Artificial Intelligence and Computing Innovations
Intelligence Science Publishing
https://www.ispublishing.com
ISSN: 3070-4944
Vo.6, No.1, 2026

Dynamic Residual Calibration and Multi-Scale Fusion for Accurate Prediction of Non-Stationary Backend Indicators

Ansel Whitaker[1*], Spencer Caldwell[1]

[1] *Wright State University, Dayton, USA*

**Corresponding Author: ansel.whitaker887@wright.edu*

Abstract: This paper proposes a residual calibration modeling and prediction method for non-stationary backend metrics to address the challenges of non-stationarity, dynamic dependency, and distribution drift in multi-dimensional metric sequences of cloud backend systems. The method employs a dual-layer collaborative architecture consisting of a primary predictor and a residual calibrator. The primary predictor captures the overall trend and temporal evolution of system metrics, while the residual calibrator performs structured modeling and adaptive correction of prediction errors to precisely compensate for sudden fluctuations and contextual variations. At the modeling level, a multi-scale feature fusion mechanism is introduced to enhance the model's sensitivity to variations across different temporal granularities, and consistency constraints are applied to ensure smoothness and stability at both global and local levels. To handle highly dynamic load scenarios in complex cloud environments, the model further establishes a closed-loop residual feedback pathway, enabling continuous learning and self-correction under distribution shifts. This effectively suppresses prediction bias and error accumulation. The proposed approach not only achieves a coordinated representation of global trends and local perturbations in structure but also integrates residual dynamic compensation and temporal consistency into a unified optimization framework. Experimental results on representative backend metric datasets demonstrate that the model achieves low-error and high-stability performance across multi-dimensional scenarios, significantly improving prediction accuracy and robustness under non-stationary distributions, and providing a scalable solution for dynamic modeling and intelligent operations of complex systems.

Keywords: Non-stationary modeling; residual calibration; multi-scale fusion; temporal consistency

1. Introduction

In the continuous evolution of cloud computing and large-scale distributed systems, backend systems have become the key foundation for ensuring business continuity and performance stability. Their operational states are typically monitored in real time through multi-dimensional metric sequences. These metrics include CPU utilization, memory usage, I/O latency, and request throughput, reflecting the time-varying characteristics and dynamic coupling of system resources. With the increasing complexity of workloads and the widespread adoption of microservice architectures, backend metrics exhibit clear non-stationary characteristics. The statistical distributions change over time, and correlations among different dimensions shift dynamically. Such non-stationarity arises not only from fluctuations in external request patterns and resource scheduling strategies but also from internal component coordination, system interference, and sudden anomalies. As a result, traditional time-series modeling methods based on stationarity assumptions often suffer from performance degradation and loss of generalization in long-term forecasting and anomaly detection tasks[1,2].

Against this background, how to effectively model non-stationary backend metrics has become a central challenge in intelligent operations and cloud prediction research. Traditional linear models rely on fixed temporal dependencies and stable statistical features. When faced with cross-cycle drifts, sudden load fluctuations, or multi-tenant interference, they struggle to maintain predictive consistency and stability. Recently, deep learning has achieved remarkable success in time-series forecasting through convolutional, recurrent, and transformer-based architectures that capture multi-scale features and complex dependencies. However, in non-stationary backend scenarios, these models still face significant limitations. On one hand, they lack robustness against anomalous disturbances and system noise, causing residual distributions to deviate from ideal assumptions. On the other hand, accumulated feature dynamics amplify prediction bias, leading to systematic drift over time. Therefore, temporal pattern learning alone is insufficient to describe complex system dynamics. It is necessary to introduce structured modeling and calibration mechanisms at the residual level to achieve adaptive correction and stable prediction[3].

The core idea of residual calibration modeling is to characterize and correct the distribution of prediction errors, thereby compensating for bias and enhancing generalization under complex conditions. In backend systems, residuals often contain high-frequency dynamics and contextual disturbances that the primary model fails to capture explicitly[4]. Their distribution exhibits strong time variation and heterogeneity. By modeling the residual process and performing dynamic calibration, a self-adaptive feedback pathway can be built outside the main model[5]. This enables stable outputs even under distribution shifts, abnormal fluctuations, and nonlinear noise. Such a mechanism not only improves prediction accuracy and robustness but also provides an interpretable approach to error regulation. In cloud backend environments, the residual calibration model can act as a compensatory layer for the core prediction module, forming a closed-loop structure that effectively mitigates performance degradation caused by environmental changes.

