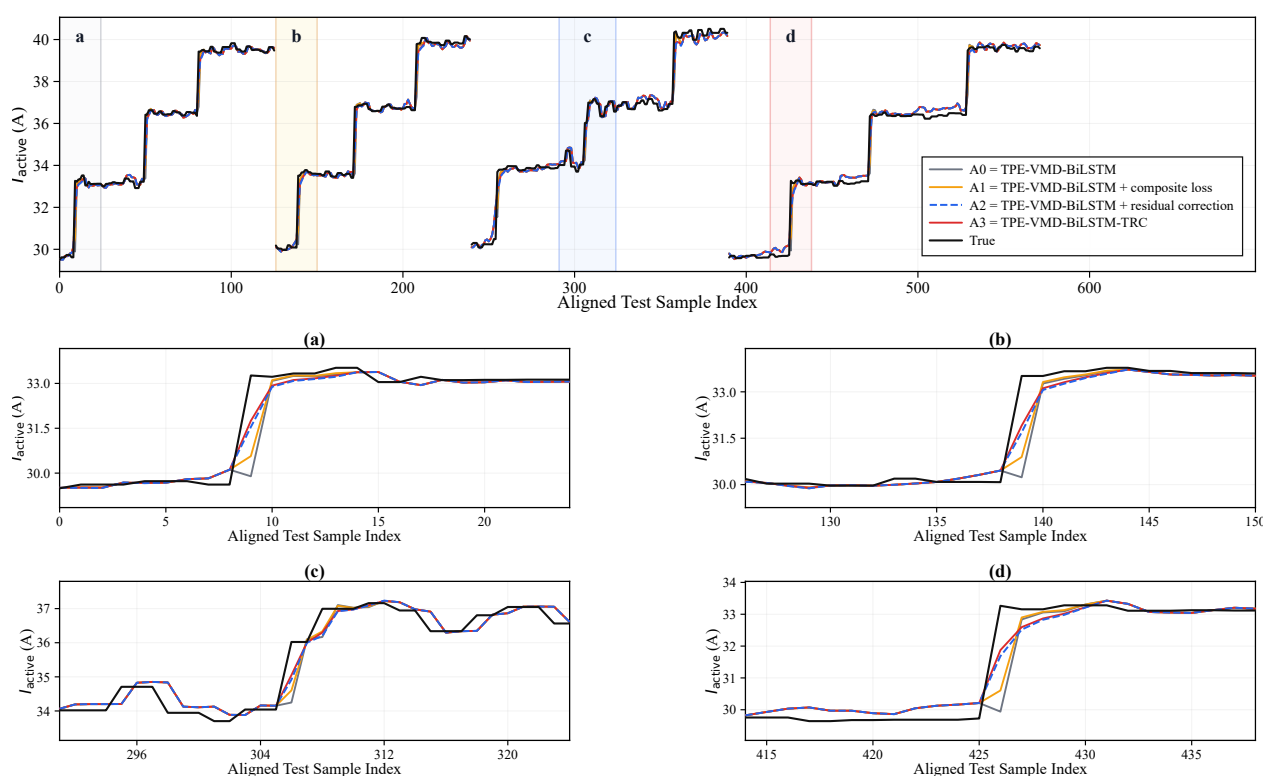


Figure 4. Overall prediction curves and local zoom-in results of the ablation models on the primary test segments: (a) segment 1; (b) segment 2; (c) segment 3; (d) segment 4; lower panels show the corresponding local zoom-in windows.

Table 5 shows that both enhancement mechanisms improve the same TPE-VMD-BiLSTM backbone. Compared with the baseline A_0 , adding the switching-aware composite loss in A_1 reduces the overall RMSE from 0.4227 to 0.3557. Adding only the residual correction branch in A_2 further reduces it to 0.2954. When both mechanisms are used, the complete model A_3 obtains the lowest overall RMSE and MAE of 0.2742 and 0.1737, confirming that the two modules provide complementary gains.

