Electronic supplementary materials

For https://doi.org/10.1631/jzus.A2500337

Real-time degradation modeling for automotive PEMFC stacks: a multi-scale fusion network validated on an industrial 215-channel system

Zifei WANG[1,2], Xiangxian ZHU[3], Congxin LI[4], Daidai CHEN[3], Zhitao LIU[2], Longhua MA[1], Jili TAO[1], Hongye SU[2]

[1]*School of Information Science and Engineering, NingboTech University, Ningbo 315100, China*
[2]*State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou 310027, China*
[3]*Ningbo Joyson Electronic Corp., Ningbo 315040, China*
[4]*Ningbo Green Power Hydrogen Technology Research Institute Co., Ltd., Ningbo 315033, China*

S1 Experimental system details

The experimental platform is an industrial-grade gas–heat–electricity (GHE) co-simulation system developed for accelerated degradation testing of large-scale automotive PEMFC stacks. The system is centered on a 215-channel PEMFC stack with a rated power of 95 kW, designed to reproduce realistic vehicle operating conditions under controlled laboratory environments. A schematic overview of the test platform is shown in Fig. S1.

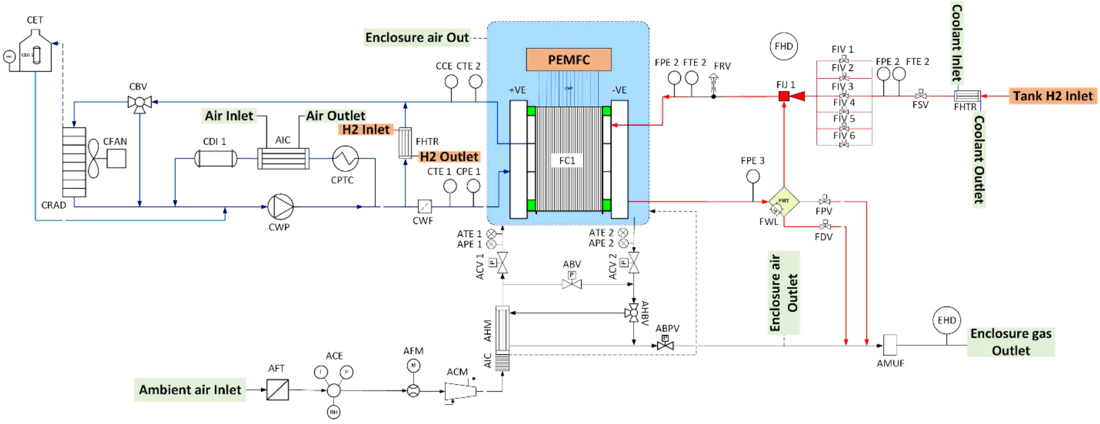

Fig. S1 Structure of the test platform

The PEMFC stack physically consists of 430 individual fuel cells. For monitoring and data acquisition purposes, the cells are electrically arranged into 215 series-connected pairs, with each pair treated as one voltage monitoring channel. This configuration enables simultaneous acquisition of per-channel voltage signals while maintaining compatibility with industrial stack architectures. The platform integrates multiple subsystems to enable coupled electro-thermal-fluidic control. Independent hydrogen and air supply lines are employed to regulate reactant pressure, flow rate, and humidity under dynamic operating conditions. A distributed thermal management system, incorporating coolant circulation, heat exchangers, and closed-loop temperature control, ensures stable stack operating temperatures during long-term aging tests. Electrical loading is provided by a high-performance programmable electronic load capable of delivering dynamic current profiles up to 500 A, allowing accurate reproduction of automotive load transients. A dedicated data acquisition system synchronously records electrical, thermal, and fluidic variables, including per-channel voltage, stack current, gas pressure, temperature, humidity, and flow rate. The system achieves a spatio-temporal synchronization accuracy better than 5 ms across all sensor channels, ensuring reliable correlation analysis among multi-physical signals. Key system components and sensor specifications are summarized in Table S1. In addition to the system configuration, the temporal evolution and statistical characteristics of the collected voltage data are summarized to provide context for subsequent modeling and validation. Fig. S2 illustrates the overall stack voltage profile over the full 327 h aging test, together with the partitioning of training and testing datasets. Fig. S3 presents the statistical distribution of voltage values across the dataset, highlighting the non-stationary characteristics induced by accelerated aging and dynamic operating conditions.

This experimental platform provides a reliable and scalable foundation for investigating degradation mechanisms and validating data-driven prediction models under industrially relevant automotive operating conditions.

Table S1 PEMFC System Components and Sensors Glossary

Name	Description	Name	Description
CET	Compressor Exit Temperature	CWL	Coolant Water Level
CRAD	Coolant Radiator	CFAN	Cooling Fan
CBV	Coolant Bypass Valve	CDI	Condensate Drain Indicator
AIC	Air Inlet Controller	CWP	Coolant Water Pump
CPTC	Compressor Temperature Controller	FHTR	Fuel Heater
CCE	Cathode Coolant Exchanger	CTE	Coolant Temperature
CPE	Coolant Pressure	AFT	Air Flow Transmitter
ACE	Air Control Valve	AFM	Air Flow Meter
ACM	Air Control Module	AHM	Air Humidity Sensor
ATE	Air Temperature	APE	Air Pressure
ABV	Air Bypass Valve	ACV	Air Control Valve
AHBV	Air Humidifier Bypass Valve	ABPV	Air Back Pressure Valve
FPE	Fuel Pressure	FTE	Fuel Temperature
FRV	Fuel Regulator Valve	FWL	Fuel Water Separator
FPV	Fuel Pressure Valve	FDV	Fuel Drain Valve
AMUF	Air Muffler	EHD	Exhaust Heat Exchanger
FIV	Fuel Inlet Valve	FSV	Fuel Shut-off Valve