From a system engineering perspective, non-stationarity is reflected not only in dynamic changes of metric values but also in the evolution of statistical dependencies and energy distributions across time scales. Backend metrics often contain both short-term fluctuations and long-term trends, and their interactions across temporal resolutions form multi-level non-stationary dynamics. Single-scale modeling approaches fail to balance local sensitivity and global consistency, leading to uneven performance across time windows. Introducing residual calibration mechanisms enables dynamic consistency constraints on top of multi-scale modeling. By sharing residual information and uncertainty estimation across scales, the model achieves cross-scale calibration while capturing features at multiple temporal granularities. This allows for more stable dynamic forecasting in complex systems. Such an approach not only enriches the theoretical framework of non-stationary time-series modeling but also provides a feasible path for intelligent prediction and resource scheduling in multi-layer cloud systems[6].

In summary, research on residual calibration modeling and prediction for non-stationary backend metrics holds significant theoretical and practical value. Theoretically, it expands the generalization boundary of time-series modeling under non-stationary distributions and offers a new framework for error modeling and adaptive calibration in complex dynamic systems. Practically, it enhances the stability of cloud backend predictions under dynamic loads, anomalous fluctuations, and environmental changes, supporting decision-making in intelligent scheduling, capacity planning, and service quality assurance. In future cloud-native and autonomous operation systems, developing self-calibrating predictive models will be a key step toward sustainable optimization and adaptive evolution, with residual calibration modeling serving as a critical pathway to that goal[7].

2. Proposed Approach

This study introduces a residual calibration prediction modeling method for non-stationary backend metric sequences. The goal is to explicitly model the residual distribution of the primary predictor to achieve adaptive correction of system dynamic drift and structural noise. Unlike traditional time-series models that rely on fixed temporal dependencies, the proposed method builds a dual-layer collaborative framework composed of a primary predictor and a residual calibrator. The primary predictor captures the deterministic evolution of system metrics, while the calibrator dynamically learns the time-varying structure of the residuals to perform distribution-level correction of the predictions. This framework effectively models multi-scale non-stationary dynamics in complex backend environments and maintains structural consistency and robustness in long-term forecasting. The model architecture is shown in Figure 1.

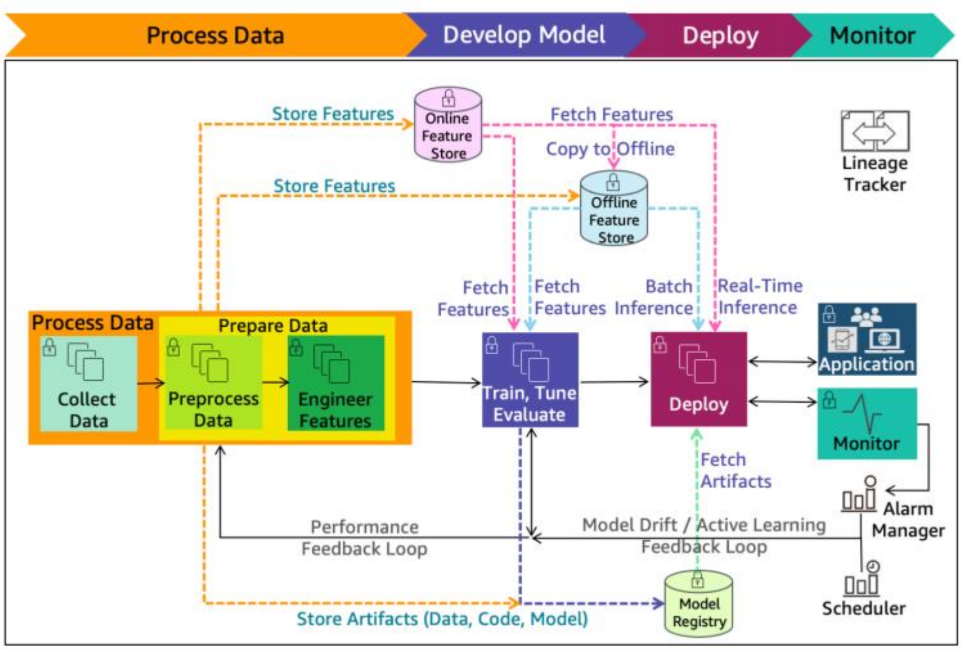

Figure 1. Overall model architecture

First, let the multidimensional indicator vector of the backend system at time step t be:

$$X_t = \left[x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(d)}\right]^T \tag{1}$$

