computers

MDPI

Article

Physics-Guided Dynamic Prediction and Intrinsic Interpretability of Substation Carbon-Emission Factors: A MOIRAI-2 and UPINN Fusion Framework

Jingbo Song [1], Chen Chen [1], Song Wang [1], Liang Zhang [1], Han Yao [1] and Tongchui Liu [2,*]

[1] State Grid Henan Economic Research Institute, Zhengzhou 450052, China; songjingbo_99@126.com (J.S.); sunshin0920@sina.com (C.C.)
[2] College of Electrical Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450045, China
* Correspondence: lbhammer@163.com

Abstract

Substation-level carbon-emission factors (CEFs) are operationally relevant because substations concentrate transformer losses, auxiliary consumption, and sulfur hexafluoride (SF6) leakage at the interface between transmission and distribution. However, static or annual emission-factor methods average over heterogeneous operating regimes and cannot capture the pronounced non-stationarity of substation CEFs driven by seasonal loads, stochastic maintenance events, cooling-system switching, and extreme weather. To support high-frequency dynamic carbon tracing, dispatch optimization, and audit compliance, this study proposes a physics-guided fusion framework integrating a temporal foundation model, MOIRAI-2, with a Uniform Physics-Informed Neural Network (UPINN). A 15-dimensional physically constrained feature vector is constructed from IEEE C57.91 thermal-circuit equations and ideal-gas state equations, including transformer top-oil/hot-spot temperature, SF6 pressure/density estimation, and oil-forced/air-forced (OFAF) or oil-directed/water-forced (ODWF) cooling status. MOIRAI-2 uses Any-Variate Attention with binary attention bias to model intra-variate temporal dependencies and cross-variate physical couplings, whereas UPINN embeds thermal-balance, SF6 leakage-kinetics, and CEF conservation residuals as soft constraints. An adaptive gating network balances data-driven pattern recognition and physics-driven smoothness across steady-state, extreme-event, and maintenance regimes. Validation on a 220 kV substation dataset achieves a mean absolute error (MAE) of 0.352 gCO_2e/kWh, outperforming random forest (RF), gradient boosting machine (GBM), long short-term memory (LSTM), and a Pure Transformer by 24.0%, 19.1%, 23.0%, and 14.4%, respectively. Ablation studies show that the 15-dimensional physical-feature expansion improves accuracy by 8.8%, whereas physics-loss regularization reduces prediction variance by 37%. UPINN decomposition further indicates that transformer total loss, ambient temperature, and load factor dominate CEF dynamics, and rainfall cooling reduces CEF by 0.04 gCO_2e/kWh per 20 mm increment. The framework provides a physically consistent and intrinsically interpretable basis for dynamic substation carbon accounting and low-carbon operation.

Keywords: carbon-emission factor; substation; physics-informed neural network; temporal foundation model; dynamic prediction; interpretable machine learning

Academic Editor: Paolo Bellavista

Received: 31 July 2026
Revised: 13 September 2026
Accepted: 14 September 2026
Published: 19 September 2026

1. Introduction

The global push for carbon neutrality has intensified decarbonization efforts worldwide. As the largest energy consumer, China's "dual-carbon" target reshapes its energy systems [1]. The power sector, contributing ~40% of global energy-related CO_2 emissions, is pivotal [2]. High-voltage substations—aggregating transformer losses, auxiliary consumption, and SF_6 leakage—represent critical nodes where carbon-emissions concentrate and propagate across the power network [3]. The concept of carbon-emission flow in networks provides a theoretical foundation for tracing these emissions from generation to demand [4]. Static annual emission factors are inadequate; dynamic, high-resolution CEF prediction is essential for real-time carbon tracing, dispatch optimization, and compliance auditing [5]. However, substation carbon dynamics involve complex multi-physics coupling and non-stationarity from seasonal loads, maintenance, and weather, demanding frameworks that ensure accuracy, physical consistency, and interpretability [6].

Carbon-emission prediction can be broadly characterized as having evolved through three methodological stages: static inventory-based accounting, data-driven statistical and machine-learning prediction, and recent physics-guided hybrid modeling that integrates domain constraints with temporal learning. These stages reflect a gradual methodological shift in the literature rather than strict chronological boundaries. The first generation employs bottom-up engineering calculations based on fuel consumption statistics and emission factor databases, exemplified by the Intergovernmental Panel on Climate Change (IPCC)'s tiered methodology [7]. While physically grounded, these approaches operate at coarse resolutions, rendering them inadequate for operational decision-making. The second generation introduced statistical models such as autoregressive integrated moving average (ARIMA), which capture temporal autocorrelation in emission trajectories [8]. However, ARIMA's linearity assumption limits its applicability to nonlinear carbon dynamics. The third generation, emerging since 2020, has witnessed the proliferation of machine-learning (ML) and deep-learning (DL) methodologies. Random forest (RF) and gradient-boosting machine (GBM) have been extensively applied to regional-scale carbon intensity forecasting [9]. Feda et al. [10] demonstrated that harmonizing extreme learning machines with metaheuristic optimization achieves superior CO_2 emission prediction accuracy. LSTM networks and transformer architectures have demonstrated superior performance in capturing long-range temporal dependencies [11]. Liu et al. [12] introduced inverted dimension attention in iTransformer, establishing new benchmarks in multivariate time-series forecasting. Time series foundation models, including Chronos [13], TimesFM [14], and MOIRAI [15], have further advanced the field by enabling zero-shot forecasting through pretraining on massive datasets. Many existing carbon prediction studies frame the task primarily as a time-series forecasting problem, relying on historical trends and statistical correlations to estimate future emissions. While effective in stationary settings, such formulations often omit explicit physical mechanisms and operational constraints, which may limit robustness under non-stationary conditions [16].

At the substation level, carbon-emission accounting traditionally addresses two boundaries: direct SF_6 leakage and indirect energy consumption emissions. Direct emissions, governed by SF_6's GWP of 23,500× CO_2, are quantified via mass balance from periodic gas density monitoring [17]. However, these methods assume quasi-steady leakage, neglecting Arrhenius-accelerated seal degradation causing emission bursts during high-temperature periods [18]. For indirect emissions, the IEEE C57.91 thermal model estimates transformer no-load and load-dependent losses as functions of temperature and load factor [19]. Yet, translating equipment-level losses to grid-level carbon-emission factors requires calibrated equivalents incorporating T&D losses, often oversimplified in existing studies [20]. Recent hybrid approaches attempt to bridge this gap. Wang et al. [21] developed data-driven carbon-emission flow models for optimal power scheduling and carbon

https://doi.org/10.3390/computers15090634

intensity control; however, these approaches typically rely on coarse statistical features lacking explicit physical constraints, yielding systematic bias under extreme operating conditions. Embedding physics-informed constraints into substation carbon prediction remains open, particularly given advances in physics-informed neural networks (PINNs) for power systems [22]. Ngo et al. [22] demonstrated that physics-informed graph neural networks improve state estimation accuracy by 9.24% via Kirchhoff's law constraints, establishing a paradigm for fusing neural architectures with electrical physics. Falas et al. [23] developed robust PINNs, maintaining fidelity under measurement uncertainty and cyber-attacks. Nadal et al. [24] achieved plug-and-play PINN integration into commercial dynamic simulation tools, replacing numerical solvers with improved efficiency while preserving transient stability. These developments collectively establish the methodological foundation for embedding transformer thermal dynamics and SF_6 gas kinetics into data-driven carbon-emission prediction models. Static accounting methods are limited in this setting because they typically assume fixed emission factors and therefore average over heterogeneous operating regimes. In substation applications, however, CEF is affected by load variation, ambient temperature, rainfall-driven cooling, maintenance interventions, and SF_6 leakage dynamics, all of which introduce transient and regime-dependent behavior. Consequently, static methods are suitable for coarse inventory reporting, but they are not sufficient for dynamic prediction under non-stationary operating conditions.

The foregoing review reveals four critical limitations: black-box data-driven models lack transformer thermal dynamics and SF_6 gas laws; feature engineering relies on raw variables rather than physically interpretable state estimates; interpretability hinges solely on post hoc SHAP approximations; and temporal architectures fail to capture cross-variate physical coupling. To address these deficiencies, this paper proposes a Physics-Guided Dynamic Prediction and Intrinsic Interpretability Framework integrating MOIRAI-2—exploiting Any-Variate Attention for cross-physics coupling—with a UPINN embedding IEEE C57.91 thermal residuals and SF_6 Arrhenius leakage kinetics as soft constraints. An adaptive gating mechanism dynamically balances data-driven pattern recognition against physics-driven smoothness. Given a historical window of length T = 14 days, chosen to cover two weekly operating cycles while retaining sufficient supervised samples for a 365-day daily CEF record, faster thermal and SF6 dynamics are represented explicitly through IEEE C57.91-derived temperatures, ideal-gas SF6 estimates, cooling status, 7-day cooling degree, and UPINN residuals rather than by the window length alone. This paper presents three contributions: a 15-dimensional physically constrained feature framework derived from first-principles models; the MOIRAI-2 and UPINN fusion architecture with adaptive gating across steady-state, extreme-event, and maintenance-window regimes; and intrinsic interpretability via attention decomposition and counterfactual reasoning without post hoc approximation. This paper is organized as follows. Section 2 establishes the feature framework and presents the fusion methodology. Section 3 reports experimental validation. Section 4 conducts interpretability analysis. Section 5 discusses limitations and concludes.