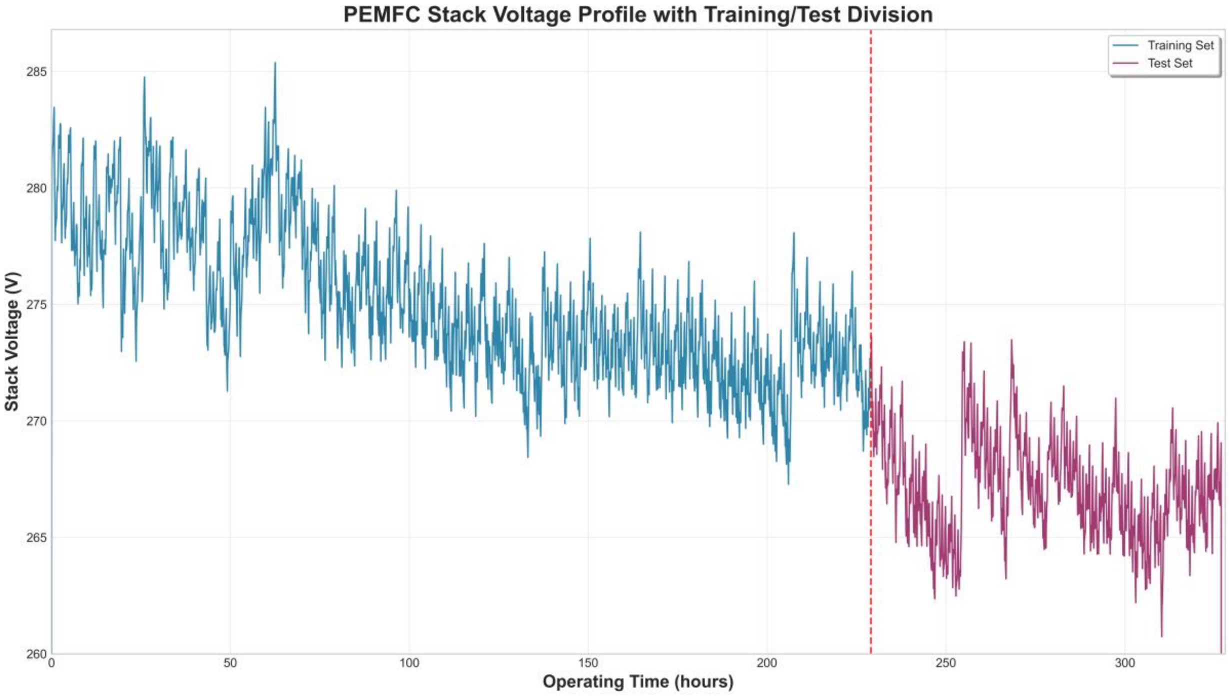

Fig. S2 Overall voltage profile and division

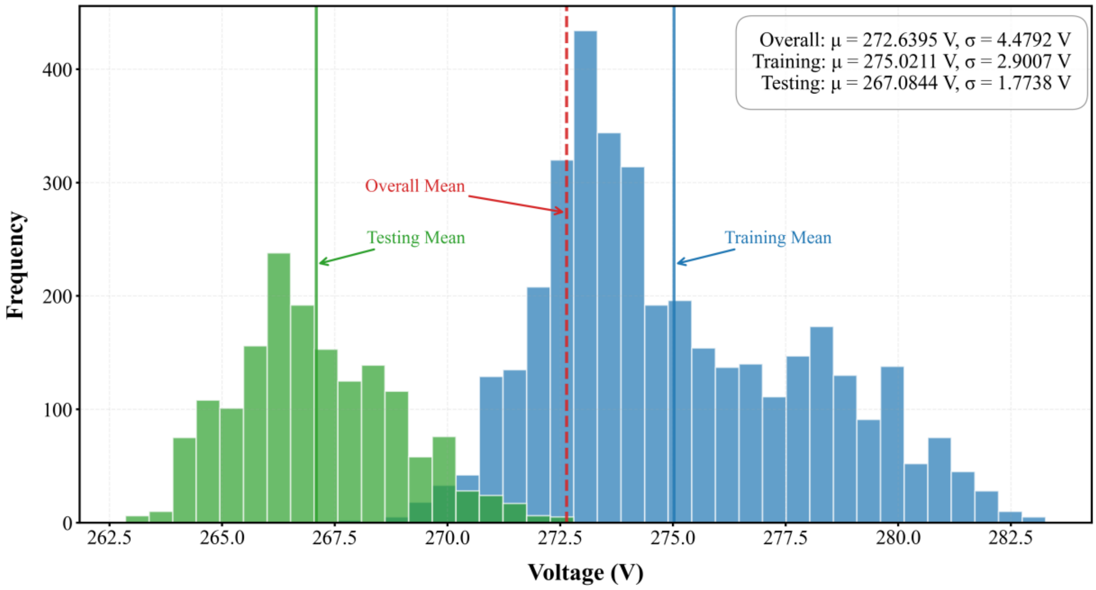

Fig. S3 Voltage distribution histogram

Additional illustrations of the industrial test platform, onsite experimental configuration, and voltage degradation trend analysis are provided in Fig. S4–S6.