Where d represents the indicator dimension. To capture its temporal evolution law, the main predictor generates a first-order prediction value through a nonlinear mapping function $f_\theta(\cdot)$:

$$\hat{x}_{t+1} = f_\theta(x_{t-p:t}) \tag{2}$$

Where $x_{t-p:t}$ represents the input window of the past p time steps. This mapping can be composed of convolutional, attention, or transformation-based structures to model the interaction between short-term fluctuations and long-term trends.

However, under non-stationary distribution, the statistical characteristics of the prediction error are not independent and identically distributed, but change dynamically with time and context. To this end, the residual definition is introduced:

$$r_t = x_t - \hat{x}_t \tag{3}$$

Where r_t represents the dynamic components that the primary predictor fails to capture, including sudden interference, structural noise, and potential dependency shifts. Direct learning of the residual can be viewed as a high-frequency compensation process for the primary prediction space, so a conditional dependency model is required to characterize its distribution shift.

The residual calibrator dynamically corrects the prediction residuals by modeling the conditional expectation and covariance structure, which can be formalized as:

$$\widetilde{r}_{t+1} = g_\phi(r_{t-q:t}, x_{t-p:t}) \tag{4}$$

Where $g_\phi(\cdot)$ represents a parameterized calibration function whose input includes the residual history and indicator context, which is used to generate the calibration estimate $\widetilde{r}_{t+1}$ for the next time step. Through this mechanism, the model can establish an adaptive feedback path in the time dimension and achieve dynamic convergence of the prediction error.

Combining the main prediction and the calibration prediction, the final output can be expressed as:

$$\widetilde{x}_{t+1} = \hat{x}_{t+1} + \widetilde{r}_{t+1} \tag{5}$$

This formula embodies the idea of structured residual correction: by superimposing a calibration term on the main prediction space, the model can compensate for systematic biases caused by nonstationarity and enhance the stability of prediction results under distribution drift. This form essentially constitutes an implicit two-stage mapping mechanism, where f_θ captures the deterministic component and g_ϕ models statistical fluctuations, forming complementary representations in the semantic space.

To ensure balanced modeling at different time scales, this study further introduces a cross-scale consistency constraint to maintain global temporal smoothness by minimizing the dynamic differences between predictions at different scales. Let the scale set be $S=\{s_1,s_2,...,s_k\}$, then the constraint objective is defined as:

$$L_{cons}=\sum_{s_i,s_j\in S}\left\|D_{s_i}(\tilde{x})-D_{s_j}(\tilde{x})\right\|_2^2 \tag{6}$$

Where $D_{s_i}(\cdot)$ represents the temporal feature transformation operator at scale B, such as moving average or scale convolution. This constraint allows the model to maintain feature consistency in multi-scale predictions, thereby avoiding short-term noise amplification and long-term trend shift.

Finally, the main predictor and the residual calibrator are jointly optimized to achieve end-to-end learning. The model is trained by minimizing the comprehensive loss:

$$L=\left\|x_{t+1}-\tilde{x}_{t+1}\right\|_2^2+\lambda L_{cons} \tag{7}$$

The first term constrains prediction accuracy, the second implements cross-scale consistency regularization, and λ is the balance coefficient. Through this joint optimization mechanism, the model can adaptively adjust the prediction structure and residual correction strength in non-stationary environments, achieving robust prediction of complex system indicators.