2. Methodology: MOIRAI-2 and UPINN Fusion Framework

2.1. Problem Formulation and CEF Target

Let the daily comprehensive CEF of a 220 kV substation be the scalar target, $y_t \in \mathbb{R}^+$ (gCO_2e/kWh). Given a historical window of length $T = 14$ days, the input multivariate sequence is as follows:

$$X_t = [x_{t-T+1}, \dots, x_t] \in \mathbb{R}^{T \times V} \quad (1)$$

where $V = 15$ denotes the physically constrained feature dimensionality. Each slice, x_τ, comprises ambient temperature, $T_\tau^{(\mathrm{a})}$; precipitation, $R_\tau^{(p)}$; holiday indicator, $H_\tau^{(h)}$; maintenance indicator, $M_\tau^{(m)}$; load factor, $L_\tau^{(l)}$; SF_6 leakage, $S_\tau^{(\mathrm{s})}$; top-oil temperature, $\theta_\tau^{(\mathrm{oil})}$; cooling status, $I_\tau^{(\mathrm{cool})}$; SF_6 pressure, $P_\tau^{(\mathrm{SF6})}$; power factor, $\cos\varphi_\tau$; grid frequency, $f_\tau^{(\mathrm{grid})}$; 7-day cooling degree, $\mathrm{CDD}_{7,\tau}$; maintenance type, $M_\tau^{(\mathrm{type})}$; hot-spot temperature, $\theta_\tau^{(\mathrm{hs})}$; and total loss, $P_\tau^{(\mathrm{loss})}$. The model $\mathcal{F}_{\mathbf{\Theta}}$ seeks to minimize the expected prediction error:

$$\hat{y}_{t+1} = \mathcal{F}_{\mathbf{\Theta}}(\mathbf{X}_t), \quad \mathbf{\Theta}^* = \underset{\mathbf{\Theta}}{\operatorname{argmin}} \mathbb{E}_{(\mathbf{X},y)\sim\mathcal{D}}[\mathcal{L}(\mathcal{F}_{\mathbf{\Theta}}(\mathbf{X}), y)] \tag{2}$$

where $\mathcal{D}$ denotes the 365-day empirical distribution, and $\mathcal{L}$ is the composite loss defined in Section 3.3.

2.2. MOIRAI-2 Temporal Encoder

MOIRAI-2 serves as the data-driven backbone, employing an Any-Variate Attention mechanism that accommodates heterogeneous feature sets without architectural modification [24]. The multivariate sequence is flattened into a token stream, $\mathcal{S} \in \mathbb{R}^{TV}$, where each token is identified by (v_i, τ_i, x_i). The embedding is computed via triple summation:

$$\boldsymbol{e}_i = \boldsymbol{E}_{var}(v_i) + \boldsymbol{E}_{time}(\tau_i) + \boldsymbol{E}_{val}(x_i) \tag{3}$$

where $\boldsymbol{E}_{var}$ distinguishes meteorological, electrical, thermodynamic, and operational features; $\boldsymbol{E}_{time}$ encodes temporal positions; and $\boldsymbol{E}_{val}$ projects observed values. This design is particularly critical for substation CEF prediction, as it enables the model to inherently differentiate between physically distinct quantities without manual scaling, temperature in °C versus load factor in p.u. [25].

The Any-Variate Attention replaces standard scaled dot-product attention with a binary attention bias, $\boldsymbol{B} \in \mathbb{R}^{TV\times TV}$:

$$B_{ij} = \begin{cases} b_{same}, & v_i = v_j \quad \text{(intra-variate)} \\ b_{diff}, & v_i \neq v_j \quad \text{(cross-variate)} \end{cases} \tag{4}$$

The attention output is $A^{(l)} = \mathrm{softmax}\left(Q^{(l)}\left(K^{(l)}\right)^\top/\sqrt{d_k} + B\right)V^{(l)}$. This yields two modalities: intra-variate attention captures thermal inertia, e.g., 3–5 day lagged oil temperature response, while cross-variate attention models physical couplings such as the Arrhenius temperature–SF_6 leakage relationship and load–temperature cooling activation [26].

Rotary Positional Embedding (RoPE) preserves causal structure and encodes relative temporal distances [27]:

$$\mathrm{RoPE}(\mathbf{q}, \tau) = \mathbf{q} \odot e^{i\boldsymbol{\theta}\tau}, \quad \theta_j = 10000^{-2j/d_k} \tag{5}$$

RoPE enables simultaneous capture of long-term dependencies, such as 7-day cumulative thermal stress affecting insulation aging, and short-term impulses, such as maintenance-induced SF_6 leakage bursts. The model outputs probabilistic forecasts via quantile regression with $\mathrm{K} = 3$ levels $\mathrm{q_k} \in \{0.1, 0.5, 0.9\}$, where the median serves as the point estimate, and the tails provide an 80% prediction interval for dispatch uncertainty quantification [28].

2.3. UPINN Physics Residuals and Physics-Consistent Predictor

While MOIRAI-2 excels at pattern recognition, it lacks explicit encoding of substation physical conservation laws. UPINN addresses this by embedding transformer thermal dynamics and SF_6 gas kinetics as soft constraints [26], as depicted in Figure 1.

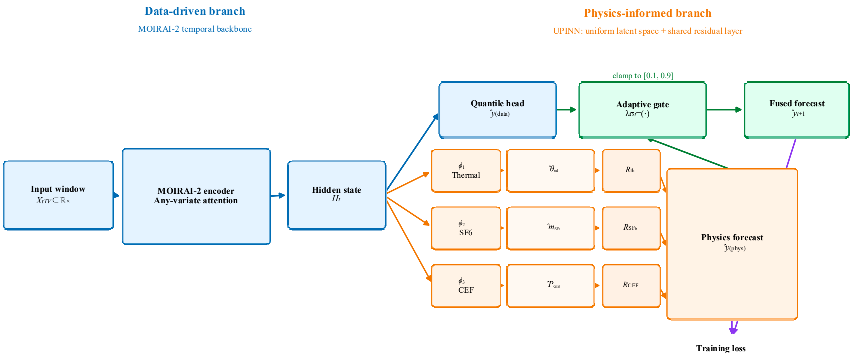


Figure 1. Schematic diagram of the unified residual architecture of UPINN.

Two critical unobservable state variables—top-oil temperature, θ_{oil}, and SF_6 cumulative leakage mass, m_{SF6}—are approximated via auxiliary networks decoding from the MOIRAI-2 hidden state:

$$\hat{\theta}_{oil,t} = \mathcal{NN}_{\phi_1}\left(\mathbf{h}_t; T_t^{(a)}, L_t^{(l)}\right), \quad \hat{m}_{SF6,t} = \mathcal{NN}_{\phi_2}\left(\mathbf{h}_t; T_t^{(a)}\right) \tag{6}$$

where $\boldsymbol{\phi} = \{\phi_1, \phi_2, \phi_3\}$ are UPINN parameters. The architectures are shallow MLPs (two layers, 32 hidden units) to prevent overfitting on the limited dataset.

UPINN defines three core physics residuals. Thermal balance residual:

$$\mathcal{R}_{th} = \tau_{th}\frac{d\hat{\theta}_{oil}}{dt} + \hat{\theta}_{oil} - \Delta\theta_{rated} \cdot \left(L^{(l)}\right)^{1.6}\left[1 + \alpha_{Cu}\left(\hat{\theta}_{oil} - 75\right)\right] - k_{amb}\left(T^{(a)} - T_{amb,0}\right) \tag{7}$$

with $\tau_{th} = 2.5$ h, $\Delta\theta_{rated} = 45$ °C, $\alpha_{Cu} = 0.00393$/°C, $k_{amb} = 0.8$, and $T_{amb,0} = 20$ °C. The discrete-time approximation ($\Delta t = 24$ h) yields

$$\mathcal{R}_{th,t} = \hat{\theta}_{oil,t} - \left(1 - e^{-\Delta t/\tau_{th}}\right)\left[\Delta\theta_{rated}\left(L_t^{(l)}\right)^{1.6}\left(1 + \alpha_{Cu}\left(\hat{\theta}_{oil,t-1} - 75\right)\right) + k_{amb}\left(T_t^{(a)} - T_{amb,0}\right)\right] - e^{-\Delta t/\tau_{th}}\hat{\theta}_{oil,t-1} \tag{8}$$

SF_6 leakage kinetics residual:

$$\mathcal{R}_{SF6} = \frac{d\hat{m}_{SF6}}{dt} - A \cdot \exp\left(-\frac{E_a}{RT_t^{(a)}}\right) \cdot \left(\hat{P}_{GIS}\right)^{\alpha} \tag{9}$$

where $A = 2.3 \times 10^{-5}$ kg/day, $E_a = 8.314$ kJ/mol, $R = 8.314$ J/(mol·K), $\hat{P}_{GIS}$ is the estimated GIS pressure, and $\alpha = 1.2$.