Fig. S4 Industrial 215-channel system

Fig. S5 Onsite testing equipment

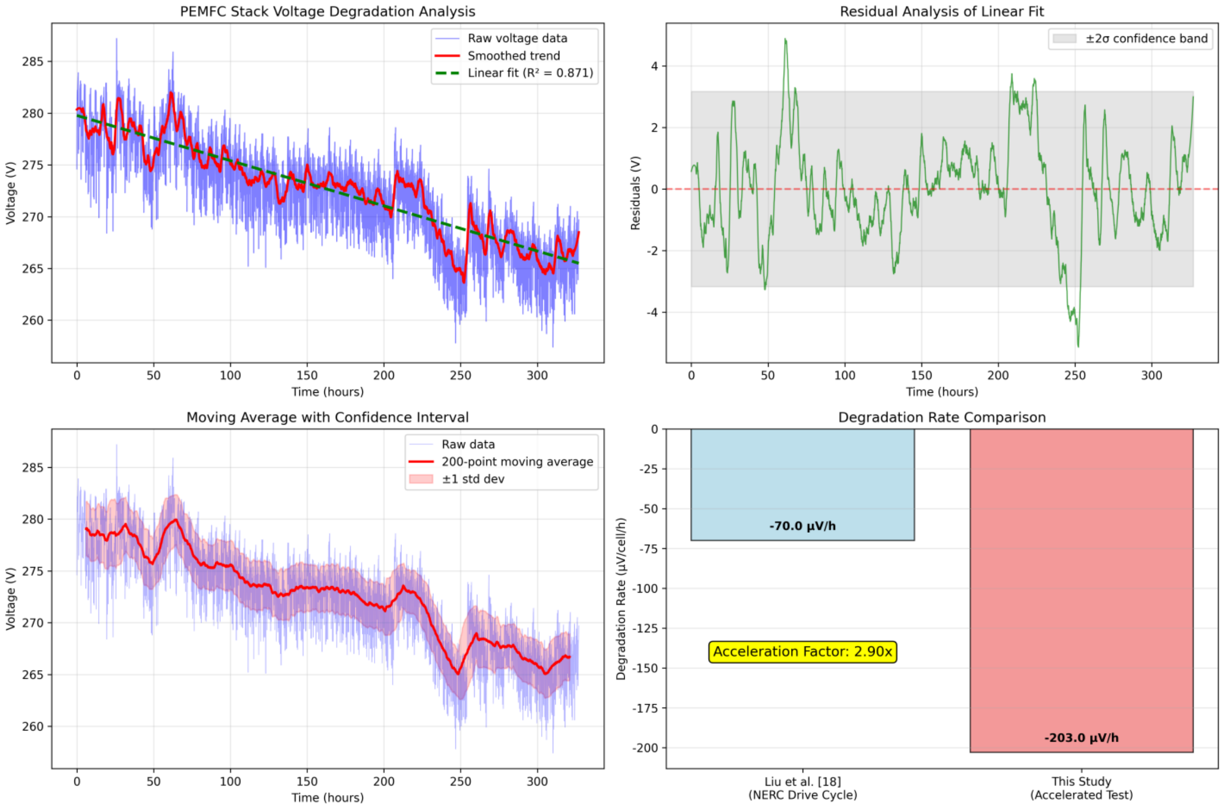

Fig. S6 Local fitting results for voltage degradation

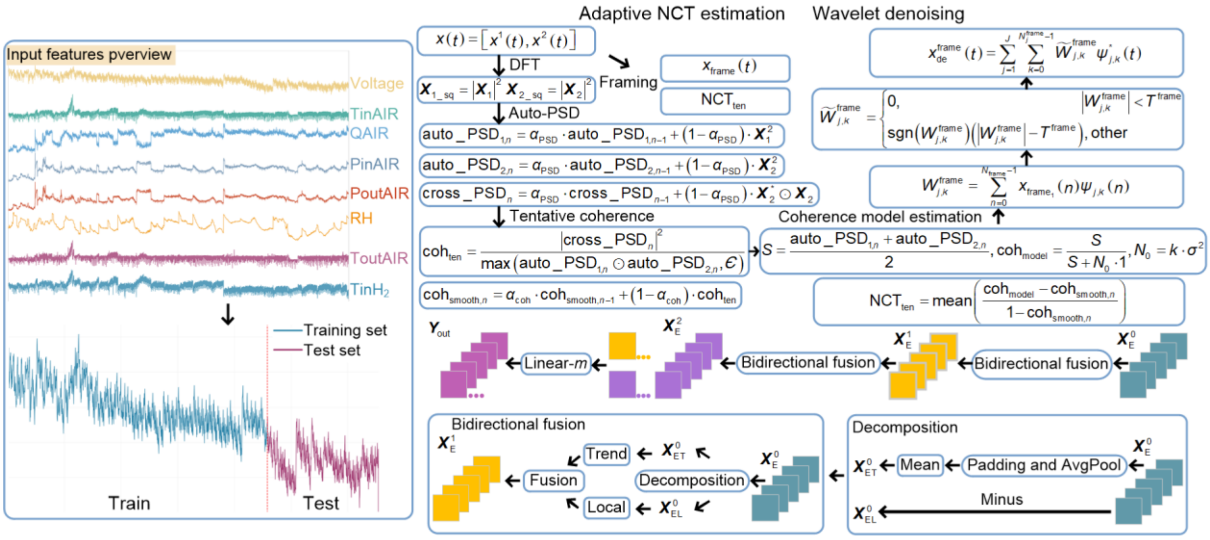

Fig. S7 **Flowchart of the proposed multi-scale bidirectional fusion network**

The proposed MBFNet is illustrated in Fig. S7, comprising: (1) a channel-joint adaptive NCT algorithm for coupled noise suppression, (2) multi-scale decomposition for disentangling transient/long-term features, and (3) parameter-efficient bidirectional fusion reconciling stack-level trends with cell-level degradation.

S2 NCT mathematical derivation

This section provides the detailed mathematical formulation of the channel-joint adaptive Noise Correlation Threshold (NCT) algorithm used in the MBFNet framework for coupled noise suppression.

S2.1 Dual-channel spectral estimation

The dual-channel signals are processed in a frame-wise manner to capture transient noise characteristics. Let the dual-channel signal be:

$$\boldsymbol{x}(\boldsymbol{t}) = [\boldsymbol{x}^{\mathbf{1}}(\boldsymbol{t}), \boldsymbol{x}^{\mathbf{2}}(\boldsymbol{t})] \in \mathbb{R}^{2\times T} \#(S1)$$

which is segmented into frames $\boldsymbol{x}_{\mathbf{frame}}(\boldsymbol{t}) \in \mathbb{R}^{2\times T_f}$. The two channels represent physically correlated sensor signals, such as inlet/outlet pressures or co-located temperature measurements. After discrete Fourier transform (DFT), the spectral coefficients are denoted as $\boldsymbol{X}_{\mathbf{1}} \in \mathbb{R}^{T_f}$ and $\boldsymbol{X}_{\mathbf{2}} \in \mathbb{R}^{T_f}$, with corresponding power spectra $|\boldsymbol{X}_{\mathbf{1}}|^{\mathbf{2}}$ and $|\boldsymbol{X}_{\mathbf{2}}|^{\mathbf{2}}$. To balance estimation stability and transient responsiveness, exponentially smoothed auto- and cross-power spectral densities (PSD) are computed as:

$$\boldsymbol{auto\_PSD}_{\mathbf{1},\boldsymbol{n}} = \alpha_{\mathrm{PSD}} \cdot \boldsymbol{auto\_PSD1}_{\boldsymbol{n-1}} + (1-\alpha_{\mathrm{PSD}}) \cdot \boldsymbol{X}_{\mathbf{1}}^{\mathbf{2}} \#(S2)$$
$$\boldsymbol{auto\_PSD}_{\mathbf{2},\boldsymbol{n}} = \alpha_{\mathrm{PSD}} \cdot \boldsymbol{auto\_PSD2}_{\boldsymbol{n-1}} + (1-\alpha_{\mathrm{PSD}}) \cdot \boldsymbol{X}_{\mathbf{2}}^{\mathbf{2}} \#(S3)$$
$$\boldsymbol{cross\_PSD}_{\boldsymbol{n}} = \alpha_{\mathrm{PSD}} \cdot \boldsymbol{cross\_PSD}_{\boldsymbol{n-1}} + (1-\alpha_{\mathrm{PSD}}) \cdot \boldsymbol{X}_{\mathbf{1}}^{*} \odot \boldsymbol{X}_{\mathbf{2}} \#(S4)$$

S2.2 Coherence modeling and tentative NCT estimation

The magnitude-squared coherence is estimated as:

$$coh_{\mathrm{ten}} = \frac{|\boldsymbol{cross\_PSD_n}|^2}{max(\boldsymbol{auto\_PSD_{1,n}} \odot \boldsymbol{auto\_PSD_{2,n}}, \epsilon)} \#(S5)$$

where $\epsilon = 10^{-4}$ avoids numerical instability. Temporal propagation of coherence under transient conditions is captured via exponential smoothing:

$$coh_{\mathrm{smooth},n} = \alpha_{\mathrm{coh}} \cdot coh_{\mathrm{smooth},n-1} + (1 - \alpha_{\mathrm{coh}}) \cdot coh_{\mathrm{ten}} \#(S6)$$