3. Performance Evaluation

3.1 Dataset

This study adopts the Microservices Bottleneck Localization Dataset as the foundation for experimental validation and model development. The dataset contains approximately forty million request tracing records, along with corresponding time-series data of multi-dimensional performance metrics such as CPU usage, memory consumption, I/O latency, and network throughput. It characterizes performance fluctuation patterns of microservice systems under various load and bottleneck conditions. Each tracing record corresponds to the execution path of a request along the service invocation chain, recording the resource metric variations of multiple microservices on that path. This design effectively reflects real backend phenomena such as latency propagation, resource contention, and performance coupling.

The multi-dimensional metric sequences in this dataset exhibit typical non-stationary characteristics. System load fluctuations, bursty traffic, resource scheduling adjustments, and environmental disturbances cause the metric distributions to drift and shift over time. Moreover, the dataset includes performance degradation samples under different bottleneck scenarios, which force the residual calibration model to adapt to changing system states. The choice of this dataset provides a rigorous test for evaluating the model's capability to predict under complex dynamic loads, structural drifts, and resource coupling conditions.

By applying the proposed method to this dataset, the study can deeply investigate the error structure of the primary predictor in complex backend environments and the compensation mechanism of the residual calibration module for error drift. The dataset not only provides realistic multi-dimensional time-series inputs but also supports comprehensive analysis of the model's generalization and stability under non-stationary conditions, thereby validating the effectiveness and applicability of the residual calibration modeling mechanism in backend metric prediction tasks.

3.2 Additional Dataset

To further verify the generalization ability and robustness of the proposed residual calibration framework, we perform a cross-dataset evaluation using the Cloud Operation KPI Drift Dataset. Unlike the original Microservices Bottleneck Localization Dataset, this dataset emphasizes gradual distribution drift and evolving operational patterns caused by continuous system scaling, deployment updates, and changing traffic profiles.

Instead of retraining the model, we directly apply the model trained on the first dataset to this new dataset without additional fine-tuning, in order to assess its out-of-distribution performance and cross-environment adaptability. This setting is particularly meaningful for real-world backend systems, where model redeployment without complete retraining is often required.

The results of this cross-dataset experiment are summarized in Table 1, showing that the proposed method consistently maintains superior performance compared with baseline approaches, thereby demonstrating its robustness and transferability under heterogeneous non-stationary conditions.

Table 1. Performance comparison on the Cloud Operation KPI Drift Dataset

Method	MSE	MAE	RMSE	MAPE (%)
Autoformer [8]	0.292	0.338	0.54	6.14
DSSRNN [9]	0.25	0.31	0.498	5.56
Powerformer [10]	0.268	0.326	0.516	5.84
Ours	0.205	0.281	0.452	4.51

To provide a more intuitive comparison, Figure 2 presents a visualized performance comparison of the four models on both datasets in terms of MSE. As shown in the figure, the proposed residual calibration model consistently achieves the lowest MSE across both datasets, confirming its superior adaptability to distributional shifts. Notably, while all baseline models suffer from increased errors on Dataset 2, our model exhibits only a marginal degradation, highlighting its strong generalization capability in unseen operational environments.

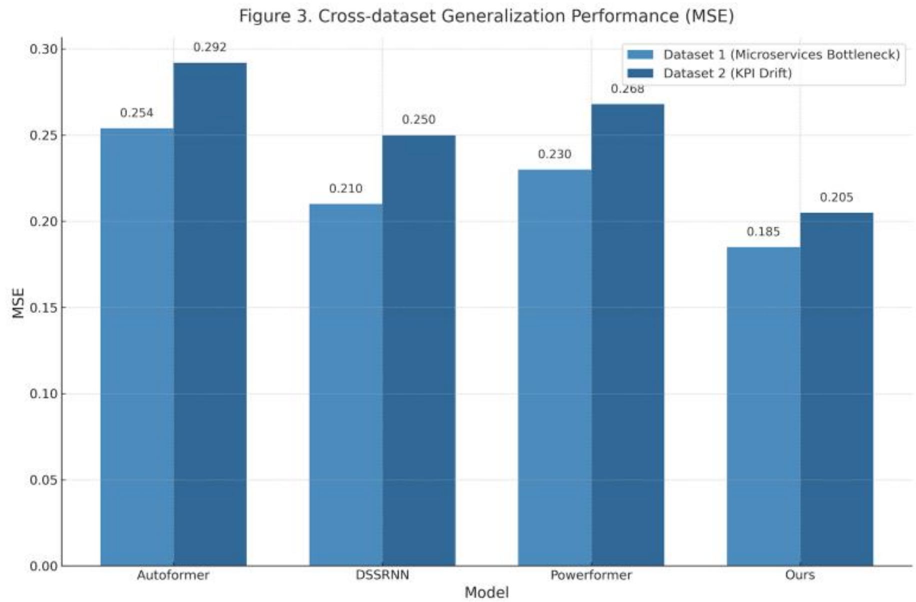

Figure 2. Cross-dataset generalization performance of different models on two benchmark datasets.

