LoG-ADFormer: A Battery Long-Horizon Capacity Forecasting Method

Yue Liu*
School of Computer Engineering and Science
Shanghai University
Shanghai, China
yueliu@shu.edu.cn

Jiawei Ni
School of Computer Engineering and Science
Shanghai University
Shanghai, China
njw0218@shu.edu.cn

Wei Zuo
School of Computer Engineering and Science
Shanghai University
Shanghai, China
zuowei@shu.edu.cn

Xinyue Jin
Materials Genome Institute
Shanghai University
Shanghai, China
812255427@qq.com

Siqi Shi
State Key Laboratory of Materials for Advanced Nuclear Energy & School of Materials Science and Engineering
Shanghai University
Shanghai, China
sqshi@shu.edu.cn

***Abstract*—Long-horizon forecasting for lithium-ion batteries is essential for practical health management. Yet reliable multi-horizon prediction remains challenging under non-stationary operation, rest-induced recovery, and nonlinear aging. We propose LoG-ADFormer, a local-to-global forecasting framework that improves long-horizon stability by mitigating level drift. It introduces an Abs-Delta dual-view prediction design to jointly model the absolute capacity level and incremental evolution, and uses horizon-aware fusion to adaptively balance the two views across future steps. This design enhances robustness to non-stationary operating conditions, rest-induced capacity recovery, and nonlinear aging dynamics. We conduct experiments on four public battery degradation datasets, spanning different cell chemistries and temperature conditions, to evaluate the model under multiple prediction horizons and lookback lengths. LoG-ADFormer achieves the best performance in 75% of long-horizon forecasting settings, with an overall average MAE of 0.3468. These results suggest that LoG-ADFormer offers a robust and practical solution for long-horizon capacity prognosis, enabling more reliable maintenance planning and risk-aware battery operation.**

***Keywords*—Battery life prediction, Long-term time series, Time series forecasting, Horizon-aware fusion**

I. INTRODUCTION

Long-horizon forecasting for lithium-ion batteries is important for practical battery health management, as many real-world decisions depend not only on the current state but also on how the degradation trajectory will evolve over an extended future horizon. In applications such as electric vehicles, energy storage systems, and maintenance scheduling, the forecasting model is expected to provide stable multi-step predictions from a limited observation window [1]. However, this task remains difficult because battery degradation is a long-term, nonlinear, and stage-dependent process. Under realistic operating conditions, the observed trajectories are further affected by temperature variation, dynamic load profiles, rest-induced recovery, and measurement noise [2, 3]. As a result, battery long-horizon forecasting requires not only low point-wise error, but also a reliable future degradation trajectory.

A key difficulty in this setting is trajectory-level drift. Even when short-range predictions appear reasonable, small deviations at earlier forecasting steps may accumulate over the horizon, gradually shifting the predicted capacity curve away from the true degradation path [4]. This issue is especially critical for battery prognosis, since long-horizon errors can directly affect remaining useful life estimation, threshold-crossing judgment, and risk-aware maintenance decisions. At the same time, battery degradation data contain two coupled temporal characteristics: local short-term variations, such as transient fluctuations and recovery-like perturbations, and long-term degradation trends associated with stage-wise capacity decline. Effective forecasting therefore requires both sensitivity to informative local dynamics and stability in long-range extrapolation. Existing methods, however, often emphasize one aspect at the expense of the other, leading either to noisy forecasts that overreact to local disturbances or to oversmoothed forecasts that fail to preserve the true degradation evolution [5].

To this end, a long-horizon battery degradation forecasting method tailored for application-oriented robustness, termed the Local-to-Global Absolute–Delta Fusion Transformer (LoG-ADFormer), is proposed. The model follows a local-to-global representation learning paradigm: it compresses long lookback sequences into patch tokens and employs a lightweight local feature encoder to capture short-term variations and turning points while suppressing high-frequency noise. The resulting representations are then refined by a global dependency module to summarize long-term degradation dynamics. Built upon the learned global representation, an Abs-Delta dual-view forecasting head is introduced, through which future capacity is predicted from two complementary views: an absolute-level view that provides global calibration, and an increment view that constrains step-wise evolution to reduce drift accumulation. A horizon-aware fusion mechanism further balances these views

* Corresponding author

across future steps, which is particularly important for long-horizon battery forecasts where error characteristics vary with prediction distance.

Evaluation is conducted on four publicly available battery degradation datasets under long-horizon forecasting settings to assess the performance of LoG-ADFormer. The results show that LoG-ADFormer achieves higher robustness and better accuracy, particularly at extended horizons, compared with strong long-term forecasting baselines. Our contributions are summarized as follows:

- A local-to-global modeling method is proposed to progressively summarize long input sequences from short-term micro-patterns to long-range degradation dynamics, tailored for battery degradation forecasting.
- A horizon-aware Absolute–Delta Fusion mechanism is developed to synergize level prediction with increment-based recovery, thereby significantly improving stability and reducing drift in long-term outputs.
- Comprehensive evaluations on multiple battery degradation datasets, together with ablation studies, confirm the effectiveness of the proposed components in long-term time series forecasting.

The remainder of this paper is organized as follows. Section II reviews related work. Section III details the proposed LoG-ADFormer. Section IV presents the experimental setup, results, and analysis. Finally, Section V concludes the paper.