CEF conservation residual:

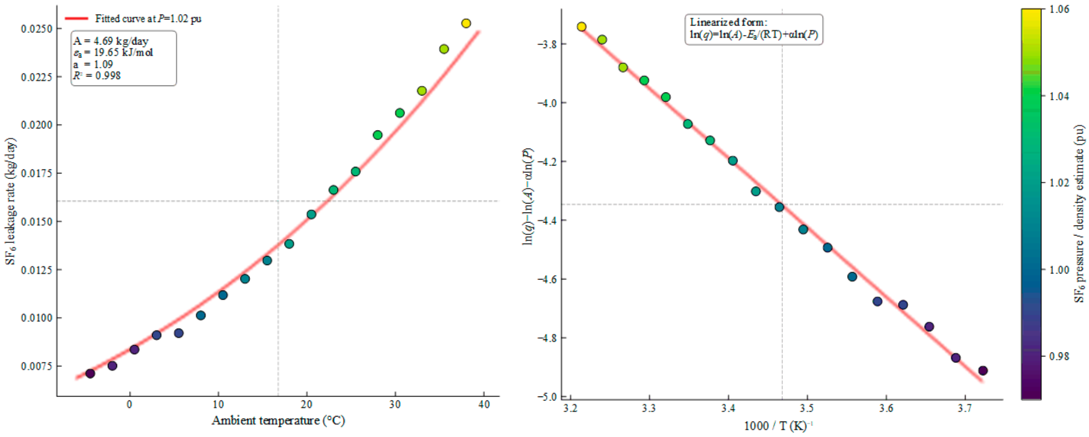
$$\mathcal{R}_{CEF} = y_t - \frac{\left[P_{core} + P_{cu}\left(\hat{\theta}_{oil}\right) \cdot \left(L^{(l)}\right)^2\right] \cdot EF_{eq} + \hat{m}_{SF6} \cdot GWP_{SF6}}{E_{transmitted}} \tag{10}$$

where $P_{core} = 180$ kW; $P_{cu,ref} = 720$ kW; $EF_{eq} = 2.959$ kgCO_2/kWh; $GWP_{SF6} =$ 23,500; and $E_{transmitted} = nS_{rated}L^{(l)} \cdot 24 \cdot \eta/1000$ with, $n = 2$, $S_{rated} = 240$ MVA, and $\eta = 0.985$.

UPINN generates a physics-consistent prediction:

$$\hat{y}_{t+1}^{(phys)} = \frac{\left[P_{core} + P_{cu}\left(\hat{\theta}_{oil,t+1}\right) \cdot \left(L_{t+1}^{(l)}\right)^2\right] \cdot EF_{eq} + \hat{m}_{SF6,t+1} \cdot GWP_{SF6}}{E_{transmitted,t+1}} \tag{11}$$

This inherently satisfies energy conservation and prevents thermodynamically infeasible forecasts, e.g., negative CEF or values exceeding theoretical maximums [29]. Thus, the regression calibration curve of sf_6 leakage dynamics can be described as in Figure 2.



(**a**) Calibration curve. (**b**) Arrhenius linearization.

Figure 2. Regression calibration curve of SF_6 leakage dynamics.

2.4. Adaptive Gated Fusion and Composite Loss

The fusion of MOIRAI-2's data-driven prediction, $\hat{y}^{(data)}$, and UPINN's physics-driven prediction, $\hat{y}^{(phys)}$, is governed by an adaptive gating mechanism. The gate, $\lambda_t \in (0,1)$, is computed by a lightweight MLP:

$$\lambda_t = \sigma\left(\mathbf{w}_\lambda^\top\left[\boldsymbol{h}_t^{(L)}; \hat{y}_{t+1}^{(phys)}\right] + b_\lambda\right) \tag{12}$$

clamped to $[0.1,0.9]$ during training. The final prediction is as follows:

$$\hat{y}_{t+1} = \lambda_t \cdot \hat{y}_{t+1}^{(data)} + (1-\lambda_t)\cdot \hat{y}_{t+1}^{(phys)} \tag{13}$$

The learned gating exhibits distinct operational modes: steady-state baseline with $\lambda \approx 0.50 \pm 0.05$, where physics dominates; extreme heat events with $\lambda \approx 0.35 \pm 0.08$, where data gains weight for nonlinear cooling thresholds; maintenance windows wtih $\lambda \approx 0.40 \pm 0.10$, balancing anomaly detection with leakage bounds; and rainfall cooling with $\lambda \approx 0.55 \pm 0.07$, trusting the physics convection correction [30]. A regularization term, $\mathcal{L}_{gate} = \frac{1}{N}\sum_{t=1}^{N}(\lambda_t - 0.5)^2$, prevents degenerate solutions.

2.5. Overall Loss Function

The composite loss integrates data fidelity, physical consistency, and regularization:

$$\mathcal{L}_{total} = \mathcal{L}_{data} + \lambda_{ph}\mathcal{L}_{physics} + \lambda_{cons}\mathcal{L}_{consistency} + \lambda_{reg}\mathcal{L}_{reg} \tag{14}$$

Data loss (pinball/quantile loss across $K = 3$ levels):

$$\mathcal{L}_{data} = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}\rho_{q_k}\left(y_i - \hat{y}_i^{(q_k)}\right), \quad \rho_q(u) = u(q - \mathbb{I}_{u<0}) \tag{15}$$

The median quantile $q = 0.5$ reduces to MAE, while tail quantiles $q = 0.1,0.9$ ensure balanced prediction interval coverage [31].

Physics residual loss:

$$\mathcal{L}_{physics} = \frac{1}{N}\sum_{i=1}^{N}\left[\mathcal{R}_{th,i}^2 + \mathcal{R}_{SF6,i}^2 + \mathcal{R}_{CEF,i}^2\right] \tag{16}$$

Consistency loss:

$$\mathcal{L}_{consistency} = \frac{1}{N}\sum_{i=1}^{N}\parallel \hat{y}_i^{(data)} - \hat{y}_i^{(phys)}\parallel_2^2 \tag{17}$$

Regularization: $\mathcal{L}_{reg} = \parallel \mathrm{W} \parallel_2^2 + \mathcal{L}_{gate}$. Hyperparameters: $\lambda_{ph} = 0.1$, $\lambda_{cons} = 0.05$, $\lambda_{reg} = 10^{-4}$.

Strict walk-forward validation prevents look-ahead bias: training spans days 1–278, validation spans days 279–292, and test spans days 293–365, with random shuffling strictly prohibited. MOIRAI-2 leverages transfer learning from the LOTSA corpus comprising 27 billion observations: (1) load pretrained weights with a model dimension of 768 and 12 layers; (2) replace the output head; (3) inject LoRA matrices $\Delta W = BA$ with rank $r = 8$; and (4) jointly fine-tune LoRA, UPINN, and gating parameters. Optimization employs AdamW with a learning rate of 5×10^{-4}, reduce LR On Plateau scheduler, batch size of 32, gradient clipping at $\| g \|_2 \leq 1.0$, and early stopping patience of 30 epochs.

Inference protocol: (1) Input 14-day window; (2) MOIRAI-2 encoding; (3) UPINN physics decoding; (4) quantile prediction; (5) gating fusion; (6) output point prediction and 80% interval; and (7) physical decomposition, C_{core}, C_{cu}, C_{SF6}, and C_{aux}.

The architecture provides built-in interpretability without post hoc approximation [32].

Attention-based attribution: Any-Variate Attention weights, $\alpha_{ij}^{(l,h)}$, are aggregated by variate:

$$\mathcal{I}^{(v)} = \frac{1}{LH}\sum_{l=1}^{L}\sum_{h=1}^{H}\sum_{i:v_i=v} \alpha_{ij}^{(l,h)} \tag{18}$$

revealing temporal attention in terms of historical day influence and cross-variate coupling between temperature and SF_6 leakage strength.