The channel-averaged PSD is defined as:

$$\boldsymbol{S} = (\boldsymbol{auto\_PSD_{1,n}} + \boldsymbol{auto\_PSD_{2,n}})/2 \#(S7)$$

To compensate for load-induced noise amplification, an adaptive noise term $N_0 = k \cdot \sigma^2$ is introduced, yielding the modeled coherence:

$$\boldsymbol{coh}_{\mathbf{model}} = \frac{\boldsymbol{S}}{\boldsymbol{S} + N_0 \cdot \mathbf{1}} \#(S8)$$

where σ^2 denotes the short-term variance of the framed signal, estimated in the time domain. The tentative noise correlation threshold is then defined as:

$$NCT_{\mathrm{ten}} = mean(\frac{\boldsymbol{coh}_{\mathbf{model}} - \boldsymbol{coh}_{\mathbf{smooth},\boldsymbol{n}}}{\mathbf{1} - \boldsymbol{coh}_{\mathbf{smooth},\boldsymbol{n}}}) \#(S9)$$

The denominator $\mathbf{1} - \boldsymbol{coh}_{\mathbf{smooth},\boldsymbol{n}}$ normalizes the coherence discrepancy by the deviation from perfect coherence in the frame, which improves numerical stability when $\boldsymbol{coh}_{\mathbf{model}}$ approaches unity and suppresses excessive threshold fluctuation under transient conditions. A larger discrepancy indicates a higher proportion of uncorrelated noise, requiring stronger suppression.

S2.3 Threshold bounding and temporal smoothing

To ensure numerical stability under low-load and high-noise conditions, NCT_{ten} is bounded using nonlinear thresholding:

$$NCT_{\mathrm{ten}} = \begin{cases} \gamma, & NCT_{ten} \leq 0 \\ thres(NCT_{\mathrm{ten}}, NCT_{\mathrm{limits}}), & \text{otherwise} \end{cases} \#(S10)$$

where γ and NCT_{limits} are empirically selected based on observed SNR ranges.
Temporal smoothing across start–stop cycles is performed as

$$NCT_{\mathrm{smooth},n} = \alpha_{\mathrm{NCT}} NCT_{\mathrm{smooth},n-1} + (1 - \alpha_{\mathrm{NCT}}) NCT_{\mathrm{ten}} \#(S11)$$

The final frame-level threshold is obtained by averaging over a selected frequency band:

$$NCT_{\mathrm{frame}} = mean\left(NCT_{\mathrm{smooth},n}, NCT_{\mu_{\mathrm{range}}}\right) \#(S12)$$

S2.4 Adaptive wavelet-domain denoising

Each signal frame is decomposed using a J-level discrete wavelet transform (The Db4 wavelet is selected to balance compact support and symmetry):

$$W_{j,k}^{\text{frame}} = \sum_{n=0}^{N_{\text{frame}}-1} x_{\text{frame}_1}(n)\psi_{j,k}(n)\#(S13)$$

A time-varying threshold $T^{\text{frame}} = \alpha \times NCT_{\text{frame}}$ is applied to perform adaptive soft-thresholding:

$$\widetilde{W}_{j,k}^{\text{frame}} = \begin{cases} 0, \ |W_{j,k}^{\text{frame}}| < T^{\text{frame}} \\ sgn(W_{j,k}^{\text{frame}})(|W_{j,k}^{\text{frame}}| - T^{\text{frame}}), \text{ other} \end{cases} \#(S14)$$

$$\boldsymbol{x}_{\mathbf{de}}^{\mathbf{frame}}(\boldsymbol{t}) = \sum_{j=1}^{J} \sum_{k=0}^{N_j^{\text{frame}}-1} \widetilde{W}_{j,k}^{\text{frame}} \psi_{j,k}^{*}(t)\#(S15)$$

The denoised signal is reconstructed via inverse wavelet transform. Both channels are processed identically, yielding the final denoised signal $\boldsymbol{x}_{\mathbf{de}}(\boldsymbol{t}) \in \mathbb{R}^{2\times T}$. The notation used in this section is summarized in Table S2.

Table S2 Notation definitions for NCT algorithm

Symbol	Description
$\boldsymbol{x}(\boldsymbol{t})$, $\boldsymbol{x}_{\mathbf{frame}}(\boldsymbol{t})$	Raw and framed dual-channel input signal
$\boldsymbol{X}_{\mathbf{1}}$, $\boldsymbol{X}_{\mathbf{2}}$	Discrete Fourier Transform (DFT) coefficients of two channels
$\boldsymbol{auto\_PSD}_{\mathbf{1},\boldsymbol{n}}$, $\boldsymbol{auto\_PSD}_{\mathbf{2},\boldsymbol{n}}$	Smoothed auto-PSD of channel 1 and 2 at frame n
$\boldsymbol{cross\_PSD}_{\boldsymbol{n}}$	Smoothed cross-PSD between the two channels
α_{PSD}, α_{coh}, α_{NCT}	Smoothing factors ($0 < \alpha < 1$) for PSD, coherence and NCT
coh_{ten}, $coh_{\text{smooth},n}$	Tentative and smoothed magnitude-squared coherence
$\boldsymbol{S}$	Channel-averaged PSD
$N_0 = k \cdot \sigma^2$	Estimated noise power
$\boldsymbol{coh}_{\mathbf{model}}$	Modeled coherence under the estimated noise
NCT_{ten}, NCT_{frame}	Tentative and final Noise Correlation Threshold
T^{frame}	Time-varying threshold applied in the wavelet domain
$W_{j,k}^{\text{frame}}$	wavelet coefficients at different scales j and positions k
$NCT_{\mu_{\text{range}}}$	Selected frequency band for NCT computation

S3 Additional preprocessing details

This section provides supplementary details on the data preprocessing procedures that support the main modeling framework, with a particular focus on feature selection based on mutual information (MI). These steps are included here to improve reproducibility while keeping the main text concise. Given two random variables X and Y, the mutual information is defined as:

$$MI(X;Y) = \sum_{y\in Y} \sum_{x\in X} p(x,y)\, log\left(\frac{p(x,y)}{p(x)p(y)}\right)\#(S16)$$

where $p(x, y)$ denotes the joint probability distribution of X and Y, and $p(x)$ and $p(y)$ are their marginal distributions, respectively. MI measures the reduction in uncertainty of one variable given knowledge of the other and can capture both linear and nonlinear dependencies. In this study, X represents each candidate sensor variable (e.g., temperature, pressure, humidity, and flow rate), and Y corresponds to the stack voltage. MI values were estimated using histogram-based probability density estimation with adaptive binning to ensure numerical stability under non-stationary operating conditions.