Table 5. Mechanism ablation results.

Model	Description	Overall RMSE	Overall MAE	Switching RMSE	Switching MAE	Steady-State RMSE	Steady-State MAE
A_0	TPE-VMD-BiLSTM	0.4227	0.1937	1.0693	0.5074	0.1965	0.1481
A_1	A_0 + composite loss	0.3557	0.1808	0.8554	0.4059	0.1965	0.1481
A_2	A_0 + residual correction	0.2954	0.1797	0.6497	0.3973	0.1965	0.1481
A_3	TPE-VMD-BiLSTM-TRC under ablation setting	0.2742	0.1737	0.5718	0.3496	0.1965	0.1481

The region-wise metrics reveal the main source of improvement. The switching RMSE decreases from 1.0693 in A_0 to 0.8554 in A_1 , indicating that the first-difference constraint and switching-neighborhood weighting strengthen training emphasis on rapidly varying samples. With the residual correction branch, A_2 further reduces the switching RMSE to 0.6497, suggesting that the backbone output still contains learnable systematic residuals near switching positions. The complete model A_3 achieves the lowest switching RMSE and MAE of 0.5718 and 0.3496. Meanwhile, all four models have identical steady-state RMSE and MAE values of 0.1965 and 0.1481, showing that the improvement is concentrated in switching neighborhoods rather than in repeated steady-state fitting.

Figure 4 provides the corresponding visual evidence. The backbone model A_0 already follows the main trend, but it retains response lag and local residual bias near several switching positions. The composite loss improves trend tracking near switching onsets, while the residual correction branch further suppresses post-switching persistent bias. In Panels (a) and (b), the complete model gives a more balanced prediction during both the rising process and the post-switching platform stage. In Panels (c) and (d), which include local fluctuations and short transient peaks, A_3 better follows the local variation trend, transition slope, and plateau stability.

Overall, the ablation results demonstrate that the composite loss and residual correction branch act at different levels. The former strengthens variation-rate consistency and key transition samples in the training objective, whereas the latter learns a compensation term between the backbone output and the actual value. Their combination improves switching-neighborhood prediction without sacrificing steady-state fitting ability.

4.3. Additional Validation Results on a Held-Out Operating-Condition Segment

To examine the prediction stability of the model across different operating-condition segments, a test segment that was not used in the main result presentation is selected as an additional validation object. This held-out segment differs from the four representative test segments in terms of local variation pattern and current level. Figure 5 shows the overall prediction curve and the local zoom-in result on this segment, while Table 6 reports the corresponding error statistics over all samples, switching-neighborhood samples, and steady-state samples.

It should be noted that this experiment is a supplementary segment-level validation within the same primary dataset, rather than an external validation on an independent dataset. Its purpose is to examine whether the improvement observed on the representative test segments can still be maintained on another operating-condition segment with different current levels and switching characteristics. Therefore, the same preprocessing rules, normalization parameters, non-cross-segment window construction strategy, and region-wise evaluation criteria are used.

In Figure 5, the horizontal axis keeps the globally aligned test sample index to preserve the original position of this segment in the aligned test sequence. The local zoom-in window

further highlights the response near a typical switching position, where prediction lag, peak attenuation, and post-switching platform bias are more likely to occur.