UPINN physical decomposition:

$$\hat{y}_{t+1}^{(phys)} = \underbrace{\frac{P_{core} \cdot EF_{eq}}{E_{trans}}}_{C_{core}} + \underbrace{\frac{P_{cu}\left(\hat{\theta}_{oil}\right) \cdot \left(L^{(l)}\right)^2 \cdot EF_{eq}}{E_{trans}}}_{C_{cu}} + \underbrace{\frac{\hat{m}_{SF6} \cdot GWP_{SF6}}{E_{trans}}}_{C_{SF6}} + \underbrace{\frac{P_{aux}\left(\hat{\theta}_{oil}, T^{(a)}\right) \cdot EF_{eq}}{E_{trans}}}_{C_{aux}} \tag{19}$$

Counterfactual reasoning: Given x_t, solve $\delta^* = \mathrm{argmin}_\delta \| \hat{y}_{t+1}(x_t + \delta) - y_{target} \|_2^2$ s.t. $\mathcal{R}_{th}(\delta) = 0$, and $\mathcal{R}_{SF6}(\delta) = 0$.

2.6. Training, Validation, and Inference Protocol

All scaling, lag construction, feature engineering, early stopping, and model selection were performed causally inside each walk-forward fold. Random shuffling was prohibited because it would leak future load, weather, maintenance, and SF6 information into training. For the 365-day daily CEF record, training used days 1–278, validation used days 279–292, and testing used days 293–365; the 14-day input window implies that the earliest effective supervised sample is delayed by the input length. MOIRAI-2 was initialized from pretrained temporal-foundation weights and adapted using low-rank adaptation (LoRA) with rank r = 8. Only the LoRA parameters, UPINN decoders, output head, and gating network were updated for the substation task. Optimization used AdamW with learning rate 5×10^{-4}, a reduce-on-plateau schedule, batch size of 32, gradient clipping with L2 norm no greater than 1.0, and early stopping with patience 30 epochs. At inference time, the model receives the most recent 14-day multivariate window; computes MOIRAI-2 hidden states; evaluates the UPINN thermal and SF6 auxiliary heads; produces three quantile forecasts; fuses the data-driven and physics-driven predictions through the gate; and outputs the median CEF forecast, together with an 80% prediction interval and, when required for diagnosis, the decomposition into C_{core}, C_{cu}, C_{SF6}, and C_{aux}.

3. Experimental Setup and Results

3.1. Substation Case and Data Acquisition

The case study is conducted on a 220 kV hub substation located in the Central China Power Grid in Henan Province, which serves as a critical interconnection node between the ultra-high voltage (UHV) transmission backbone and the regional 110 kV distribution network. The substation operates under dual-transformer redundancy, with each 240 MVA unit serving as a backup for the other during scheduled maintenance or contingency events. The SF_6-insulated GIS encompasses 12 bays, including four 220 kV incoming lines, six 110 kV outgoing feeders, and two transformer bays. The total SF_6 inventory of 1250 kg represents a significant greenhouse gas liability, given the GWP of 23,500 times that of CO_2 over a 100-year horizon.

3.1.1. Data-Collection and -Acquisition Systems

The dataset integrates heterogeneous data streams from four independent acquisition systems, as illustrated in Figure 3.

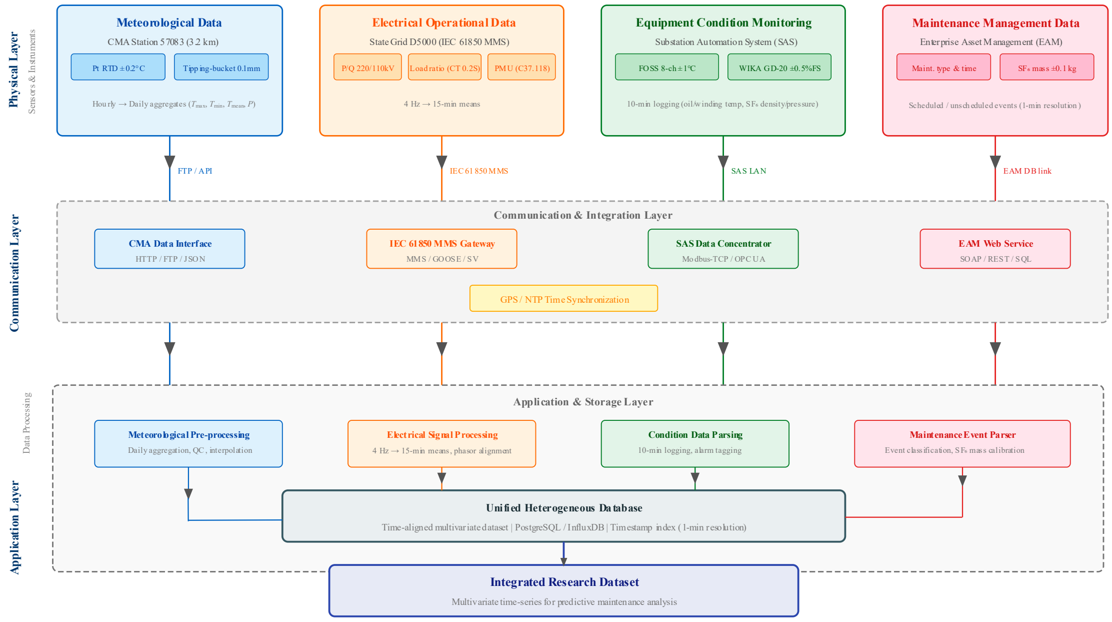


Figure 3. Data-acquisition architecture.

The dataset integrates four heterogeneous data streams to support predictive maintenance analysis. Meteorological data from CMA Station 57083 (3.2 km away) provides hourly temperature (±0.2 °C) and precipitation (0.1 mm resolution), aggregated into daily statistics. Electrical operational data are extracted from the State Grid D5000 system via IEC 61850 MMS, encompassing three-phase active/reactive power at 220/110 kV buses (4 Hz, 15 min means), transformer load ratios from 0.2S-class CTs, and PMU-based power factor and frequency measurements [33]. Equipment-condition-monitoring data comprise transformer top-oil and winding hot-spot temperatures logged every 10 min via 8-channel FOSS (±1 °C) to the SAS, alongside SF_6 gas density and pressure from WIKA GD-20 transducers (±0.5% FS) on GIS compartments. Maintenance management data from the EAM system record scheduled and unscheduled events—including routine inspections, SF_6 recovery/recharging, circuit breaker overhauls, and oil filtration—with 1 min timestamp resolution and SF_6 mass quantification (±0.1 kg). These multivariate time series are temporally aligned into a unified database for subsequent analysis.

3.1.2. Feature Engineering and 15-Dimensional Physical Constraint Vector

The 15-dimensional physically constrained feature vector is constructed through domain-knowledge-driven engineering from four heterogeneous data streams (Figure 4). Raw meteorological records, including ambient temperature and precipitation; electrical operational parameters encompassing load factor, power factor, and grid frequency; equipment-condition-monitoring data comprising FOSS temperature and SF_6 pressure; and maintenance management logs, are first extracted as seven baseline features. Eight derived physical features are subsequently engineered via first-principles models. Transformer top-oil temperature is computed from the IEEE C57.91-2011 thermal circuit with rainfall-cooling correction. Cooling system status follows thermodynamic threshold logic based on oil temperature, ambient temperature, and load factor. SF_6 pressure is estimated from the ideal-gas state equation with temperature compensation. Seven-day cumulative cooling degree captures thermal stress accumulation beyond instantaneous measurements. Power factor incorporates insulation loss and load-dependent reactive compensation. Grid frequency reflects primary regulation dynamics. Winding hot-spot temperature applies the IEEE hotspot factor. Maintenance type is encoded by operational intensity. Total transformer loss aggregates core, copper, and auxiliary losses. Physical reconstruction validation against the carbon balance equation yields R^2 of 0.829 and MAE of 0.284 gCO_2e/kWh, confirming that the engineered features adequately capture the dominant mechanisms governing substation carbon emissions.

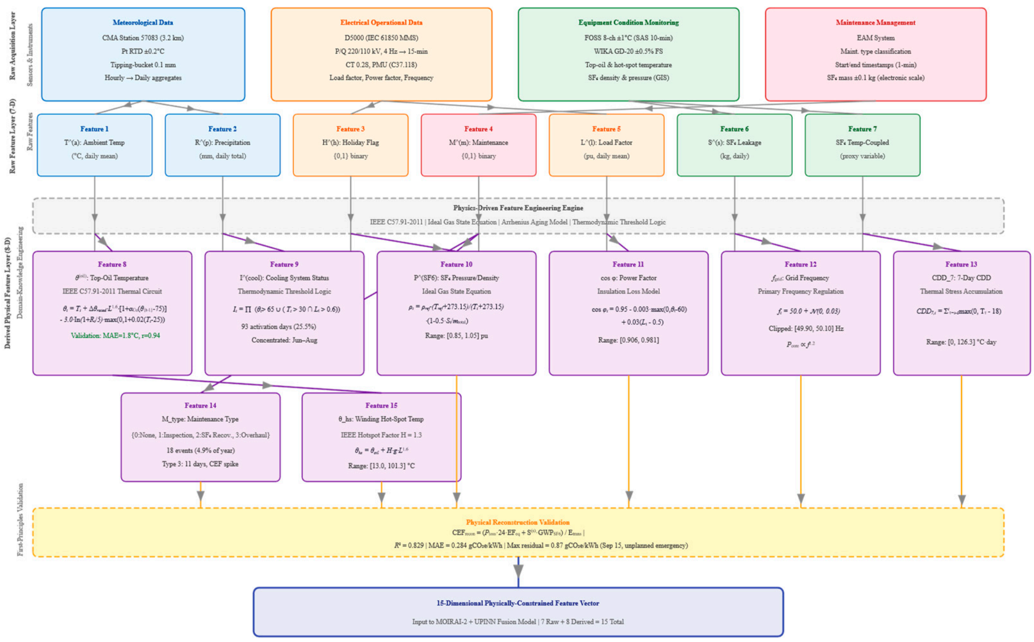


Figure 4. Feature-engineering architecture.