A threshold of $\mathrm{MI} > 0.15$ was adopted to select degradation-relevant variables. This threshold was chosen empirically to balance information retention and dimensionality reduction. Variables below this threshold exhibited weak or unstable dependency with voltage and were excluded from subsequent modeling. As a result, seven key variables were retained, achieving approximately 42% reduction in input dimensionality while preserving physically meaningful degradation-related information.

S4 MBFNet Architecture Details

This section provides detailed descriptions of the baseline model selection principles and the complete architectural configuration of the proposed MBFNet, ensuring reproducibility while maintaining a concise main text.

S4.1 Baseline Model Selection Principles

The selection of baseline models follows three fundamental principles aligned with the requirements of industrial-grade PEMFC degradation prediction.

(1) Representative temporal modeling paradigms. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are adopted as canonical recurrent architectures for sequential dependency learning (Hochreiter and Schmidhuber, 1997; Chung et al., 2014). Their gating mechanisms enable effective modeling of electrochemical dynamics with characteristic time constants ranging from seconds to hours. In addition, one-dimensional convolutional neural networks (1D-CNN) are included to complement recurrent models. Depthwise separable convolutions with kernel sizes $\{5, 10, 20\}$ are employed to extract local voltage fluctuation patterns at different temporal bandwidths, which have been shown to correlate with degradation behavior (Jouin et al., 2014; Chollet, 2017).

(2) Attention mechanism for temporal importance weighting. To evaluate the effect of adaptive temporal focus, attention-enhanced LSTM/GRU variants are considered. The attention weights are computed as:

$$\alpha_t^{(i)} = softmax\left(\boldsymbol{W_q} h_t^{(i)} \cdot \boldsymbol{W_k} H_{1:T}^{(i)}\right) \#(S17)$$

which enables the model to emphasize critical voltage segments associated with accelerated degradation phases (Ren et al., 2022).

(3) Automotive deployment constraints. Considering onboard computational limitations, memory usage is explicitly constrained according to

$$Memory = \frac{4 \times Params}{10^6} + \frac{Activates \times Batch}{256} < 4000 \#(S18)$$

where the threshold corresponds to typical memory budgets of automotive electronic control units (ECUs). This constraint ensures a fair comparison between models in terms of both prediction accuracy and deployment feasibility.

Overall, the baseline selection framework covers recurrent (LSTM/GRU) and convolutional (1D‑ CNN) temporal paradigms, incorporates attention mechanisms to assess local versus global temporal focus, and explicitly accounts for hardware constraints to balance accuracy, memory footprint, and inference latency.

S4.2 MBFNet Architectural Configuration

To address model transparency and reproducibility, the complete layer-wise configuration of MBFNet is summarized in Table S3. The architecture consists of three main components: channel-joint NCT denoising, multi-scale temporal decomposition, and bidirectional feature fusion.

Table S3 MBFNet Layer Configuration and Hyperparameters

Module	Sub-module	Layer Type	Input Dim	Output Dim	Kernel/Params
NCT Denoising	Framing	-	$2 \times T$	$2 \times T_f$	$[16, 256]$
	PSD	Exponential Smoothing	T_f	T_f	$\alpha_{PSD} = 0.8$
	Wavelet Transform	DWT (Db4)	-	-	$J = 4$
	Thresholding	Soft Threshold	-	-	$\alpha = 1.2$
Multi-scale Decomposition	Average Pooling	Pool-1D	$\mathrm{I}/2^{m} \times 8$	$I/2^{m+1} \times 8$	Kernel=2, Stride=2
	Encoder	Linear	8	128	1024
			128	64	8,192
Bidirectional Fusion	$\mathcal{F}_{trend}$	Linear	64	64	4,096
	$\mathcal{F}_{local}$		64	64	
	$\mathcal{P}_m$		128	64	8,192

S5 Supplementary experiments

This section provides additional experimental results and visual analyses that support the conclusions presented in the main text. Fig. S8-S11 present voltage trajectory comparisons for 5-step, 12-step, 24-step, and 100-step predictions, respectively.