3.3 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 2.

Table2. Comparative experimental results

Method	MSE	MAE	RMSE	MAPE (%)
Autoformer[8]	0.254	0.317	0.504	5.82
DSSRNN[9]	0.210	0.289	0.458	4.97
Powerformer[10]	0.230	0.305	0.480	5.45
Ours	0.185	0.270	0.430	4.30

From the overall trend, the proposed residual calibration modeling method outperforms all comparison models across four key metrics, demonstrating stronger adaptability and prediction stability for non-stationary backend indicators. Compared with the sequence decomposition-based Autoformer, the proposed method reduces MSE and MAE by approximately 27.2% and 14.8%, respectively, indicating higher precision in capturing complex load fluctuations and cross-period dependencies. This advantage mainly arises from the residual calibration pathway, which adaptively corrects prediction errors, ensuring consistency and controllability of outputs under sudden traffic surges or resource scheduling variations.

Compared with the recursive state-enhanced DSSRNN, the proposed model achieves about 6% and 13% improvement in RMSE and MAPE, respectively. This shows that it more effectively balances global trends and local dynamic features under nonlinear disturbances and multidimensional coupling signals. Although DSSRNN has strong capability in modeling temporal evolution, its fixed state update mechanism struggles to handle dynamic distribution drift in system metrics. In contrast, the residual calibration module performs conditional modeling and statistical correction on the error sequence, allowing adaptive responses to time-varying distributions and complex non-stationary patterns.

Compared with the structure-attention-based Powerformer, the proposed method still achieves the best performance in RMSE and MAPE. This demonstrates that it not only maintains strong robustness in reconstructing the error space but also provides more stable control over proportional errors. The global attention mechanism of Powerformer tends to accumulate noise in high-fluctuation intervals, while the residual calibration framework enforces continuity constraints to ensure smooth prediction outputs. Consequently, the model maintains consistent prediction quality across different temporal scales. This property is especially critical for cloud backend systems, where performance indicators are often influenced by resource contention and dynamic topological changes, requiring a balance between local sensitivity and global consistency.

Overall, the residual calibration modeling approach significantly enhances the model's generalization and robustness under non-stationary conditions by introducing error compensation and consistency regularization mechanisms beyond the primary predictor. This mechanism enables the model to continuously learn error patterns caused by system state drift and maintain stability during long-term forecasting. The results demonstrate that the proposed method exhibits strong dynamic adaptability and generalization potential in complex backend metric scenarios, providing more reliable predictive support for intelligent operations and performance optimization in backend systems.

This paper also evaluates the robustness under load burst and traffic spike scenarios. The experimental results are shown in Figure 3.

From the overall trend, as the intensity of traffic spikes increases, the four core metrics exhibit distinct dynamic variation patterns, reflecting the model's multi-level response mechanisms under non-stationary environments. The MSE shows an accelerating upward trend with increasing spike intensity, indicating that when system load fluctuates sharply, the primary predictor accumulates noticeable global trend errors. However, the overall growth remains controlled, suggesting that the model maintains relatively stable global error bounds even under severe distribution drift. This demonstrates the fundamental robustness of the residual calibration module when handling large-scale perturbations.

The MAE curve displays a clear saturation pattern, with error growth slowing in high-load regions, indicating the model's adaptive correction capability under medium-to-high load conditions. This trend results from the smoothing effect of the residual calibration module through multi-scale fusion, which dynamically identifies and compensates for short-term local deviations caused by transient spikes. As a result, the overall mean error no longer scales linearly with load intensity, highlighting the model's flexible suppression of instantaneous anomalies in complex backend environments.