II. Related Work

A. Long-Term Time-Series Forecasting

Early deep forecasting methods mainly relied on sequential architectures such as LSTMs/GRUs [6] and later convolutional models such as TCNs to capture temporal dependencies. More recently, Transformer-based models have become a dominant paradigm for long-term time-series forecasting due to their ability to model long-range interactions. At the same time, DLinear [7] shows that simple decomposition-based designs can remain highly competitive, indicating that accurate long-horizon forecasting depends not only on stronger attention mechanisms but also on effective temporal structure modeling.

Building on this trend, recent methods have explored more structured representations for LTSF. PatchTST [8] improves efficiency and stability through patch-based tokenization, while Crossformer [9] and iTransformer [10] further enhance multivariate dependency modeling. Other methods, such as TimesNet [11] and TimeMixer [12], aim to capture multi-scale temporal patterns and periodic structures [13, 14]. However, battery degradation trajectories are dominated by non-stationary and stage-dependent trends rather than strong periodicity. Therefore, directly applying generic pattern-oriented models may not adequately address the trajectory-level drift problem in battery prognosis.

Overall, effective LTSF solutions must preserve fine-grained local patterns while maintaining globally consistent extrapolation. Unlike prior work that focuses on generic temporal representation or periodicity modeling, our work targets a specific failure mode critical to battery prognosis—trajectory-level drift. A local-to-global design, together with an Abs-Delta dual-view formulation and horizon-aware fusion, is introduced to explicitly enhance the stability of long-horizon forecasts, rather than merely improving generic context modeling.

B. Long-Horizon Battery Life Prediction

Recent battery prognosis research is increasingly moving beyond single-point SOH/RUL estimation toward long-horizon forecasting of degradation trajectories (e.g., capacity/SOH) [15], which is directly relevant to maintenance scheduling and RUL-oriented planning. Compared with generic forecasting benchmarks, battery degradation time series are particularly challenging due to strong non-stationarity and stage dependency, as well as intermittent perturbations induced by operating-condition variability, rest-related regeneration, and measurement noise. These characteristics make long-range extrapolation highly susceptible to error accumulation and level drift, which can systematically bias future trajectories and compromise downstream decisions [16].

To improve long-range modeling capacity, recent studies have adopted stronger temporal backbones, especially Transformer-style architectures for battery health modeling and forecasting, leveraging attention mechanisms to capture longer dependencies and cross-cycle correlations [17, 18]. In parallel, emerging work has started to explore foundation-model and generative perspectives for battery degradation, aiming to learn transferable temporal priors and support earlier or more generalizable prediction of long-term aging trends [19].

Despite these advances, two issues remain prominent. First, horizon-conservative evaluation can underexpose drift-dominant failure modes that only emerge under extended prediction windows. Second, even with modern sequence models, forecasts often become either overly sensitive to local disturbances or excessively smoothed, obscuring stage transitions and turning points critical for maintenance planning [5]. Therefore, long-horizon battery forecasting should be assessed not only by point-wise accuracy, but also by trajectory stability and drift behavior. In contrast to prior work that largely improves backbone capacity or pursues generic transferability, our method explicitly targets long-horizon drift as a critical practical challenge in battery forecasting, using a local-to-global design with an Abs-Delta dual-view predictor and horizon-aware fusion to jointly stabilize level calibration and incremental evolution under non-stationary regimes.

III. Methodology

A. Problem Formulation

Given a multivariate battery measurement time series $\{\mathrm{x}_t\}$, where $\mathrm{x}_t \in \mathbb{R}^N$ denotes N observed variables at time t, our goal is to forecast the future capacity trajectory over a long horizon. An input segment $\mathbf{X}_t = [\mathrm{x}_{t-L+1}, \dots, \mathrm{x}_t] \in \mathbb{R}^{L\times N}$, is constructed using a sliding-window strategy with a lookback length L and stride s. The forecasting target is the future capacity sequence $\boldsymbol{y}_t = [y_{t+1}, \dots, y_{t+T}] \in \mathbb{R}^T$, where T is the prediction horizon and y_{t+k} denotes the capacity at step $t + k$. The model learns a mapping $f_\theta(\cdot)$ to produce a one-shot multi-step forecast $\hat{y}_t = f_\theta(\mathbf{X}_t) \in \mathbb{R}^T$.

B. Model Overview

Fig. 1 presents the overall architecture of our forecasting model. The framework follows a local-to-global pipeline: it tokenizes the multivariate degradation sequence into patch

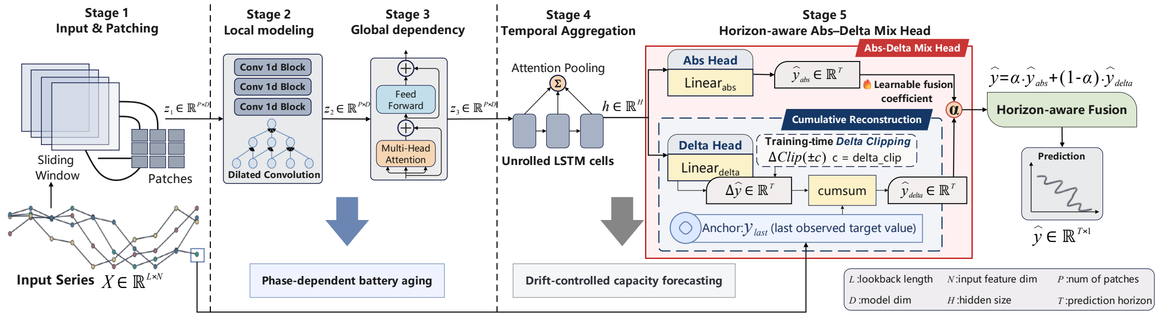

Fig. 1. Overview of LoG-ADFormer for long-horizon battery capacity forecasting. The degradation sequence is segmented into patches to preserve local aging patterns while reducing sequence length. A TCN captures noise-robust local dynamics, and a Transformer models cross-patch global dependencies. The resulting tokens are aggregated by an LSTM with attention pooling into a compact degradation state. The Abs-Delta Mix Head then predicts an absolute trajectory for level calibration and an increment trajectory for step-wise evolution. The Delta branch is cumulatively summed and anchored to the last observed capacity to reduce long-horizon drift, and horizon-aware fusion combines both views for final multi-step forecasting.