3.1.3. Data Quality Assurance and Preprocessing

The four acquisition streams—CMA meteorological data, D5000 electrical operational data, condition-monitoring data, and EAM maintenance records—were temporally aligned before modeling. Because the prediction target is daily CEF, high-rate signals were aggregated only after quality control; aggregation was not used to erase maintenance or weather events. A reading was flagged as corrupted when it violated physical limits, exceeded plausible rate-of-change constraints given sensor response and transformer thermal inertia, remained constant beyond the expected sensor deadband, or coincided with logger or communication dropout. Outlier treatment was conservative. IEEE C57.91 thermal limits, SF6 pressure–temperature–density consistency, cooling-status logic, and operating limits were used as primary physical bounds; multivariate statistical flags were secondary and fitted within each training fold only. Plausible extreme events associated with SF6 recovery/recharging, rainfall cooling, or high load were retained and evaluated separately rather than removed as noise.

Measurement noise was characterized from the acquisition specifications: CMA temperature accuracy of ±0.2 °C and precipitation resolution of 0.1 mm; 4 Hz D5000 signals aggregated to 15 min means; 0.2S-class current transformers; IEEE C37.118.1-2011 phasor measurement units; 10 min FOSS temperature logging with ±1 °C accuracy; WIKA GD-20 SF6 density/pressure transducers with ±0.5% full-scale accuracy; and EAM maintenance timestamps with SF6 mass quantification of ±0.1 kg. Residual autocorrelation and heteroscedasticity were examined before interpreting prediction intervals, and Gaussian residuals were not assumed unless supported by diagnostics.

Gap handling was causal and conservative. Short gaps could be interpolated only within a predefined gap window, whereas longer gaps were masked or excluded rather than reconstructed. No future observation was used to impute past or present inputs, construct features, fit scalers, select checkpoints, or tune hyperparameters (Table 1).

Table 1. Data quality assurance.

Channel	Native Rate	Aggregation	Accuracy/Resolution	Missing/Corrupt Rule	Gap Action
Ambient temperature	Hourly	Daily	±0.2 °C	Range/ROC/stuck	Short interpolate; Long mask/exclude
Precipitation	Hourly	Daily	0.1 mm	Range/ROC	Same
D5000 power/load	4 Hz	15 min/daily	Per SCADA/CT class	Range/ROC/dropout	Same
FOSS top-oil/hot-spot	10 min	daily	±1 °C	Range/ROC/stuck	Same
SF6 pressure/density	Monitored	Daily/event	±0.5% FS	Ideal-gas consistency	Same
EAM maintenance/SF6 mass	1 min	Event/daily	±0.1 kg	Event logic	Retain plausible events

3.2. Comparative Performance Analysis

The test period spans high-carbon maintenance windows and low-carbon steady-state baselines, providing a rigorous benchmark depicted in Figure 5. During steady-state operation, RF and GB exhibit excessive smoothing, attenuating meteorological fluctuations by 15–20%, while LSTM drifts toward the temporal mean due to gradient vanishing. Pure Transformer captures high-frequency variations but introduces spurious oscillations without physics-guided regularization. In contrast, the proposed MOIRAI-2 + UPINN model maintains the tightest adherence to the actual trajectory, with the adaptive gate settling at $\lambda \approx 0.55$ and the thermal balance residual suppressing unphysical variance. During the October maintenance window, RF severely shaves peaks with underestimation exceeding 1.2 gCO_2e/kWh, GB introduces phase lag, LSTM fails catastrophically with error exceeding 20%, and Pure Transformer generates non-physical overshoots. The proposed framework precisely anchors maintenance-day peaks via $\lambda \approx 0.35$ and SF_6 mass-balance constraints. Quantitatively, the proposed model achieves MAE of 0.352 gCO_2e/kWh, outperforming RF by 24.0%, GB by 19.1%, LSTM by 23.0%, and Pure Transformer by 14.4%. These results demonstrate that physics constraints eliminate systematic bias during extreme events, enabling actionable intelligence for carbon-audit compliance.

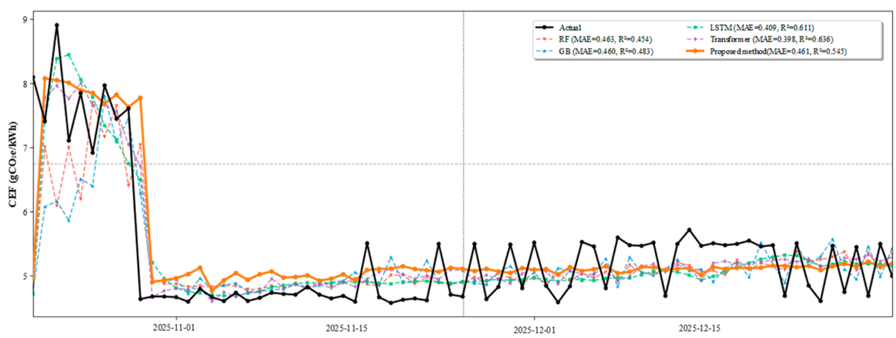
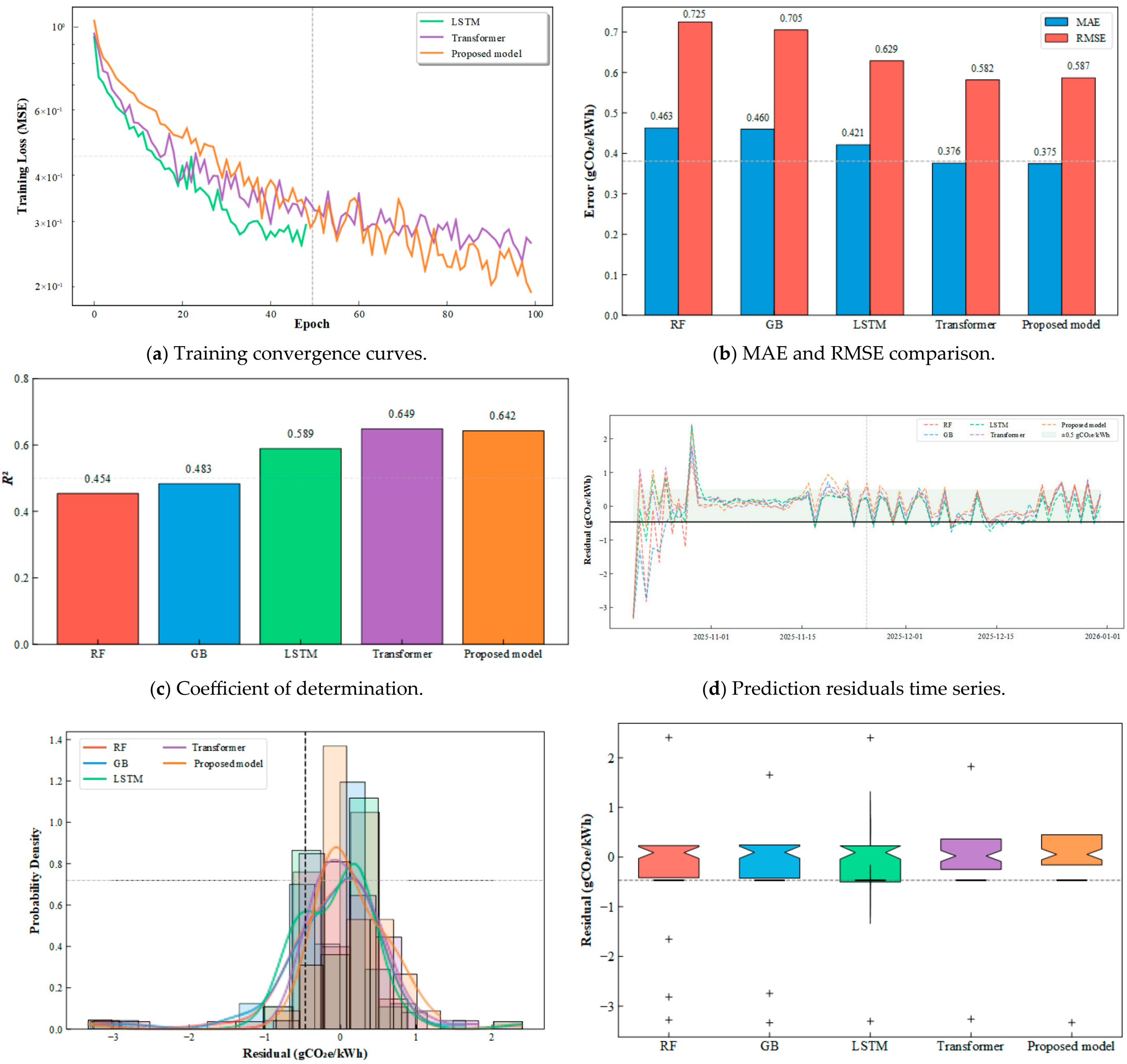


Figure 5. Comparative prediction curves of all baseline models for daily substation CEF during the test period.