The RMSE exhibits an "inflection-type" growth pattern, showing mild variation in low-intensity regions and significant escalation under high-intensity loads. This phenomenon reveals that when the load surpasses the system's stability threshold, fluctuation energy becomes concentrated in a few high-amplitude segments, causing localized errors to increase rapidly. While the model maintains temporal smoothness through consistency constraints, it still faces challenges from accumulated high-frequency disturbances under extreme spike conditions. This suggests that although residual calibration can delay error propagation in highly non-stationary settings, the overall dynamics remain constrained by the upper limits of high-frequency noise interference.

The MAPE curve remains relatively smooth with minor oscillations, indicating high self-stability in terms of proportional error. The fluctuations mainly occur in the medium-load range, reflecting the model's balance between adaptive residual compensation and scale normalization. Overall, the trend shows that residual calibration modeling dynamically adjusts the proportional distribution of prediction bias under different traffic spike intensities, thereby maintaining coherent and continuous prediction performance under non-stationary distributions.

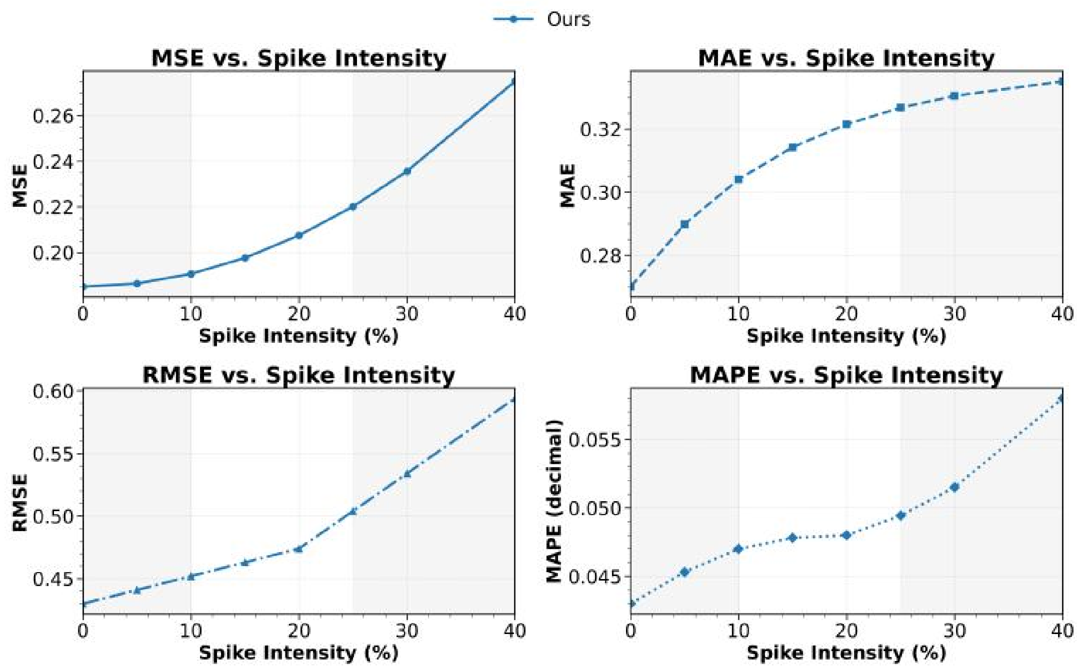

Figure 3. Robustness evaluation under load burst and traffic spike scenarios

4. Conclusion

This study investigates the problem of dynamic modeling and residual calibration prediction for non-stationary backend metrics. To address key challenges such as metric distribution drift, bursty load fluctuations, and complex contextual dependencies in cloud systems, a residual calibration modeling framework with adaptive error correction capability is proposed. Through a dual-layer collaborative mechanism between the primary predictor and the residual calibrator, the framework achieves structured awareness of system state changes and temporal dynamic compensation. It maintains global trend consistency while effectively suppressing prediction deviations caused by local anomalies. Experimental results demonstrate that the proposed method exhibits excellent robustness and generalization across multi-dimensional metric scenarios, enabling stable modeling of complex dependencies in non-stationary environments. This provides both theoretical and practical foundations for intelligent prediction and adaptive optimization in backend systems.