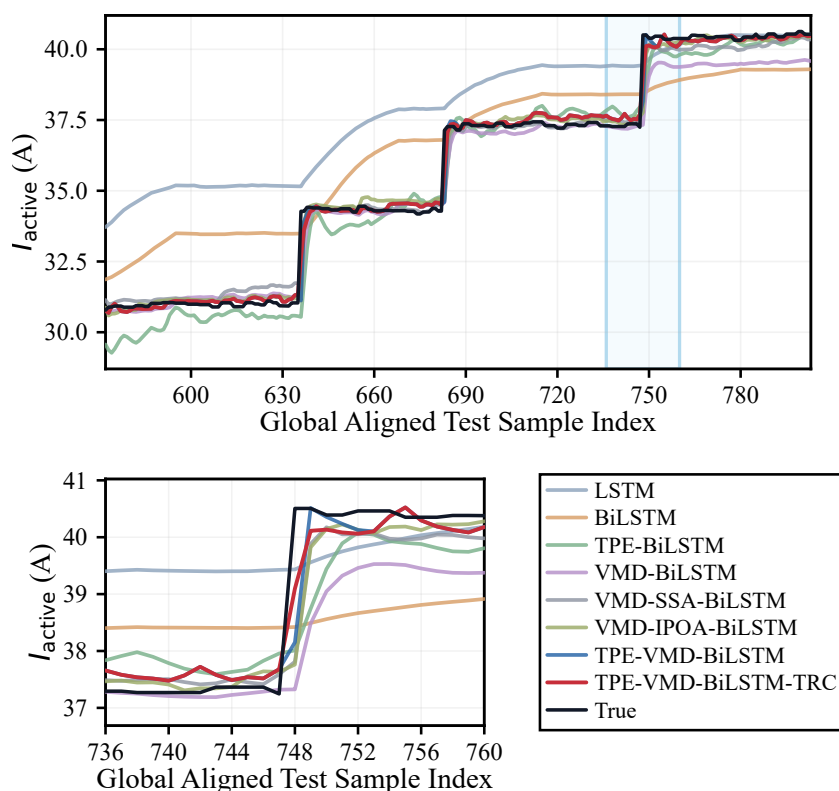


Figure 5. Prediction results on the held-out operating-condition segment: full-segment prediction curve and local zoom-in result near a representative switching position.

Table 6. Additional validation results on the held-out operating-condition segment.

Model	Overall RMSE	Overall MAE	Switching RMSE	Switching MAE	Steady-State RMSE	Steady-State MAE
LSTM	2.6822	2.2996	1.7252	1.4586	2.7554	2.3027
BiLSTM	1.6639	1.4969	1.1487	0.9627	1.7045	1.5473
TPE-BiLSTM	0.6667	0.4980	1.6082	1.3286	0.4924	0.4196
VMD-BiLSTM	0.6315	0.4220	1.3845	0.9723	0.5056	0.3701
VMD-SSA-BiLSTM	0.4074	0.2479	1.0473	0.5980	0.2795	0.2149
VMD-IPOA-BiLSTM	0.3812	0.2107	1.1772	0.6151	0.2030	0.1725
TPE-VMD-BiLSTM	0.3548	0.1781	1.0574	0.5089	0.1797	0.1469
TPE-VMD-BiLSTM-TRC	0.2525	0.1689	0.6302	0.4014	0.1797	0.1469

As shown in Table 6, the performance ranking on the held-out operating-condition segment is generally consistent with the ranking on the representative test segments. The basic recurrent models produce relatively large errors, while the decomposition- and optimization-based models achieve lower RMSE and MAE values. Among all models, TPE-VMD-BiLSTM-TRC obtains the best overall performance, with an overall RMSE of 0.2525 and an MAE of 0.1689. Compared with TPE-VMD-BiLSTM, the overall RMSE is reduced by 28.84% and the overall MAE by 5.17%, indicating that the proposed correction mechanism remains effective on another operating-condition segment.

The region-wise results further show that the main improvement is concentrated in the switching neighborhood. Although TPE-BiLSTM substantially reduces the overall and steady-state errors compared with BiLSTM, its switching RMSE and MAE increase to 1.6082 and 1.3286, respectively. This indicates that automatic parameter optimization of the basic

BiLSTM improves steady-state fitting but does not guarantee better prediction in switching neighborhoods. In contrast, TPE-VMD-BiLSTM-TRC achieves a switching RMSE of 0.6302 and a switching MAE of 0.4014, whereas the corresponding values of TPE-VMD-BiLSTM are 1.0574 and 0.5089. The reductions are therefore 40.40% and 21.13%, respectively.