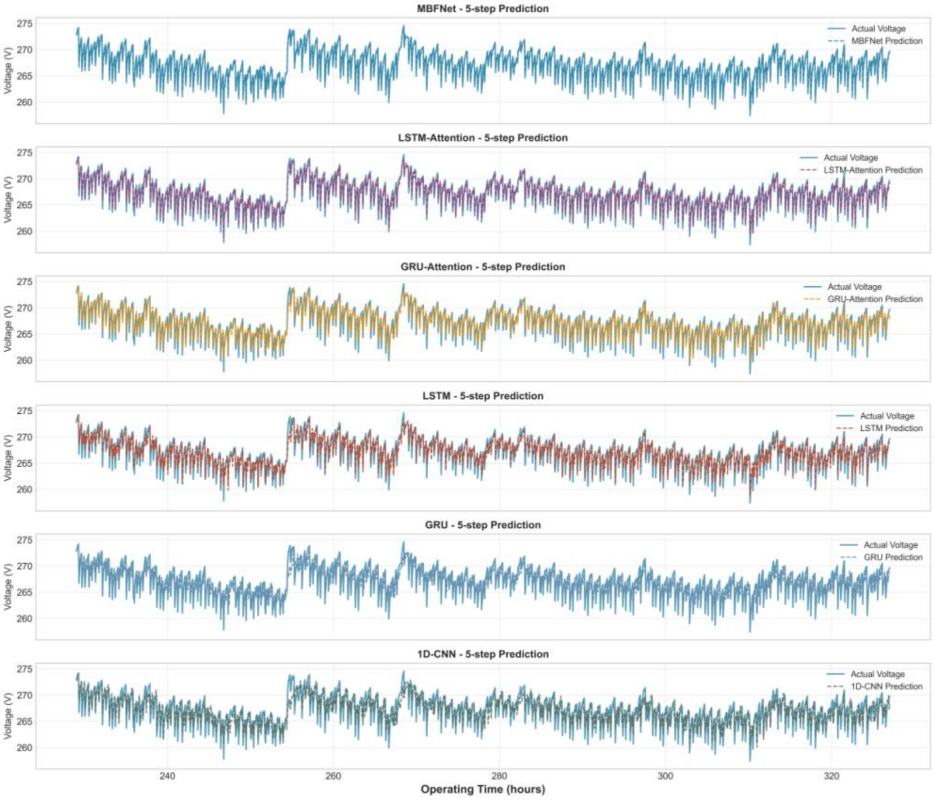

Fig. S8 5-step prediction results of each model

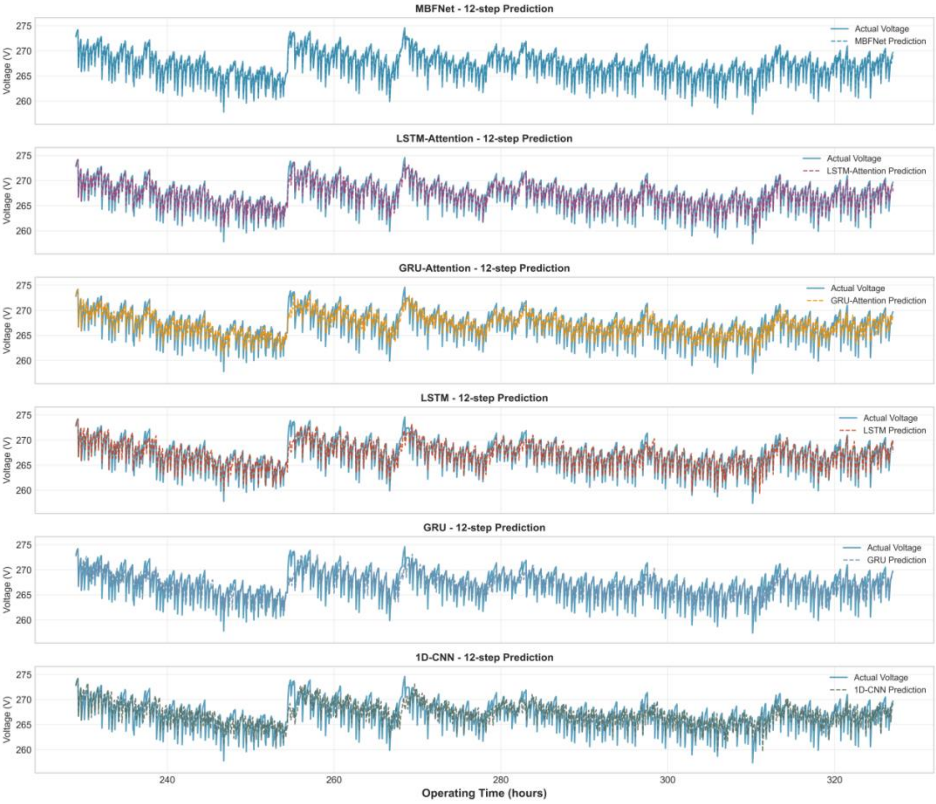

Fig. S9 12-step prediction results of each model

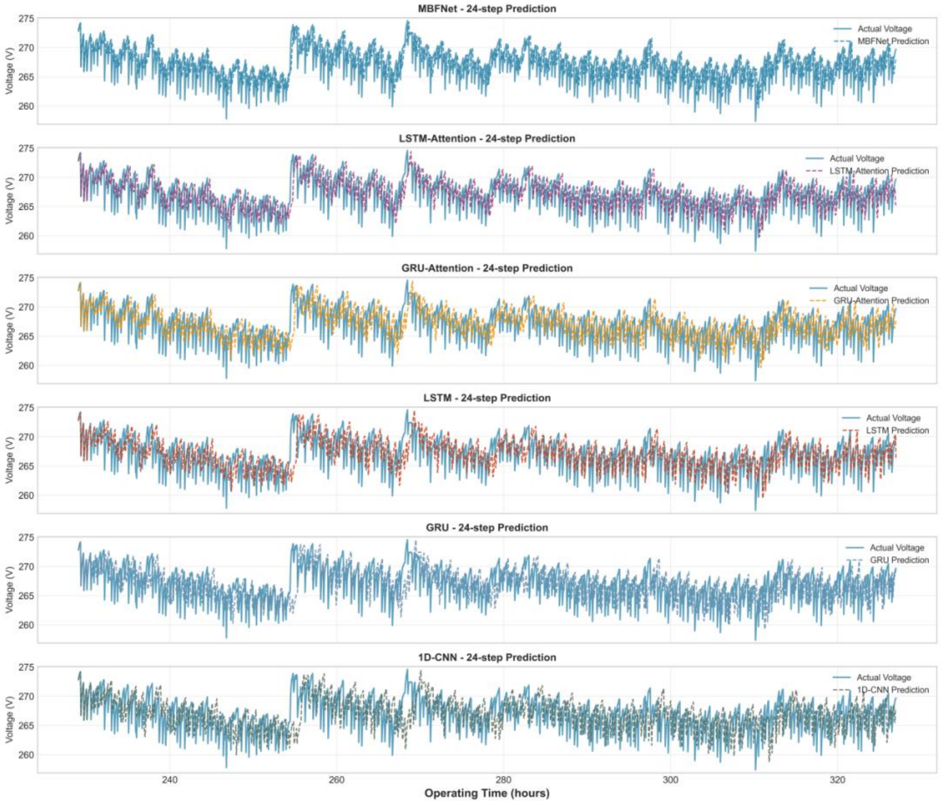

Fig. S10 24-step prediction results of each model

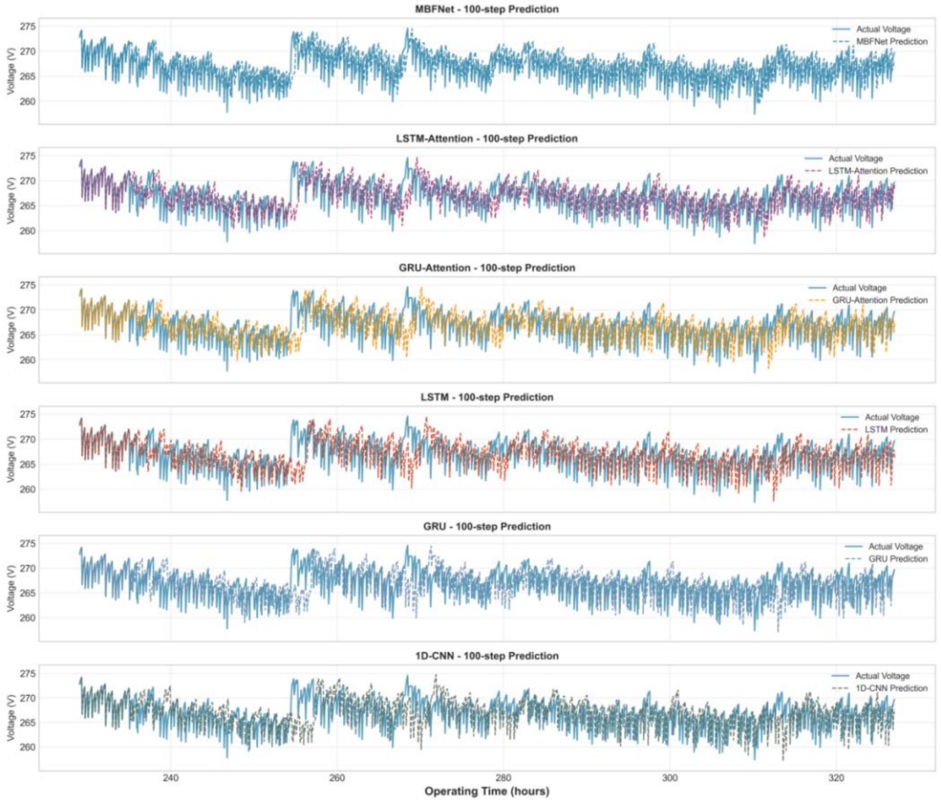

Fig. S11 100-step prediction results of each mode

Fig. S12 presents the local prediction error distribution during voltage recovery events. The results highlight the differences in error magnitude and dispersion among the compared models, illustrating the challenges associated with accurately capturing short-term recovery dynamics under non-steady-state operating conditions.