representations, extracts temporal patterns across multiple scales, and aggregates them into a global degradation state. On top of this state, a horizon-aware Abs–Delta Mix Head produces multi-step capacity forecasts by combining absolute-level prediction with increment-based correction, which is designed to mitigate long-horizon drift.

The design addresses two key requirements: robust representation under non-stationarity and drift-resistant extrapolation. Patching preserves local aging morphology while reducing sequence length, and the hybrid encoder captures both local transitions and long-range dependencies. For drift control, the Delta trajectory is reconstructed via cumulative summation anchored to the last observation, and horizon-aware fusion balances the Abs and Delta views across steps for stable long-horizon forecasts.

C. Local-to-Global Degradation Representation Learning

Given a multivariate lookback window $\mathbf{X}_{1:L} \in \mathbb{R}^{L \times N}$, a patch-based tokenization strategy is adopted to preserve local temporal morphology while reducing the effective sequence length for long-horizon forecasting. Specifically, $\mathbf{X}_{1:L}$ is partitioned into P overlapping patches using a patch length l_p and stride s_p, where the p-th patch is denoted as $\mathbf{X}^{(p)} \in \mathbb{R}^{l_p \times N}$. Overlapping patching not only improves continuity between adjacent segments, but also alleviates boundary artifacts by allowing local patterns to be captured from multiple aligned receptive windows. Each patch is then flattened into a vector $\mathrm{vec}(\mathbf{X}^{(p)}) \in \mathbb{R}^{l_p \times N}$ and projected into a D-dimensional latent token through a learnable linear projection:

$$z_p = \mathrm{vec}(\mathbf{X}^{(p)}) W_e + b_e \,, W_e \in \mathbb{R}^{(l_p N) \times D} \tag{1}$$

which yields a compact token sequence $Z = [\mathrm{z}_1, \ldots, \mathrm{z}_P] \in \mathbb{R}^{P \times D}$. Since patching transforms the original time axis into a shorter token index axis, positional encodings are added to preserve patch order and temporal consistency in global modeling.

On top of patch tokens, a Temporal Convolutional Network (TCN) is employed as a local-to-mid-range feature encoder before global attention. Concretely, the TCN consists of stacked 1D temporal convolutional blocks with dilated convolutions, residual connections, and nonlinear activations. For the k-th block, dilation expands the receptive field with depth, enabling the network to capture short-term fluctuations and local turning points while keeping computation linear in sequence length. This convolutional inductive bias is particularly beneficial for battery degradation signals, where operational perturbations (e.g., temperature/load variations), rest-induced capacity recovery, and short-term non-stationarity are common and can mislead purely attention-based modeling if treated as long-term trends. Moreover, operating on the shortened patch-token sequence rather than the raw long sequence improves efficiency and yields a robust intermediate representation that is less sensitive to measurement noise.

The TCN-refined token representations are then fed into a Transformer encoder for global dependency modeling. Self-attention relates distant patches to summarize stage-dependent aging dynamics over long horizons, while patching and the TCN provide locality priors that stabilize long-range interaction learning under non-stationary conditions. This yields a local-to-global representation pipeline that captures both fine-grained aging morphology and long-range degradation trends, supporting drift-resistant long-horizon capacity forecasting.

D. Temporal Aggregation and Drift-Resistant Abs-Delta Capacity Head

After obtaining local-to-mid representations over patch tokens, a Transformer encoder is employed to capture long-range interactions among tokens and to model global battery degradation dynamics. The resulting token sequence is then processed by an LSTM to introduce sequential inductive bias along the patch timeline, which is helpful for stage-dependent aging evolution. To avoid over-reliance on a single token under non-stationary degradation patterns induced by operating-condition shifts and rest-induced recovery, attention pooling is further applied to aggregate the LSTM hidden states into a compact global representation. Specifically, given the LSTM hidden states $\{s_p\}_{p=1}^{P}$, the attention weights and the pooled representation are computed as follows:

$$\beta_p = \frac{\exp\,(w^T s_p)}{\sum_{q=1}^{P} \exp\,(w^T s_q)} \,, \mathbf{h} = \sum_{p=1}^{P} \beta_p s_p \tag{2}$$

where $\mathbf{h}$ summarizes the current degradation state and serves as the input to the forecasting head.

Based on $\mathbf{h}$, the Abs-Delta Mix head produces two complementary predictions:

$$\hat{y}^{abs} = \mathrm{Head}_{abs}(\mathbf{h}) \,, \hat{y}^{delta} = \mathrm{Head}_{delta}(\mathbf{h}) \tag{3}$$

TABLE I.
MULTIVARIATE LONG-TERM FORECASTING RESULTS WITH LOG-ADFORMER. PREDICTION LENGTHS T ∈ {96, 192, 336, 720} ARE USED FOR ALL DATASETS. THE INPUT SEQUENCE LENGTH IS SET TO 336 FOR ALL MODELS. THE BEST RESULTS ARE IN **BOLD** AND THE SECOND BEST ARE UNDERLINED.