Meanwhile, the steady-state RMSE and MAE of TPE-VMD-BiLSTM-TRC and TPE-VMD-BiLSTM remain identical at 0.1797 and 0.1469. This indicates that the performance gain on the held-out segment does not come from repeated optimization of steady-state samples, but mainly from the suppression of local errors near switching positions. The result is consistent with the observations on the primary test segments and further supports the switching-neighborhood-oriented design of the proposed method.

The prediction curves in Figure 5 support the quantitative results. The horizontal axis retains the globally aligned test sample index, so the held-out segment keeps its original position in the aligned test sequence. In the local zoom-in window, the actual current rapidly rises from a lower platform to a higher platform and then remains near a new operating level. The basic recurrent models show visible response lag, and several decomposition-based models still exhibit deviations in the rising position or plateau stage. In contrast, TPE-VMD-BiLSTM-TRC provides the closest fit during the switching onset, transition process, and post-switching platform, demonstrating stronger local dynamic response on the held-out segment.

Overall, the additional validation confirms that the proposed method maintains stable error reduction on a segment not used in the main result presentation. These results indicate that segment-level data organization, segment-wise VMD decomposition, TPE-based joint optimization, and switching-aware residual correction have consistent effects across different operating-condition segments. However, this validation is still conducted within the same primary dataset, and external generalization under different experimental platforms, sampling conditions, or noise environments remains to be further examined.

4.4. Sensitivity and Complexity Analysis

To examine whether the conclusion depends on the selected switching-neighborhood radius, the trained models and their predictions are fixed, and the switching neighborhoods are recalculated using representative radii $r = 3, 5, 7, 9$. This analysis changes only the evaluation-stage neighborhood partition and does not involve model retraining. As shown in Table 7, the proposed method maintains the lowest switching-region RMSE under all tested radii, indicating that its relative advantage is not dependent on the specific choice of $r = 5$.

Table 7. Switching-neighborhood RMSE under different evaluation radii.

Model	$r = 3$	$r = 5$	$r = 7$	$r = 9$
VMD-IPOA-BiLSTM	1.4645	1.1472	0.9780	0.8676
TPE-VMD-BiLSTM	1.3664	1.0693	0.9098	0.8083
TPE-VMD-BiLSTM-TRC	0.8255	0.6542	0.5602	0.5024

The computational complexity of the main VMD-BiLSTM-based models is summarized in Table 8. The parameter storage is estimated from trainable parameters and is used for relative comparison. The proposed TRC branch is a lightweight correction module rather than a large global prediction model. Compared with TPE-VMD-BiLSTM, TPE-VMD-BiLSTM-TRC increases the parameter count from 185,478 to 188,263 and the inference latency from 0.3223 to 0.3345 ms/sample, while reducing the overall RMSE from 0.4228 to 0.2965. This indicates that the proposed switching-aware correction provides a favorable accuracy–complexity trade-off.

Table 8. Complexity comparison of representative models.

Model	Params	Parameter Storage (MB)	Latency (ms/Sample)	Overall RMSE
VMD-BiLSTM	43,395	0.1773	0.1062	0.7127
VMD-SSA-BiLSTM	267,525	1.0398	0.3420	0.4737
VMD-IPOA-BiLSTM	72,325	0.2951	0.2437	0.4555
TPE-VMD-BiLSTM	185,478	0.7305	0.3223	0.4228
TPE-VMD-BiLSTM-TRC	188,263	0.7424	0.3345	0.2965

Table 9 further compares the search efficiency of TPE, SSA, and IPOA. The three optimizers use the same data partition, search space, training budget, and validation criterion. The actual number of objective-function evaluations may differ because the optimizers have different candidate-generation and update mechanisms. TPE completes fewer objective-function evaluations than SSA and IPOA, although its wall-clock search time is not the shortest under the current CPU implementation. Therefore, the advantage of TPE is described as fewer candidate evaluations and better final prediction performance, rather than the shortest absolute running time.