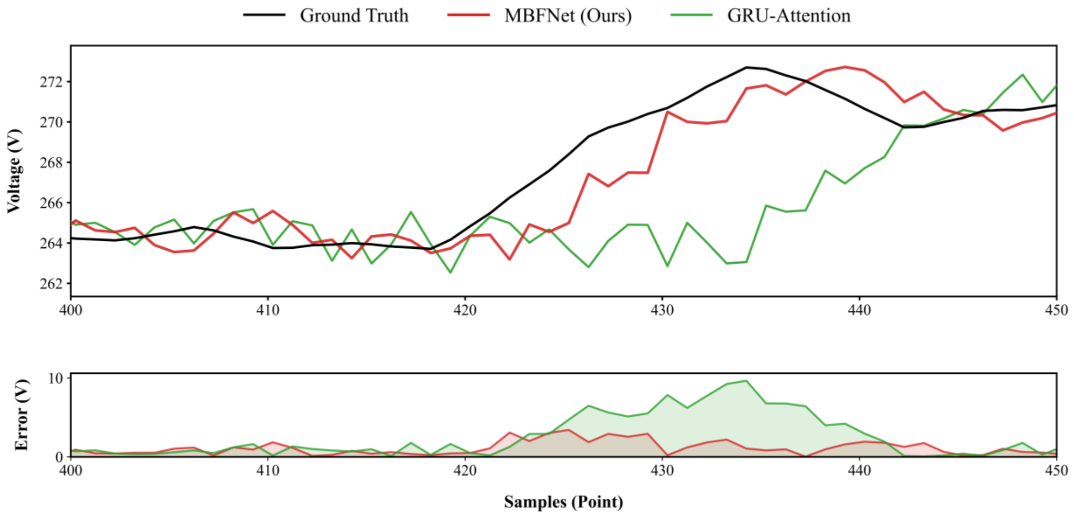

Fig. S12 Voltage recovery local error analysis

Fig. S13 compares the error distributions over a 100-second prediction window. MBFNet exhibits the smallest interquartile range ($\mathrm{IQR} = 0.659$) within the error range $[0, 3.406]$, indicating superior prediction stability during voltage recovery. In contrast, the 1D-CNN fails to reasonably capture the voltage recovery behavior, resulting in significantly larger error dispersion.

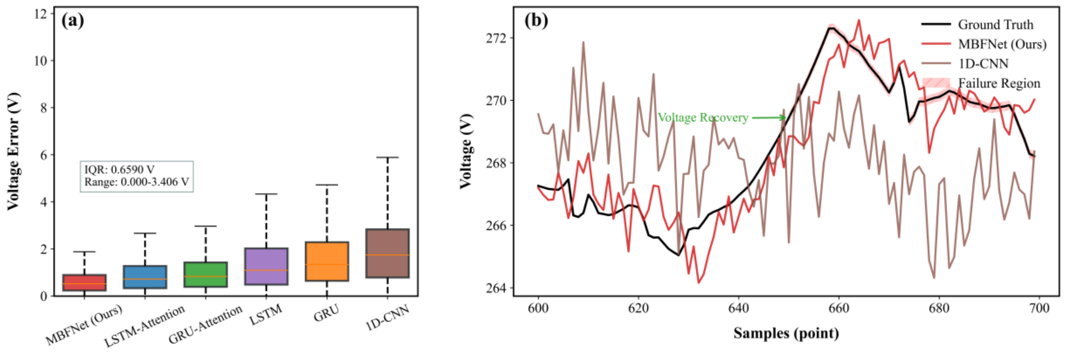

Fig. S13 (a) Voltage error boxplots (b) Voltage recovery local comparison

Fig. S14 compares the prediction accuracy and parameter count of MBFNet and its ablation variants. Removing individual submodules leads to distinct performance degradation patterns, highlighting the complementary roles of NCT denoising, multiscale decomposition, and bidirectional fusion. The results demonstrate that MBFNet achieves a favorable balance between prediction accuracy and model complexity.

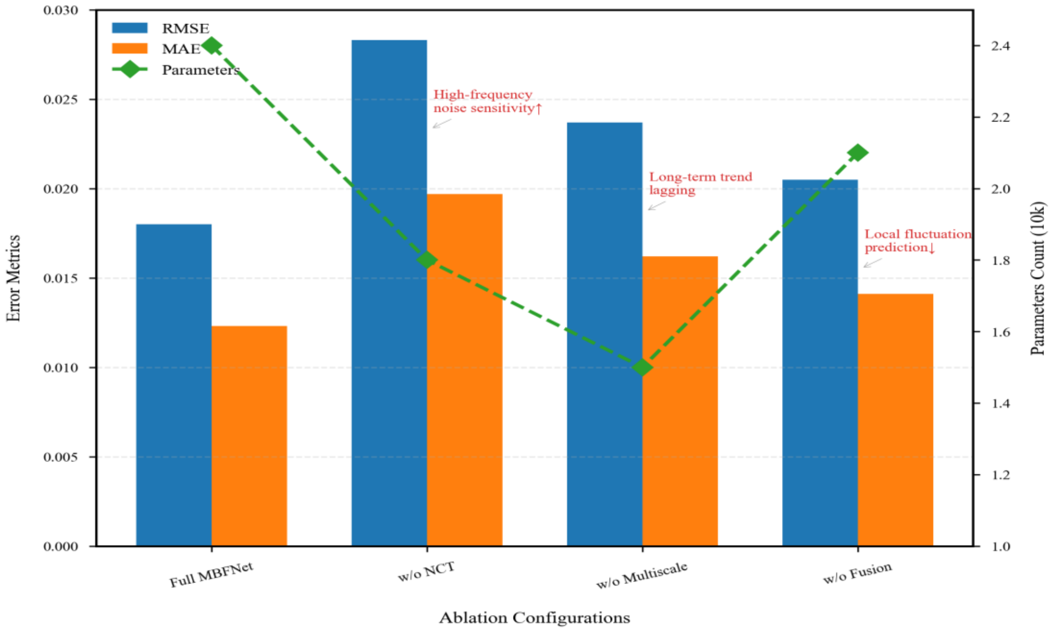

Fig. S14 Comparison of prediction errors and parameter count for ablation models

Fig. S15 illustrates the effect of the proposed NCT denoising module on raw voltage signals. After denoising, high-frequency noise and transient oscillations are effectively suppressed while the underlying degradation trend and recovery dynamics are preserved, confirming the robustness of the channel-joint adaptive noise suppression strategy.