Models		LoG-ADFormer		PatchFormer		TimeMixer		iTransformer		DLinear		PatchTST		Autoformer		Informer	
Metric		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Tongji	96	**0.2838 ±0.0160**	**0.6237 ±0.0195**	0.5433 ±0.0246	0.8495 ±0.0318	0.6099 ±0.0287	0.8848 ±0.0341	0.6135 ±0.0304	0.9182 ±0.0379	0.5146 ±0.0221	0.8523 ±0.0296	0.5485 ±0.0238	0.8912 ±0.0337	0.8117 ±0.0413	1.0002 ±0.0488	0.6857 ±0.0326	0.9549 ±0.0415
	192	**0.4157 ±0.0109**	**0.7730 ±0.0101**	0.6200 ±0.0269	0.9403 ±0.0357	0.7538 ±0.0348	0.9826 ±0.0396	0.7983 ±0.0417	1.1930 ±0.0624	0.7076 ±0.0315	1.0053 ±0.0418	0.7367 ±0.0332	1.0568 ±0.0475	0.8869 ±0.0441	1.0655 ±0.0512	0.7638 ±0.0369	1.0237 ±0.0448
	336	**0.5879 ±0.0300**	**0.9239 ±0.0362**	0.8069 ±0.0387	1.1783 ±0.0536	0.8005 ±0.0364	1.0260 ±0.0437	0.8552 ±0.0418	1.1508 ±0.0529	0.7885 ±0.0349	1.0664 ±0.0476	0.8170 ±0.0372	1.0645 ±0.0468	0.8847 ±0.0435	1.0702 ±0.0494	0.8560 ±0.0407	1.1079 ±0.0516
	720	**0.7125 ±0.0268**	**0.9686 ±0.0333**	0.7844 ±0.0368	1.0727 ±0.0497	0.8318 ±0.0396	1.0368 ±0.0453	0.8596 ±0.0415	1.0953 ±0.0506	0.7576 ±0.0344	1.0140 ±0.0439	0.8108 ±0.0381	1.0324 ±0.0446	0.9154 ±0.0469	1.1676 ±0.0578	0.8730 ±0.0428	1.0852 ±0.0499
Vehicle	96	0.1569 ±0.0258	**0.4287 ±0.0649**	0.1928 ±0.0184	0.4998 ±0.0467	0.1683 ±0.0159	0.4515 ±0.0412	0.1931 ±0.0197	0.4601 ±0.0436	0.2121 ±0.0218	0.5148 ±0.0493	**0.1490 ±0.0141**	0.4317 ±0.0408	0.3668 ±0.0335	0.6509 ±0.0586	0.2742 ±0.0262	0.5051 ±0.0479
	192	0.3057 ±0.0317	0.5732 ±0.0794	0.3413 ±0.0298	0.6160 ±0.0575	0.3074 ±0.0281	0.5841 ±0.0523	0.3572 ±0.0347	0.7791 ±0.0732	0.3463 ±0.0319	**0.5707 ±0.0516**	**0.2743 ±0.0246**	0.5890 ±0.0554	0.5550 ±0.0498	0.7718 ±0.0697	0.3765 ±0.0355	0.6160 ±0.0561
	336	**0.4137 ±0.0491**	0.7023 ±0.0863	0.4645 ±0.0427	0.7215 ±0.0668	0.5051 ±0.0468	0.7505 ±0.0694	0.4875 ±0.0442	0.8380 ±0.0781	0.4460 ±0.0403	**0.6633 ±0.0612**	0.4551 ±0.0415	0.7522 ±0.0703	0.6990 ±0.0628	0.9439 ±0.0876	0.4699 ±0.0439	0.7077 ±0.0654
	720	**0.5739 ±0.0215**	**0.8267 ±0.0267**	0.6121 ±0.0248	0.8390 ±0.0312	0.7448 ±0.0331	0.9663 ±0.0386	0.6479 ±0.0275	0.9112 ±0.0357	0.5829 ±0.0234	0.8345 ±0.0310	0.6674 ±0.0292	0.9312 ±0.0365	0.9432 ±0.0417	1.1616 ±0.0489	0.7149 ±0.0314	1.0084 ±0.0407
MIT	96	0.3258 ±0.0149	**0.6725 ±0.0264**	**0.3257 ±0.0142**	0.6901 ±0.0247	0.6488 ±0.0316	1.0540 ±0.0429	0.5468 ±0.0258	1.0070 ±0.0393	0.6022 ±0.0287	0.9763 ±0.0381	0.4620 ±0.0214	0.9789 ±0.0397	0.8046 ±0.0408	1.0682 ±0.0475	0.4460 ±0.0196	0.7859 ±0.0298
	192	0.4729 ±0.0120	**0.7967 ±0.0180**	**0.4529 ±0.0178**	0.8719 ±0.0297	0.7519 ±0.0334	1.1819 ±0.0476	0.6409 ±0.0289	1.1672 ±0.0469	0.7166 ±0.0315	1.0490 ±0.0413	0.6287 ±0.0271	1.1607 ±0.0472	1.0709 ±0.0517	1.3500 ±0.0613	0.5215 ±0.0224	0.9867 ±0.0388
	336	0.6285 ±0.0427	**0.9673 ±0.0427**	**0.6236 ±0.0419**	1.0369 ±0.0431	0.7820 ±0.0497	1.1101 ±0.0476	0.7167 ±0.0448	1.1394 ±0.0495	0.7637 ±0.0472	1.0786 ±0.0452	0.7321 ±0.0455	1.1801 ±0.0528	1.1596 ±0.0668	1.3695 ±0.0624	0.6707 ±0.0437	1.0389 ±0.0446
	720	**0.7238 ±0.0271**	**1.0795 ±0.0152**	0.7921 ±0.0316	1.1075 ±0.0227	0.8989 ±0.0374	1.1800 ±0.0251	0.8509 ±0.0346	1.2223 ±0.0279	0.8299 ±0.0331	1.1233 ±0.0238	0.8305 ±0.0337	1.2260 ±0.0284	1.0756 ±0.0284	1.3091 ±0.0316	0.7599 ±0.0308	1.1309 ±0.0242
HUST	96	**0.0052 ±0.0009**	0.0113 ±0.0031	0.0054 ±0.0008	**0.0079 ±0.0016**	0.0142 ±0.0024	0.1270 ±0.0137	0.0498 ±0.0049	0.1536 ±0.0161	0.3520 ±0.0291	0.6392 ±0.0418	0.2171 ±0.0184	0.5524 ±0.0369	0.6968 ±0.0514	0.9208 ±0.0597	0.0246 ±0.0032	0.0843 ±0.0099
	192	0.0080 ±0.0003	0.0155 ±0.0027	**0.0060 ±0.0011**	**0.0093 ±0.0018**	0.0319 ±0.0042	0.1713 ±0.0186	0.0785 ±0.0068	0.1879 ±0.0198	0.4175 ±0.0334	0.6864 ±0.0457	0.4434 ±0.0362	0.8179 ±0.0533	1.3094 ±0.0875	1.5513 ±0.1028	0.0273 ±0.0036	0.0923 ±0.0108
	336	**0.0089 ±0.0023**	**0.0153 ±0.0046**	0.0125 ±0.0021	0.0234 ±0.0044	0.0264 ±0.0037	0.1337 ±0.0158	0.0897 ±0.0075	0.1915 ±0.0206	0.4263 ±0.0341	0.6844 ±0.0451	0.4505 ±0.0367	0.8228 ±0.0542	1.1973 ±0.0814	1.4611 ±0.0987	0.0298 ±0.0041	0.0977 ±0.0116
	720	**0.0103 ±0.0009**	**0.0158 ±0.0014**	0.0138 ±0.0012	0.0214 ±0.0023	0.0198 ±0.0021	0.1406 ±0.0139	0.0987 ±0.0084	0.2095 ±0.0217	0.2988 ±0.0226	0.5751 ±0.0348	0.6547 ±0.0489	1.0249 ±0.0662	0.9364 ±0.0661	1.1741 ±0.0661	0.0307 ±0.0031	0.0821 ±0.0086
1st count		**10**	**12**	4	2	0	0	0	0	0	2	2	0	0	0	0	0