Table 9. Search efficiency of different optimizers.

Optimizer	Search Mechanism	Objective Evaluations	Search Time (s)
TPE	Sequential probabilistic model	18	1254.50
SSA	Population-based search	26	1076.36
IPOA	Population-based search	27	631.15

4.5. Discussion

The overall comparison on the primary dataset, mechanism ablation, and additional validation on the held-out operating-condition segment indicate that the performance improvement of the proposed method mainly comes from the targeted reduction of errors in switching neighborhoods, rather than from repeated optimization of steady-state regions. On the representative test segments, TPE-VMD-BiLSTM-TRC reduces the overall RMSE and switching RMSE by 29.87% and 38.82%, respectively, compared with TPE-VMD-BiLSTM, while the steady-state RMSE and MAE remain unchanged. This shows that the switching-aware composite loss and the TRC residual correction branch mainly improve the local rapidly varying regions without significantly changing the average fitting level in steady-state intervals.

From the perspective of error sources, the basic LSTM and BiLSTM models tend to show response lag near switching onsets, indicating that a pure recurrent temporal structure is insufficient to fully capture the rapid transition process after operating-condition switching. VMD decomposition and parameter optimization can reduce the overall error, but some models still retain noticeable deviations around local peaks and turning regions. The proposed method constrains the local variation trend of the predicted sequence through the first-difference error, and increases the training weight of key transition samples through the switching-neighborhood weighted error. As a result, the model obtains stronger dynamic tracking capability at turning positions. The residual correction branch further compensates for systematic residuals in the backbone output, which leads to more stable transient tracking and plateau maintenance in the local zoom-in windows.

The mechanism ablation results further show that the composite loss and the residual correction branch have different functional emphases. The composite loss strengthens variation-trend constraints and switching-sample weights from the training-objective level, whereas the residual correction branch performs secondary correction for the remaining

bias of the backbone model from the output-structure level. When the two mechanisms are combined, the model achieves a more significant improvement in switching-neighborhood metrics, while the steady-state errors remain stable. This indicates that the proposed TRC mechanism does not improve dynamic response at the expense of fitting ability in stable regions.

The additional validation on the held-out operating-condition segment shows that the proposed method still maintains lower overall and switching-neighborhood errors on a segment not used in the main result presentation. These results suggest that segment-level data organization, segment-wise VMD decomposition, TPE-based joint optimization, and switching-aware residual correction have consistent effects across different operating-condition segments. However, this validation is still conducted within the same primary dataset and cannot fully represent external generalization ability under different experimental platforms, sampling conditions, or complex noise environments. Future work can further introduce independent experimental datasets and conduct sensitivity analysis on the switching-neighborhood radius, composite loss weights, and residual branch structure.

5. Conclusions and Future Work

5.1. Conclusions

To address switching-neighborhood error concentration, turning-point lag, and the influence of inter-segment distribution discrepancies in short-term active current prediction of synchronous reluctance motors under multiple operating conditions such as frequency steps, load transients, and alternating forward-reverse rotation, this paper proposes a multi-condition prediction method based on TPE-VMD-BiLSTM-TRC. The proposed method constructs non-cross-segment one-step prediction samples using operating-condition segments as the basic units. Within each segment, VMD is used to decompose the active current sequence, and BiLSTM is employed for mode-wise prediction and reconstruction. On this basis, a switching-aware composite loss and a lightweight residual correction branch are introduced to constrain and compensate for errors in switching neighborhoods. Meanwhile, TPE is adopted to jointly search the VMD parameters and BiLSTM hyperparameters.