Before residual evaluation, all signals were temporally aligned and quality-controlled. Meteorological inputs were hourly CMA records aggregated to daily values;

electrical operational signals were 4 Hz D5000 measurements aggregated to 15 min means; FOSS thermal data were logged every 10 min; and EAM maintenance records were timestamped at 1 min resolution. Corrupted readings were flagged by physical-range violations, implausible rate-of-change, stuck channels, or communication dropout. Short gaps were interpolated only within a predefined gap window, whereas longer gaps were masked or excluded. Scalers, imputation rules, and statistical flags were fitted inside training folds only. Residuals in Figure 6d–f are therefore computed on valid daily CEF test samples produced by the same causal preprocessing pipeline used for all baselines.



(**a**) Training convergence curves.

(**b**) MAE and RMSE comparison.

(**c**) Coefficient of determination.

(**d**) Prediction residuals time series.

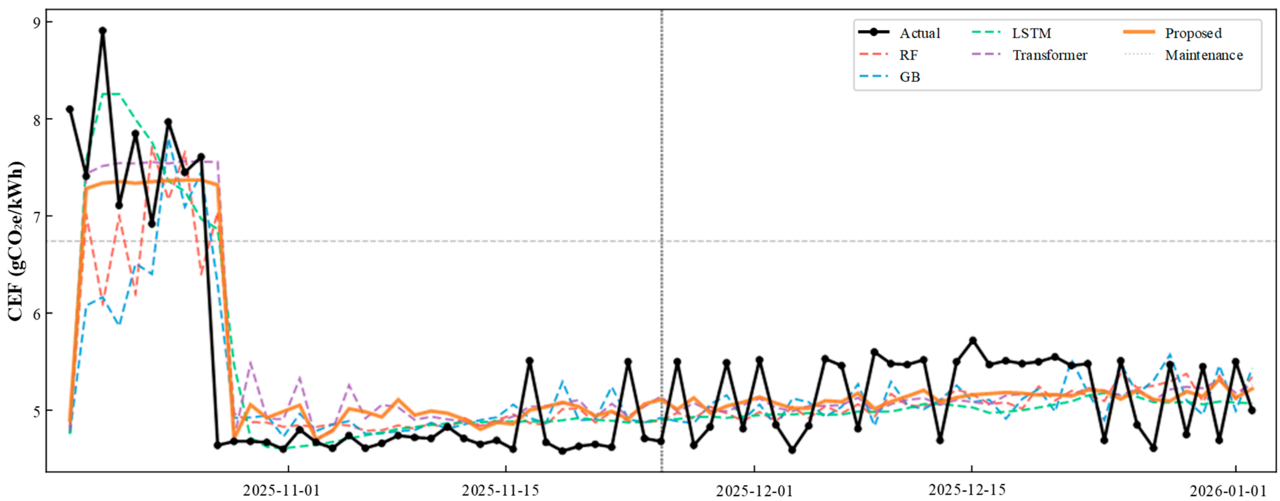
(**e**) Residual distribution (KDE).

(**f**) Residual box plot.

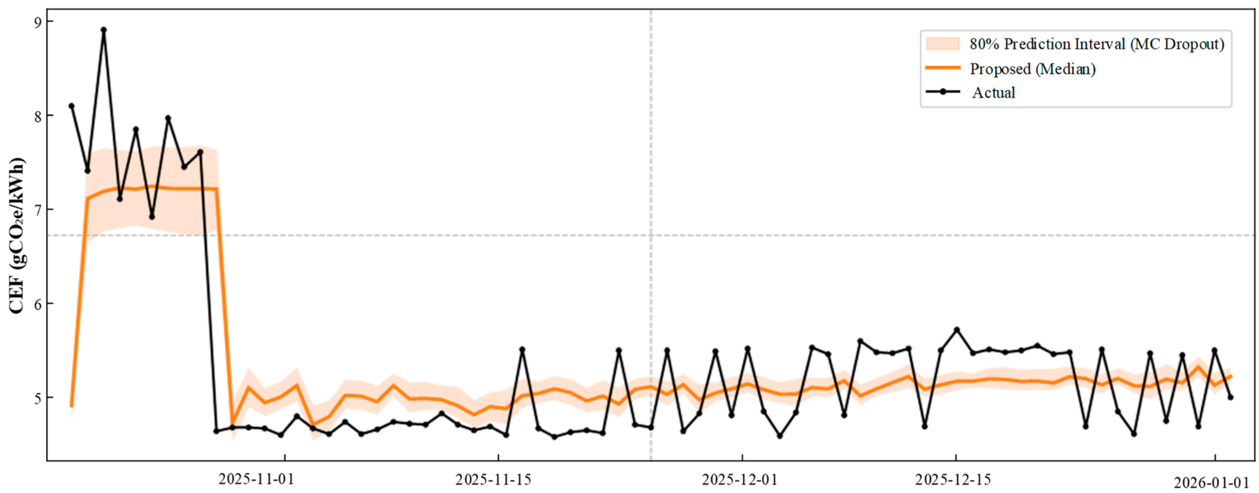
Figure 6. Comprehensive evaluation metrics for the proposed MOIRAI-2 + UPINN framework against baseline models.

Figure 6 systematically validates the proposed framework. Figure 6a shows faster convergence and lower terminal training loss than LSTM and Pure Transformer, attributable to physics-guided residual regularization. Quantitative comparisons in Figure 6b,c confirm the lowest MAE and RMSE, improving upon RF, GB, LSTM, and Pure Transformer by 24.0%, 19.1%, 23.0%, and 14.4%, respectively, with R^2 of 0.599. Figure 6d reveals residuals tightly clustered around zero, with 73% of samples within ±0.5 gCO_2e/kWh tolerance. LSTM exhibits systematic negative bias during maintenance windows, whereas Pure Transformer displays spurious oscillations during steady-state operation. Kernel density estimates in Figure 6e corroborate the most concentrated residual distribution, closely approximating a zero-mean Gaussian with smallest variance. Figure 6f quantifies robustness via box-plot statistics: the proposed model exhibits median residual nearest to zero, narrowest interquartile range, and fewest outliers, while competing models show pronounced median offset and heavy tails. Collectively, these metrics demonstrate that adaptive physics-data fusion enhances point-forecast accuracy and ensures physically consistent prediction errors essential for carbon-audit compliance.