where Head_{abs} and Head_{delta} denote the absolute-view and delta-view prediction heads, respectively.

To improve long-horizon stability in capacity trajectory forecasting, a horizon-aware Abs-Delta Mix head is designed to produce two complementary forecasts from **h**: an absolute trajectory prediction and an increment (delta) prediction. The increment prediction is first optionally stabilized by clipping its magnitude during training to suppress outlier step changes, i.e., $\Delta\hat{y} \leftarrow \mathrm{clip}(\Delta\hat{y}, c, -c)$, and then converted into a capacity trajectory via cumulative summation anchored at the last observed capacity y_{last}. This reconstruction is defined as:

$$\hat{y}_{t+k}^{delta} = y_{last} + \sum_{i=1}^{k} \Delta\hat{y}_{t+i}, k = 1, \dots, T \tag{4}$$

Meanwhile, the absolute branch directly predicts a future capacity trajectory $\hat{y}^{abs}$. Finally, the two trajectories are fused with horizon-wise learnable weights to adaptively balance short-term consistency and long-term level correction across steps:

$$\hat{y}_{t+k} = \alpha_k \hat{y}_{t+k}^{abs} + (1 - \alpha_k)\hat{y}_{t+k}^{delta}, k = 1, \dots, T \tag{5}$$

The entire model is trained end-to-end by minimizing the mean squared error over the prediction horizon between the fused forecast and the ground truth:

$$\mathcal{L} = \frac{1}{T}\sum_{k=1}^{T}(\hat{y}_{t+k} - y_{t+k})^2 \tag{6}$$

IV. EXPERIMENTS

A. Datasets

The proposed method is evaluated on four public lithium-ion battery degradation datasets that support long-horizon multivariate forecasting under the adopted sliding-window setting. These datasets cover different chemistries and operating conditions, and their statistics are summarized in Table I.

TABLE II. SUMMARY OF DATASETS.

Dataset	Features	Temperature	Chemistry	Time points
Tongji[20]	18	25,35,45	NCA	22278
Vehicle[21]	9	30 - 44	NCM	120966
MIT[2]	6	30	LFP	59259
HUST[22]	4	30	LFP	1163167

B. Experimental Settings

To ensure a rigorous evaluation, the proposed method is compared with representative baselines from both generic long-term time-series forecasting and battery prognosis. Specifically, TimeMixer, PatchTST, iTransformer, Autoformer [23], Informer [24], and DLinear are used as generic LTSF baselines, covering patch-based, attention-based, and linear forecasting paradigms. In addition, PatchFormer [25] is included as a battery-specific baseline for degradation forecasting.

Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are used as evaluation metrics for model performance:

$$MAE = \frac{1}{L}\sum_{i=1}^{L}|y_i - \hat{y}_i|, RMSE = \sqrt{\frac{1}{L}\sum_{i=1}^{L}(y_i - \hat{y}_i)^2} \tag{7}$$

To guarantee a fair comparison, all methods are evaluated under the same data preprocessing pipeline, train-validation-test split, and prediction setting. Baseline hyperparameters are selected according to the original implementations or widely used standard configurations, while all methods are trained under the same optimization protocol and early-stopping strategy for reproducibility. Specifically, LoG-ADFormer and all baselines are trained on an NVIDIA RTX A6000 GPU with 48 GB memory. Adam is used as the optimizer with an initial learning rate of 1×10^{-4}. The batch size is set to 32, and all models are trained for up to 30 epochs with early stopping, where the patience is set to 3. To avoid information leakage, the data are

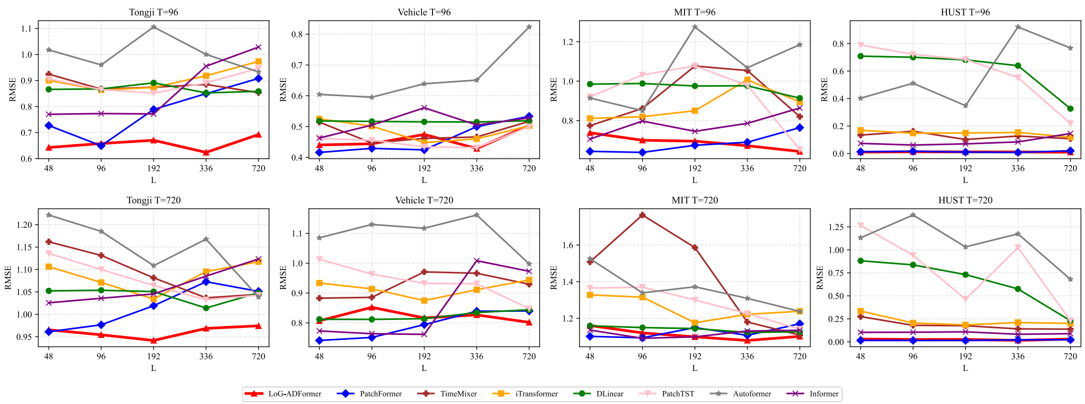

Fig. 2. RMSE scores for different lookback windows on four datasets. The X-axis shows the lookback window L = [48, 96, 192, 336, 720], the Y-axis shows the RMSE. The prediction horizons are T = 96 and T = 720.

split by cell, so that all cycles from the same cell appear in only one of the training, validation, and test sets. Unless otherwise specified, a 7:1:2 split is used for all methods.

C. Performance Comparison with State-of-the-Art Methods

The results in Table II demonstrate that LoG-ADFormer achieves the strongest overall performance across datasets and forecasting horizons. Specifically, it obtains the best MAE in 10 out of 16 settings and the best RMSE in 12 out of 16 settings, while remaining within the top-2 in all settings for both metrics. The accompanying standard deviations are also consistently small, indicating that the proposed model is not only accurate but also stable across repeated runs.

A more important observation is that the advantage of LoG-ADFormer becomes clearer as the prediction horizon increases. At T=720, it yields the lowest RMSE on all four datasets, with relative reductions of 4.5% on Tongji, 12.9% on Vehicle, 2.5% on MIT, and 2.3% on HUST compared with the best baseline in each case. This trend suggests that the proposed local-to-global representation and Abs-Delta fusion are effective in suppressing long-range error accumulation and trajectory-level drift.

The dataset-wise behavior further supports this interpretation. On Tongji, LoG-ADFormer consistently outperforms all baselines across all horizons, showing strong robustness under complex degradation patterns. On Vehicle, its advantage is less pronounced at shorter horizons, where PatchTST remains slightly competitive, but becomes substantial at T=336 and T=720, indicating stronger long-range extrapolation under distribution shifts. On MIT and HUST, the proposed method also maintains highly competitive or best performance, suggesting that its benefits are not restricted to a specific chemistry or operating regime. Overall, these results confirm that LoG-ADFormer is suited to application scenarios where long-horizon stability matters more than short-range fitting alone.

D. Ablation on Abs-Delta Forecasting Head

Table III shows that the full model achieves the best overall balance between accuracy and stability, confirming that the Abs and Delta branches are complementary rather than redundant. In most settings, removing either branch leads to clear performance degradation, especially in RMSE, which is more sensitive to long-horizon trajectory deviation.

The impact of removing the Abs branch becomes more evident as the forecasting horizon increases. Without explicit