The experimental results show that VMD decomposition can reduce the difficulty of modeling the nonstationary active current sequence, while TPE-based joint optimization improves the matching between the decomposition scale and network capacity. On the representative test segments, the proposed method reduces the overall RMSE and switching RMSE by 29.87% and 38.82%, respectively, compared with TPE-VMD-BiLSTM. The MAPE is reduced from 0.5463% to 0.5060%, corresponding to a relative reduction of approximately 7.4%. The switching-neighborhood first-difference RMSE is also reduced from 1.4728 to 0.8407, indicating improved tracking of local rate variations. This indicates that the switching-aware composite loss and the TRC residual correction branch can effectively alleviate local error concentration in switching neighborhoods.

The region-wise evaluation and mechanism ablation results further show that a reduction in overall error does not necessarily imply a simultaneous improvement in switching-neighborhood response. The main gains of the proposed method are concentrated in switching samples, while the steady-state errors remain nearly unchanged. This demonstrates that the proposed mechanisms mainly act on switching-neighborhood errors, rapid transitions, and post-switching structural bias. Additional validation on a held-out operating-condition segment shows that the proposed method still maintains stable error reduction on a segment not used in the main result presentation, with the overall RMSE reduced by approximately 28.84% compared with TPE-VMD-BiLSTM. Meanwhile, the event-level peak-amplitude error is not the best among all compared methods, suggesting that peak preservation for individual switching events remains a direction for further improvement.

5.2. Future Work

Several aspects of this study can be further improved. The current experiments are mainly based on one primary dataset and its held-out operating-condition segment. In future work, independent datasets collected from different experimental platforms, sampling conditions, and noise environments can be introduced to further examine the external generalization ability of the model. In addition, training-stage sensitivity to the switching-neighborhood radius and loss weights, segment-adaptive VMD parameter selection, and more flexible residual correction structures can be further investigated to improve robustness and applicability in practical electric drive systems.

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

Funding: This research was funded by the Natural Science Foundation of Shandong Province under Grant ZR2023ME079, and in part by the Shandong Provincial Key Research and Development Program (Competitive Innovation Platform) under Grant 2023CXPT089.

Data Availability Statement: The experimental data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT based on GPT-5 (OpenAI, San Francisco, CA, USA) to assist with language polishing and translation from Chinese to English. The authors reviewed and edited the output as needed and take full responsibility for the content of the manuscript.

Conflicts of Interest: The authors confirm that the Gree Electric Appliances, Inc. of Zhuhai affiliation corresponds to Fang Zhang’s author affiliation and does not constitute a conflict of interest for this study. The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SynRM	Synchronous reluctance motor
VMD	Variational mode decomposition
BiLSTM	Bidirectional long short-term memory
TPE	Tree-structured Parzen estimator
TRC	Switching-aware residual correction
RMSE	Root mean square error
MAE	Mean absolute error
LSTM	Long short-term memory
SSA	Sparrow search algorithm
IPOA	Improved pelican optimization algorithm
MAPE	Mean absolute percentage error
K	Number of VMD modes
α	VMD penalty factor
r	Switching-neighborhood radius
$\lambda_1, \lambda_2, \lambda_3$	Loss-weight coefficients
dir_t	Rotational direction feature