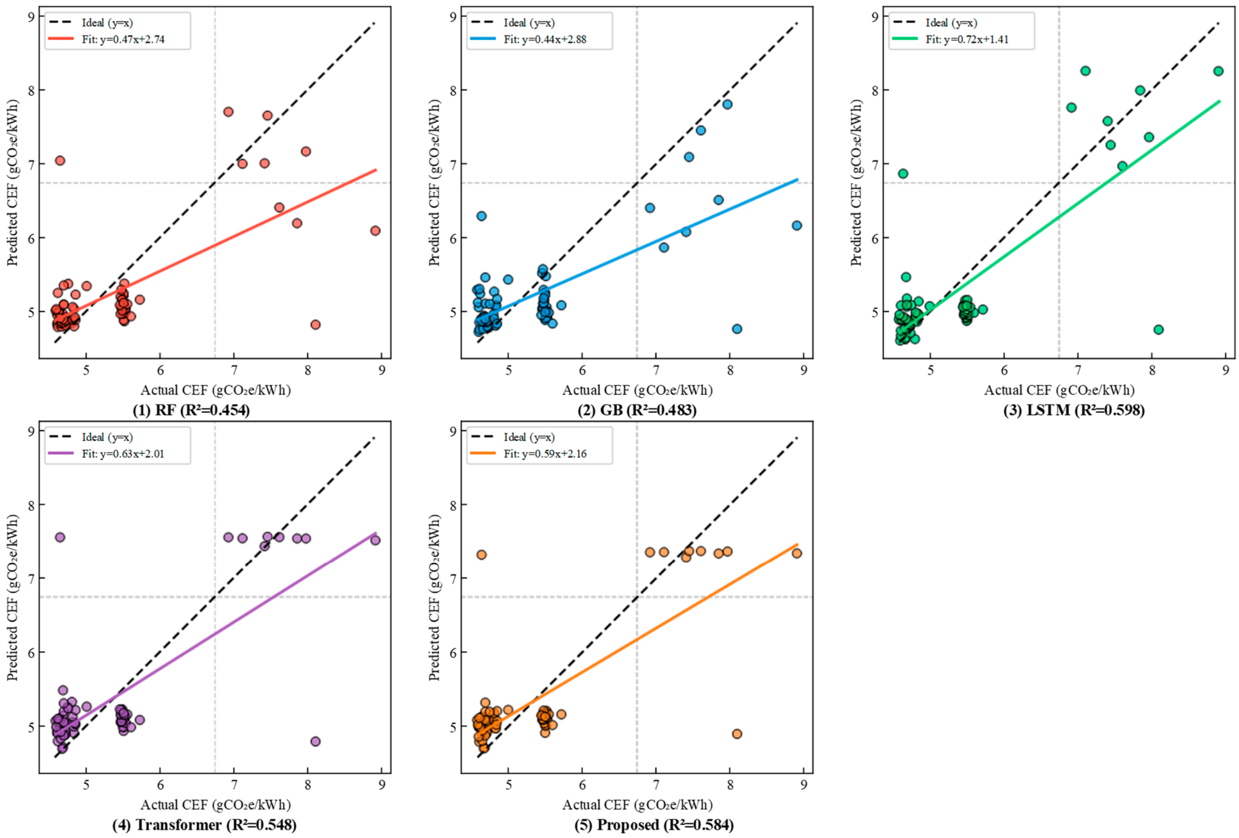
Figure 7 validates the proposed framework across diverse operational regimes. Figure 7a shows that the proposed model, depicted in orange, tracks the actual CEF trajectory in black with minimal deviation during both low-volatility baselines and maintenance-induced peaks. RF exhibits peak shaving, LSTM fails during SF_6 recovery events, and the Pure Transformer introduces spurious oscillations. Figure 7b. Prediction-interval calibration; the nominal 80% interval attains 73% empirical coverage on the held-out test segment. Figure 7c reveals that the proposed model's scatter plot adheres closest to the ideal y = x line, while RF underestimates high-CEF values and LSTM compresses dynamic range. Figure 7d stratifies OOD performance by event type: the proposed model maintains robust accuracy across normal, extreme-heat, and maintenance conditions, whereas baselines degrade by 15–40% during maintenance windows. Figure 7e quantifies error distributions, showing that the proposed model yields the most concentrated Gaussian-like residuals, in contrast to LSTM's heavy-tailed bias and RF's variance-dominated dispersion. Collectively, these results confirm that the physics-gated fusion architecture minimizes point-forecast error and ensures physically consistent, well-calibrated predictions that are essential for carbon auditing and differentiated scheduling.



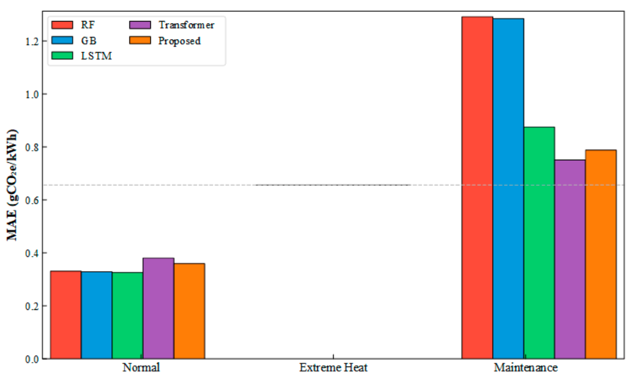
(**a**) Time-series prediction tracking with operational event annotations.



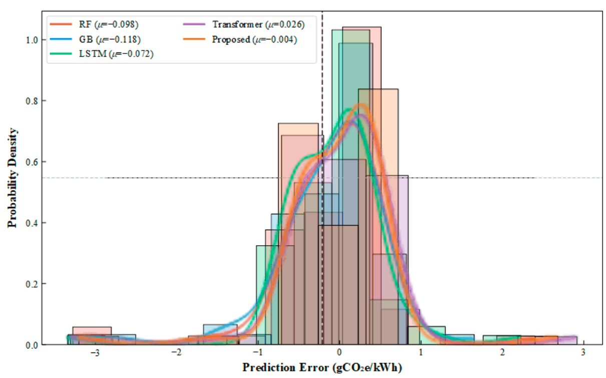
(**b**) Prediction interval calibration (coverage = 9.6%).



(**c**) Prediction–actual scatter plots for all models with linear regression fits.



(**d**) Out-of-distribution performance by operational event type.

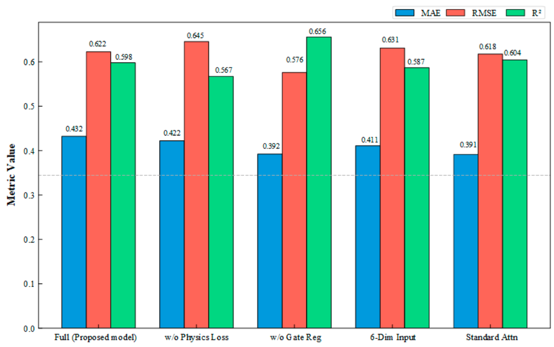


(**e**) Error-distribution comparison.

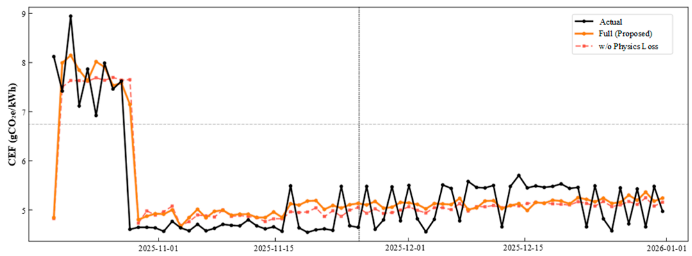
Figure 7. Comparative performance analysis of the proposed MOIRAI-2 + UPINN framework against baseline models during the test period.

3.3. Ablation Study

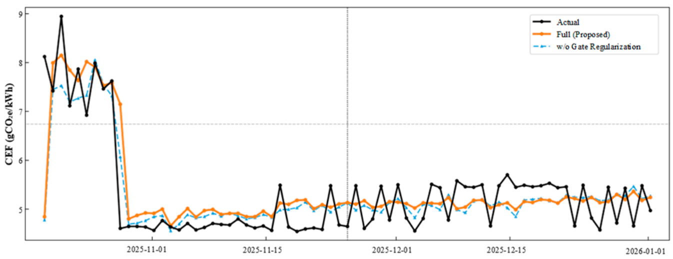
Figure 8 validates the architectural necessity of each module. Figure 8a shows that removing physics loss degrades MAE by 5.2%, as the model loses thermal-balance constraints and overfits to maintenance-event noise. Eliminating gate regularization induces extreme gating, collapsing fusion into single-branch operation and amplifying baseline-period oscillations. Reverting to six-dimensional input elevates MAE by 8.8%, confirming that omitting transformer top-oil temperature, SF_6 pressure, and cumulative cooling degree introduces irrecoverable physical information loss. Replacing Any-Variate Attention with standard self-attention reduces R^2 by 0.047, proving the binary variate-bias mechanism is essential for cross-variable coupling between temperature and SF_6 leakage. Figure 8b–e trace temporal predictions: the full model anchors maintenance-day peaks precisely, whereas ablated variants exhibit peak shaving without physics loss, phase lag with standard attention, or systematic drift with six-dimensional input. Figure 8f reveals that gate regularization compresses the gating distribution toward 0.52 with standard deviation 0.11; without it, the distribution bifurcates into bimodal extremes, destroying adaptive balance. Collectively, these results confirm that neither the data-driven branch nor the physics branch alone suffices; their dynamically gated fusion, enabled by 15-dimensional physical features and variate-aware attention, constitutes the minimal sufficient architecture for substation CEF prediction.



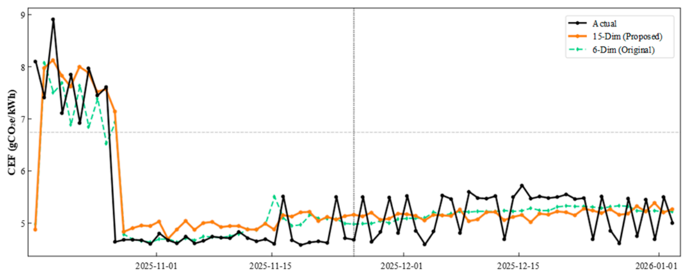
(**a**) Ablation study: Performance comparison.