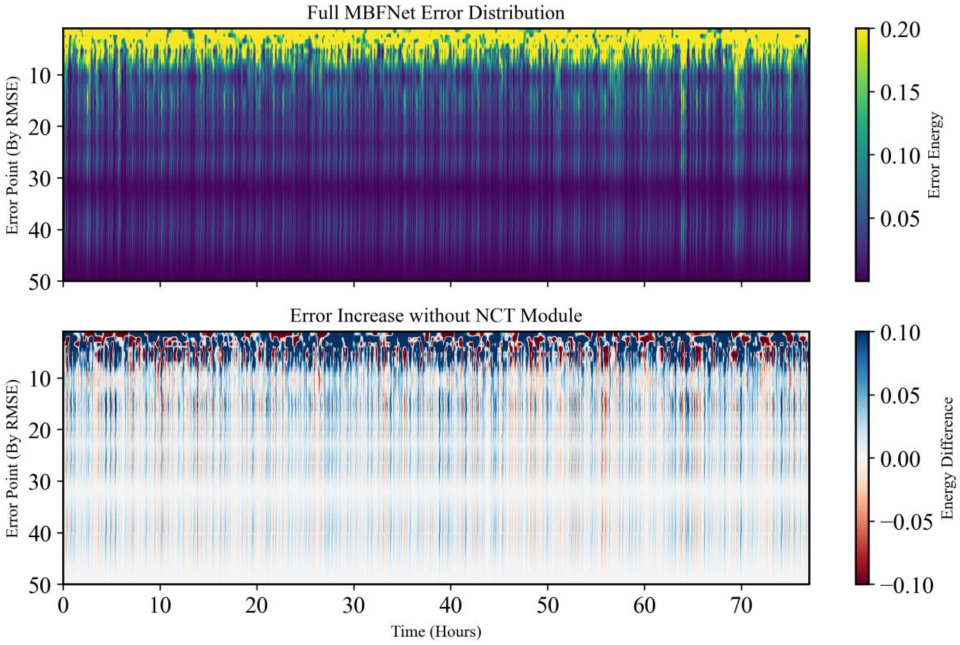

Fig. S15 Noise elimination before and after the NCT module

Fig. S16 presents the trade-off between prediction accuracy and computational resources as the number of projection layers increases. While deeper configurations yield marginal improvements in RMSE, memory usage and inference latency grow rapidly, indicating diminishing returns beyond two projection layers for edge deployment scenarios.

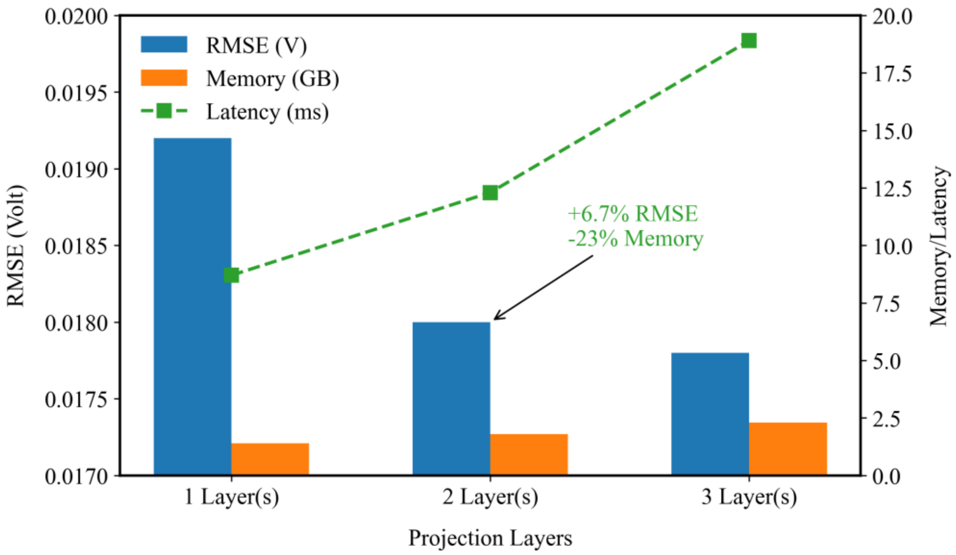

Fig. S16 Accuracy–resource trade-off under different projection layers

Fig. S17 analyzes the sensitivity of prediction performance to the number of temporal decomposition layers. Increasing the decomposition depth significantly improves long-term prediction accuracy, whereas short-term performance exhibits lower sensitivity. This result confirms the effectiveness of multiscale decomposition in capturing slow degradation processes while maintaining computational efficiency.

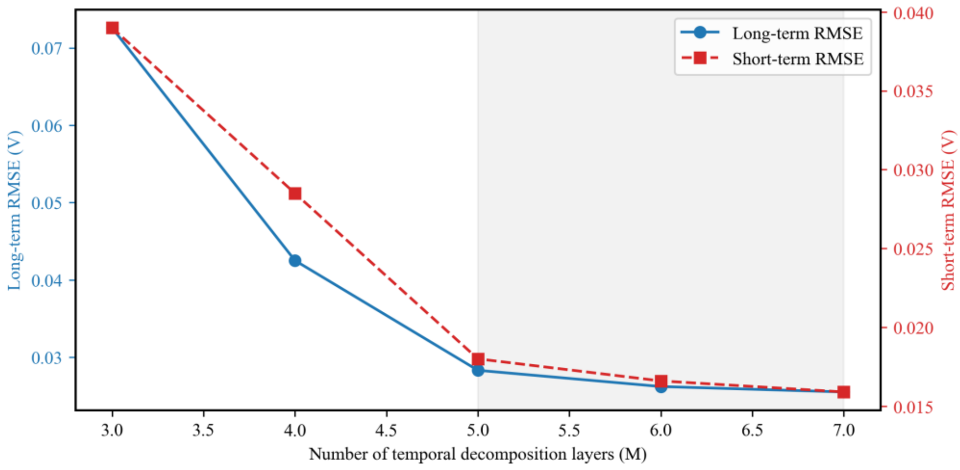

Fig. S17 Sensitivity analysis with respect to temporal decomposition depth