References

- Fadini, G.; Kumar, S.; Kumar, R.; Flayols, T.; Del Prete, A.; Carpentier, J.; Souères, P. Co-designing versatile quadruped robots for dynamic and energy-efficient motions. *Robotica* **2024**, *42*, 2004–2025. [\[CrossRef\]](#)
- Wang, Z.; Yu, F.; Zhang, J.; Qin, X. Low current ripple parameter-free predictive current control of three-level inverter-fed PMSMs with current variations correction. *J. Power Electron.* **2025**, *25*, 1526–1538. [\[CrossRef\]](#)
- Suli, S.; Oner, Y.; Senol, I. Design of SynRM and PMaSynRM using a common industrial stator platform: A new design methodology for transition from industrial applications to electric vehicles. *Eng. Sci. Technol. Int. J.* **2025**, *71*, 102204. [\[CrossRef\]](#)
- Ning, Y.; Huang, Z.; Hou, H.; Li, R.; Sun, W.; He, J. Synchronous reluctance motor speed control based on multivariable Super-Twisting. *Sci. Rep.* **2025**, *15*, 44326. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cai, H.; Luo, W.J. Full-speed sensorless control system of synchronous reluctance motor with flux saturation model. *Sci. Rep.* **2025**, *15*, 9048. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kim, J.; Kim, H.; Kim, H.G.; Lee, D.; Yoon, S. A comprehensive survey of deep learning for time series forecasting: Architectural diversity and open challenges. *Artif. Intell. Rev.* **2025**, *58*, 216. [\[CrossRef\]](#)
- Li, C.; Meng, Q.; Shi, T. A review on model predictive control strategies for AC motor drives. *IET Electr. Power Appl.* **2024**, *18*, 1584–1604. [\[CrossRef\]](#)
- Kong, X.; Chen, Z.; Liu, W.; Ning, K.; Zhang, L.; Marier, S.M.; Liu, Y.; Chen, Y.; Xia, F. Deep learning for time series forecasting: A survey. *Int. J. Mach. Learn. Cybern.* **2025**, *16*, 5079–5112. [\[CrossRef\]](#)
- Teler, K.; Orłowska-Kowalska, T. Analysis of the stator current prediction capabilities in induction motor drive using the LSTM network. *Power Electron. Drives* **2023**, *8*, 31–52. [\[CrossRef\]](#)
- Le, T.T.H.; Oktian, Y.E.; Jo, U.; Kim, H. Time series electrical motor drives forecasting based on simulation modeling and bidirectional long-short term memory. *Sensors* **2023**, *23*, 7647. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mourad, N. Statistical variational mode decomposition. *J. Frankl. Inst.* **2026**, *363*, 108287. [\[CrossRef\]](#)
- Zhang, X.Q.; Wang, X.; Li, H.Y.; Sun, S.F.; Liu, F. Monthly runoff prediction based on a coupled VMD-SSA-BiLSTM model. *Sci. Rep.* **2023**, *13*, 13149. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhong, B.; Zhou, Y.; Liu, Y.; Zhao, X. Short-term power grid load forecasting based on VMD-SE-Attention-BiLSTM. *Int. J. Low-Carbon Technol.* **2024**, *19*, 2330–2339. [\[CrossRef\]](#)
- Yuan, P.; Zhao, W.; Zhao, Y. Intelligent prediction algorithm of ship roll and pitch motion based on SSA-optimized BiLSTM network. *Ocean Eng.* **2025**, *320*, 120331. [\[CrossRef\]](#)
- Kaur, S.; Bala, A.; Parashar, A. GA-BiLSTM: An intelligent energy prediction and optimization approach for individual home appliances. *Evol. Syst.* **2024**, *15*, 413–427. [\[CrossRef\]](#)
- Hanifi, S.; Cammarono, A.; Zare-Behtash, H. Advanced hyperparameter optimization of deep learning models for wind power prediction. *Renew. Energy* **2024**, *221*, 119700. [\[CrossRef\]](#)
- Yang, D.C.; Li, M.Z.; Guo, J.E.; Du, P. An attention-based multi-input LSTM with sliding window-based two-stage decomposition for wind speed forecasting. *Appl. Energy* **2024**, *375*, 124057. [\[CrossRef\]](#)
- Dai, S.; Meng, F.; Dai, H.; Wang, Q.; Chen, X.; Bai, W.; Shi, P.; Allmendinger, R.; Zhang, Y.; Liu, J. Machine learning in peak demand forecasting: Foundations, trends, and insights. *Renew. Sustain. Energy Rev.* **2026**, *227*, 116500. [\[CrossRef\]](#)
- Zhang, Y.D.; Wang, X.; Sun, Z.Y.; Wang, P.K.; Wang, B.W.; Li, L.M.; Wang, Y. Meta Koopman decomposition for time series forecasting under temporal distribution shifts. *Adv. Eng. Inform.* **2024**, *62*, 102840. [\[CrossRef\]](#)
- Song, X.; Deng, L.; Wang, H.; Zhang, Y.; He, Y.; Cao, W. Deep learning-based time series forecasting. *Artif. Intell. Rev.* **2025**, *58*, 23. [\[CrossRef\]](#)
- Yang, X.; Li, J.; Jiang, X. Research on information leakage in time series prediction based on empirical mode decomposition. *Sci. Rep.* **2024**, *14*, 28362. [\[CrossRef\]](#) [\[PubMed\]](#)
- Lopez, E.; Gorla, G.; Etxebarria, J.; Amigo, J.M.; Seifert, A. The importance of choosing a proper validation strategy in predictive models. Part 2: Recipes for (avoiding) overfitting—A tutorial. *Anal. Chim. Acta* **2026**, *1384*, 344838. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhou, H.; Schertzer, D.; Tchiguirinskaia, I. Combining recurrent neural networks with variational mode decomposition and multifractals to predict rainfall time series. *Hydrol. Earth Syst. Sci.* **2025**, *29*, 4437–4455. [\[CrossRef\]](#)
- Wang, Q.; Sun, Q.; Wang, J.; Zhang, L. Short-term wind speed prediction of wind farm based on TSO-VMD-BiLSTM. *PeerJ Comput. Sci.* **2024**, *10*, e2032. [\[CrossRef\]](#) [\[PubMed\]](#)
- Li, J.; Li, L.; Du, Q.; Li, Y. Research on short-term prediction method of photovoltaic power based on HPO-VMD-BiLSTM. *Sci. Rep.* **2026**, *16*, 13541. [\[CrossRef\]](#) [\[PubMed\]](#)
- Fan, Y.; Tang, Q.; Guo, Y.; Wei, Y. BiLSTM-MLAM: A multi-scale time series prediction model for sensor data based on Bi-LSTM and local attention mechanisms. *Sensors* **2024**, *24*, 3962. [\[CrossRef\]](#) [\[PubMed\]](#)
- Kim, D.H.; Moon, I.J.; Lim, C.; Woo, S.B. Improved prediction of extreme ENSO events using an artificial neural network with weighted loss functions. *Front. Mar. Sci.* **2024**, *10*, 1309609. [\[CrossRef\]](#)