TABLE III.
RESULTS OF ABLATION STUDY ON FOUR DATASETS

Models		FULL		w/o Abs		w/o Delta	
Metric		MAE	RMSE	MAE	RMSE	MAE	RMSE
Tongji	96	**0.2838 ±0.0160**	**0.6237 ±0.0195**	0.3272 ±0.0140	0.6731 ±0.0224	0.3431 ±0.0264	0.7146 ±0.0309
	192	**0.4157 ±0.0109**	**0.7730 ±0.0101**	0.4754 ±0.0214	0.7811 ±0.0294	0.4562 ±0.0194	0.8288 ±0.0265
	336	0.5879 ±0.0300	**0.9239 ±0.0362**	**0.5482 ±0.0118**	0.9474 ±0.0142	0.6689 ±0.0422	1.0503 ±0.0556
	720	**0.7125 ±0.0268**	**0.9686 ±0.0333**	0.7481 ±0.0191	0.9776 ±0.0115	0.7549 ±0.0133	1.0325 ±0.0191
Vehicle	96	0.1569 ±0.0258	**0.4287 ±0.0649**	**0.1255 ±0.0118**	0.4413 ±0.0316	0.1544 ±0.0213	0.4706 ±0.0540
	192	**0.3057 ±0.0317**	**0.5732 ±0.0794**	0.3149 ±0.0205	0.7601 ±0.0410	0.3102 ±0.0476	0.7279 ±0.0939
	336	**0.4137 ±0.0491**	**0.7023 ±0.0863**	0.4359 ±0.0614	0.8553 ±0.1082	0.4293 ±0.0587	0.7810 ±0.1011
	720	**0.5739 ±0.0215**	**0.8267 ±0.0267**	0.5905 ±0.0293	0.8499 ±0.0316	0.5836 ±0.0212	0.8411 ±0.0323
MIT	96	0.3258 ±0.0149	0.6725 ±0.0264	**0.2721 ±0.0216**	**0.6558 ±0.0359**	0.3424 ±0.0203	0.6974 ±0.0390
	192	**0.4729 ±0.0120**	**0.7967 ±0.0180**	0.5012 ±0.0315	0.9397 ±0.0440	0.5153 ±0.0307	0.9080 ±0.0418
	336	**0.6285 ±0.0427**	**0.9673 ±0.0427**	0.6518 ±0.0160	1.0917 ±0.0317	0.6529 ±0.0337	1.0548 ±0.0352
	720	**0.7238 ±0.0271**	**1.0795 ±0.0152**	0.7625 ±0.0379	1.1555 ±0.0401	0.7547 ±0.0153	1.1318 ±0.0182
HUST	96	**0.0052 ±0.0009**	0.0113 ±0.0031	0.0065 ±0.0003	0.0116 ±0.0012	0.0053 ±0.0013	**0.0103 ±0.0025**
	192	**0.0080 ±0.0003**	0.0155 ±0.0027	0.0083 ±0.0008	**0.0125 ±0.0021**	0.0085 ±0.0010	0.0166 ±0.0042
	336	**0.0089 ±0.0023**	0.0153 ±0.0046	0.0101 ±0.0006	**0.0148 ±0.0011**	0.0097 ±0.0034	0.0158 ±0.0058
	720	**0.0103 ±0.0009**	**0.0158 ±0.0014**	0.0136 ±0.0016	0.0201 ±0.0028	0.0112 ±0.0003	0.0161 ±0.0009

absolute-level calibration, the model relies only on accumulated increments, making it more vulnerable to bias accumulation over future steps. This results in less stable long-range forecasts and larger drift, as reflected by the increased RMSE at longer horizons. These results indicate that the Abs branch plays a key role in anchoring the global trajectory and correcting long-range offsets. By contrast, removing the Delta branch mainly weakens the modeling of local temporal evolution. Although direct level prediction can still capture coarse degradation trends, it is less effective at representing step-wise transitions and short-term recovery-like variations. As a result, the forecasts are less consistent with local dynamics and often show inferior accuracy.

Overall, although a few settings show only small differences, the general trend is clear: the complete Abs-Delta design delivers more robust performance across datasets and horizons, and its advantage becomes particularly evident under more challenging long-horizon forecasting scenarios.

E. Parameter Sensitivity

Fig. 2 analyzes the sensitivity of LoG-ADFormer to the lookback length L on four datasets under T=96 and T=720. Overall, increasing L generally improves performance, and this effect is much clearer at the longer horizon, indicating that sufficient history is important for identifying degradation stages and stabilizing long-range extrapolation. For T=720, RMSE usually decreases as L increases, showing that long-horizon forecasting benefits from richer temporal context and multi-scale degradation cues. For T=96, however, the trend is less monotonic on some datasets. This suggests that short-horizon prediction depends more strongly on recent local dynamics, while excessively long input histories may introduce information that is less relevant for near-future forecasting. Consequently, L=336 is adopted as a robust default that balances accuracy and efficiency.

V. Conclusions

Reliable long-horizon forecasting is essential for battery maintenance but is often limited by cumulative errors and trajectory instability. To address this, LoG-ADFormer is proposed, which combines hierarchical local-to-global representation learning with a drift-resistant Abs-Delta head for stable long-term prediction. Experiments on four public datasets show that it consistently ranks among the top-2 methods, with clearer advantages at longer horizons. Looking ahead, the reliable long-horizon forecasts of LoG-ADFormer may facilitate reliability-aware battery management by supporting risk-sensitive early warning and maintenance scheduling in real-world battery systems.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (NSFC) under Grant No.92270124.

References

[1] F. Ahmadzadeh and J. Lundberg, “Remaining useful life estimation”, International Journal of System Assurance Engineering and Management, 2014, 5, (4), pp. 461-474.

[2] K. A. Severson, P. M. Attia, N. Jin, N. Perkins, B. Jiang, Z. Yang, M. H. Chen, M. Aykol, P. K. Herring, D. Fraggedakis, M. Z. Bazant, S. J. Harris, W. C. Chueh, and R. D. Braatz, “Data-driven prediction of battery cycle life before capacity degradation”, Nature Energy, 2019, 4, (5), pp. 383-391.

[3] M. Chen, G. Ma, W. Liu, N. Zeng, and X. Luo, “An overview of data-driven battery health estimation technology for battery management system”, Neurocomputing, 2023, 532, pp. 152-169.