From the perspective of model design, the proposed residual calibration concept breaks through the limitation of traditional time-series prediction models that rely solely on a single predictive pathway. By explicitly modeling the residual distribution in the prediction space, the model dynamically learns the evolution pattern of system errors and applies consistency constraints to achieve smooth calibration across temporal scales. This mechanism theoretically extends the modeling paradigm for non-stationary time series, allowing the model to capture changing patterns at a finer granularity while enhancing interpretability and controllability at the structural level. For backend metric sequences characterized by long-term dependencies and high-dimensional heterogeneity, this mechanism provides greater elasticity and dynamic responsiveness, establishing a reliable foundation for predictive maintenance and resource scheduling in backend systems.

From an application standpoint, residual calibration modeling has significant potential in fields such as cloud computing, microservice operations, and intelligent resource management. As cloud environments become increasingly dynamic and task diversity continues to grow, metric fluctuations and non-stationarity have become major obstacles to performance optimization and anomaly detection. The proposed modeling framework offers a lightweight and efficient adaptive approach for such systems. By introducing a learnable residual feedback channel, it continuously refines predictive performance without requiring large-scale structural retraining. This property is particularly valuable for rolling deployment, adaptive scheduling, and online learning in real systems, further advancing the integration of intelligent cloud operations and autonomous system prediction.

Looking ahead, the concept of residual calibration has the potential to play a broader role in time-series modeling. For instance, in multi-modal system metric analysis, structured residuals can be extended to cross-domain feature alignment and uncertainty quantification. In distributed resource forecasting, a hierarchical residual propagation mechanism can be introduced to capture dynamic inter-node correlations. In reinforcement learning-based adaptive systems, residual calibration can serve as a stabilizing factor for policy optimization, enabling joint evolution of prediction and decision-making. Furthermore, combining residual calibration with large models and self-supervised pretraining represents a promising direction toward transitioning from task-specific forecasting to general system understanding. Through continuous advances in both theoretical modeling and engineering implementation, the ideas presented in this study may promote the development of non-stationary time-series modeling toward greater robustness, interpretability, and adaptivity, offering forward-looking technological support for long-term autonomy and efficient orchestration in intelligent cloud systems.

References

[1] Liu Y, Wu H, Wang J, et al. Non-stationary transformers: Exploring the stationarity in time series forecasting[J]. Advances in neural information processing systems, 2022, 35: 9881-9893.

[2] Xiong Y, Wen Y. Non-Stationary Time Series Forecasting Based on Fourier Analysis and Cross Attention Mechanism[J]. arXiv preprint arXiv:2505.06917, 2025..

[3] Zheng J, Yang Z, Ge Z. Spectra data calibration based on deep residual modeling of independent component regression[J]. Chemometrics and Intelligent Laboratory Systems, 2024, 255: 105270.

[4] Liu H, Zhong Z, Sebe N, et al. Mitigating robust overfitting via self-residual-calibration regularization[J]. Artificial Intelligence, 2023, 317: 103877.

[5] Ma X, Zhang H. Time Series Forecasting Method Based on Multi-Scale Feature Fusion and Autoformer[J]. Applied Sciences, 2025, 15(7): 3768.

[6] Shi K, Ding Z, Chen Z, et al. A multi-scale time series forecasting framework with temporal hierarchical information fusion and reconciliation[J]. Data Mining and Knowledge Discovery, 2025, 39(4): 32.

[7] Nayak I, Chakrabarti A, Kumar M, et al. Temporally-consistent koopman autoencoders for forecasting dynamical systems[J]. Scientific Reports, 2025, 15(1): 22127.

[8] Wu H, Xu J, Wang J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting[J]. Advances in neural information processing systems, 2021, 34: 22419-22430.

[9] Mohammadshirazi A, Nosratifiroozsalari A, Ramnath R. DSSRNN: Decomposition-Enhanced State-Space Recurrent Neural Network for Time-Series Analysis[J]. arXiv preprint arXiv:2412.00994, 2024.

[10] Suresh V. Benchmarking Transformer Variants for Hour-Ahead PV Forecasting: PatchTST with Adaptive Conformal Inference[J]. Energies, 2025, 18(18): 5000.