28. Bommididi, B.S.; Teeparthi, K.; Kosana, V. Hybrid wind speed forecasting using ICEEMDAN and transformer model with novel loss function. *Energy* **2023**, *265*, 126383. [[CrossRef](#)]
29. Han, J.C.; Zeng, P. Residual BiLSTM based hybrid model for short-term load forecasting in buildings. *J. Build. Eng.* **2025**, *99*, 111593. [[CrossRef](#)]
30. Zhu, L.; Liu, W.; Tan, H.Z.; Hu, T. A novel Bayesian optimization prediction framework for four-axis industrial robot joint motion state. *Complex Intell. Syst.* **2024**, *10*, 4867–4881. [[CrossRef](#)]
31. Lopez, E.; Etxebarria-Elezgarai, J.; Amigo, J.M.; Seifert, A. The importance of choosing a proper validation strategy in predictive models: A tutorial with real examples. *Anal. Chim. Acta* **2023**, *1275*, 341532. [[CrossRef](#)] [[PubMed](#)]
32. Allgaier, J.; Pryss, R. Practical approaches in evaluating validation and biases of machine learning applied to mobile health studies. *Commun. Med.* **2024**, *4*, 76. [[CrossRef](#)] [[PubMed](#)]
33. Khoshvaght, H.; Permala, R.R.; Razmjou, A.; Khiadani, M. A critical review on selecting performance evaluation metrics for supervised machine learning models in wastewater quality prediction. *J. Environ. Chem. Eng.* **2025**, *13*, 119675. [[CrossRef](#)]

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.