[4] W. Wu, Z. Chen, W. Liu, D. Zhou, T. Xia, and E. Pan, “Battery health prognosis in data-deficient practical scenarios via reconstructed voltage-based machine learning”, Cell Reports Physical Science, 2025, 6, pp. 102442.

[5] Z. Fan, Y. Guo, L. Song, J. Du, H. Wang, and X. Li, “BatteryTSFM: Generalizable long-horizon degradation prediction across conditions and chemistries via time series foundation models”, Energy and AI, 2025, 22, pp. 100646.

[6] K. Cho, B. van Merriënboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, Oct. 25–29, 2014, pp. 1724–1734.

[7] A. Zeng, M. Chen, L. Zhang, and Q. Xu, “Are transformers effective for time series forecasting?” in Thirty-Seventh AAAI Conference on Artificial Intelligence, 2023, pp. 11121–11128.

[8] Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, “A time series is worth 64 words: Long-term forecasting with transformers,” in The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023, 2023.

[9] Y. Zhang and J. Yan, “Crossformer: Transformer utilizing cross-dimension dependency for multivariate time series forecasting,” in *The Eleventh International Conference on Learning Representations, ICLR 2023*, Kigali, Rwanda, May 1–5, 2023, 2023.

[10] Y. Liu, T. Hu, H. Zhang, H. Wu, S. Wang, L. Ma, and M. Long, “iTransformer: Inverted transformers are effective for time series forecasting,” in *The Twelfth International Conference on Learning Representations, ICLR 2024*, Vienna, Austria, May 7–11, 2024, 2024.

[11] H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, “TimesNet: Temporal 2D-Variation Modeling for General Time Series Analysis,” in *The Eleventh International Conference on Learning Representations, ICLR 2023*, Kigali, Rwanda, May 1–5, 2023, 2023.

[12] S. Wang, H. Wu, X. Shi, T. Hu, H. Luo, L. Ma, J. Y. Zhang, and J. Zhou, “TimeMixer: Decomposable Multiscale Mixing for Time Series Forecasting,” in *The Twelfth International Conference on Learning Representations, ICLR 2024*, Vienna, Austria, May 7–11, 2024, 2024.

[13] S. Wang, J. Li, X. Shi, Z. Ye, B. Mo, W. Lin, S. Ju, Z. Chu, and M. Jin, “TimeMixer++: A General Time Series Pattern Machine for Universal Predictive Analysis,” in *The Thirteenth International Conference on Learning Representations, ICLR 2025*, Singapore, Apr. 24–28, 2025, 2025.

[14] W. Yue, Y. Liu, X. Ying, B. Xing, R. Guo, and J. Shi, “FreEformer: Frequency Enhanced Transformer for Multivariate Time Series Forecasting,” in *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)*, 2025, pp. 3606–3614.

[15] W. Tang, J. Chen, and D. Chen, “Predicting EV battery state of health using long short term degradation feature extraction and FEA TimeMixer”, Scientific Reports, 2025, 15, (1), pp. 2200.

[16] R. Tan, X. Lu, M. Cheng, J. Li, J. Huang, and T.-Y. Zhang, “Forecasting battery degradation trajectory under domain shift with domain generalization”, Energy Storage Materials, 2024, 72, pp. 103725.

[17] X. Liu, C. Yang, Y. Meng, J. Zhu, and Y. Duan, “Capacity estimation of Li-ion battery based on transformer-adversarial discriminative domain adaptation”, AIP Advances, 2023, 13, (7).

[18] M. Wei, S. Yuan, L. Chen, Y. Liu, and J. Hu, “State of Health Prediction of Pure Electric Vehicle Batteries Based on Patch Wavelet Transformer”, Mathematics, 2025, 13, (13), pp. 2099.

[19] J. Hu, P. Fu, Z. Wei, Y. Huang, J. Early, A. Fly, and Y. Zhang, “Early prediction of lithium-ion battery degradation with a generative pre-trained transformer”, Nature Communications, 2025, 17, (1), pp. 126.

[20] J. Zhu, Y. Wang, Y. Huang, R. Bhushan Gopaluni, Y. Cao, M. Heere, M. J. Mühlbauer, L. Mereacre, H. Dai, X. Liu, A. Senyshyn, X. Wei, M. Knapp, and H. Ehrenberg, “Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation”, Nature Communications, 2022, 13, (1), pp. 2261.

[21] Z. Deng, L. Xu, H. Liu, X. Hu, Z. Duan, and Y. Xu, “Prognostics of battery capacity based on charging data and data-driven methods for on-road vehicles”, Applied Energy, 2023, 339, pp. 120954.

[22] G. Ma, S. Xu, B. Jiang, C. Cheng, X. Yang, Y. Shen, T. Yang, Y. Huang, H. Ding, and Y. Yuan, “Real-time personalized health status prediction of lithium-ion batteries using deep transfer learning”, Energy & Environmental Science, 2022, 15, (10), pp. 4083-4094.

[23] H. Wu, J. Xu, J. Wang, and M. Long, “Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting”, Advances in neural information processing systems, 2021, 34, pp. 22419-22430.

[24] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting”, Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35, (12), pp. 11106-11115.

[25] L. Liu, J. Huang, H. Zhao, T. Li, and B. Li, “PatchFormer: A novel patch-based transformer for accurate remaining useful life prediction of lithium-ion batteries”, Journal of Power Sources, 2025, 631, pp. 2361.