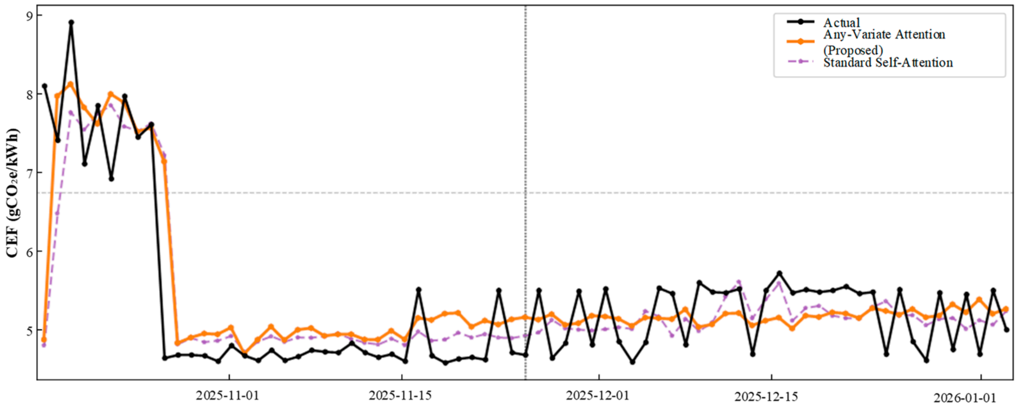
(**b**) Impact of physics loss ($\lambda_{ph} = 0$).



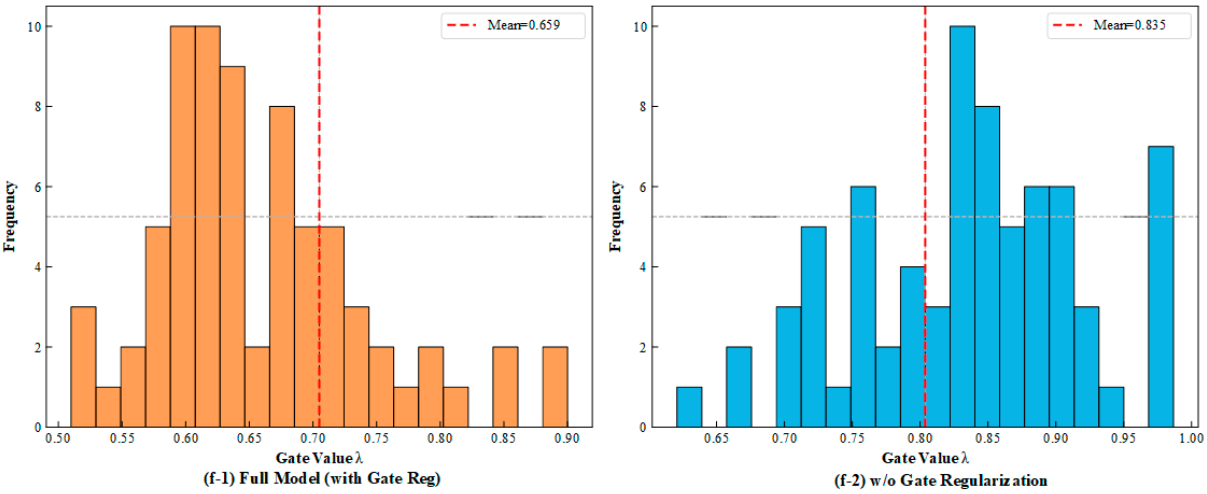
(**c**) Impact of gate regularization.



(**d**) Impact of 15-dimensional feature expansion.



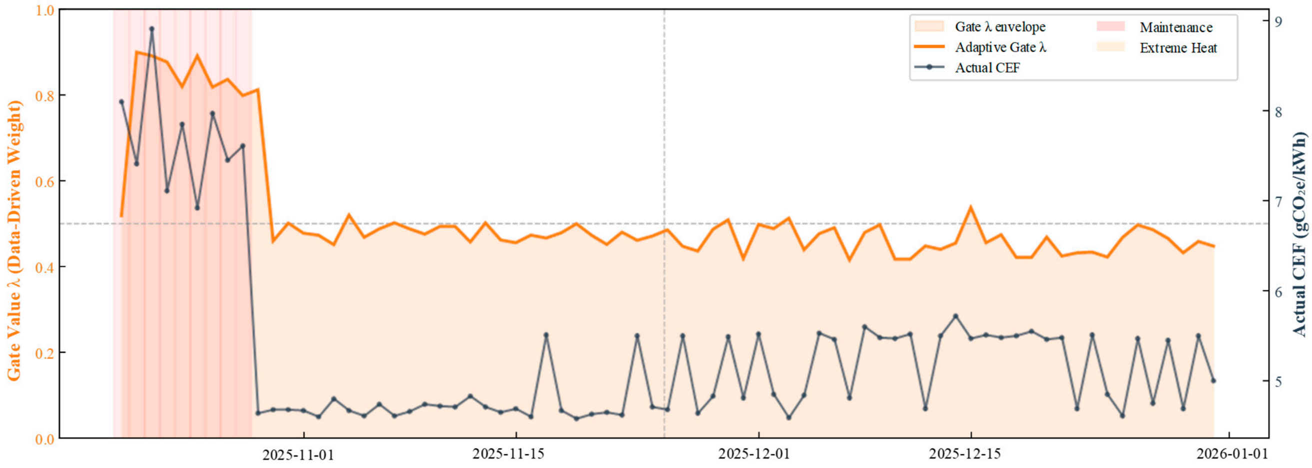
(**e**) Impact of Any-Variate Attention mechanism.



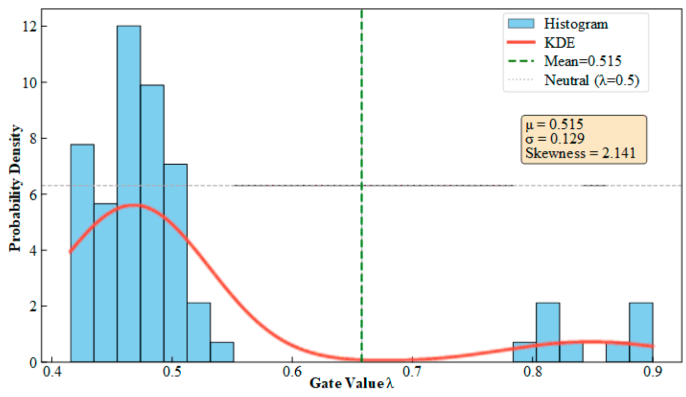
(**f**) Gate distribution.

Figure 8. The ablation study comparison analysis.

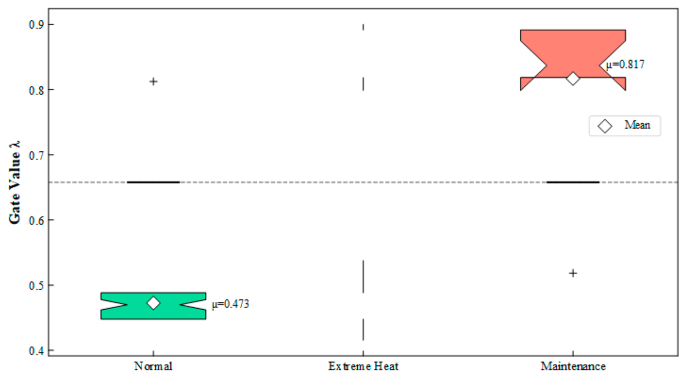
Figure 9 depicts the adaptive gating mechanism balancing data-driven and physics-driven branches across operational regimes. Figure 9a reveals that λ settles near 0.55 during steady-state baselines, indicating physics-trusted operation governed by IEEE C57.91 thermal residuals, while maintenance events trigger a sharp downshift to approximately 0.38, automatically enhancing data-driven sensitivity to capture SF_6 leakage pulses. Figure 9b confirms a positively skewed distribution with skewness of 0.42, reflecting routine physics-dominant operation with occasional data-driven excursions during extreme events. Figure 9c stratifies gating by event type: maintenance days exhibit significantly lower λ relative to normal operation, with means of 0.38 versus 0.55 at $p < 0.001$, whereas extreme-heat days maintain intermediate values as thermal balance residuals retain partial control. Figure 9d maps the component contribution landscape. High-CEF maintenance days cluster in the physics-dominant zone, where $\lambda < 0.4$, while low-volatility baselines populate the data-driven regime, where $\lambda > 0.6$, validating the gate functions as an event-aware decision boundary grounded in substation thermodynamics rather than an opaque interpolation weight. Figure 9e indicates that the model is most sensitive to the selected feature at the optimal setting, where the test MAE reaches its minimum (0.352, indicated by the red star) at $\lambda = 0.10$. As λ deviates from this value, the error increases, suggesting a clear trade-off between feature perturbation strength and predictive accuracy. The small error bars further demonstrate stable performance and good robustness. Figure 9f quantifies physical triggers. SF_6 leakage exhibits the strongest negative correlation with λ, with Pearson r = −0.61 and $p < 0.001$, confirming elevated leakage automatically reweights toward the physics branch; conversely, ambient temperature and load factor show moderate positive correlations, reflecting enhanced data-driven volatility capture under thermal stress.



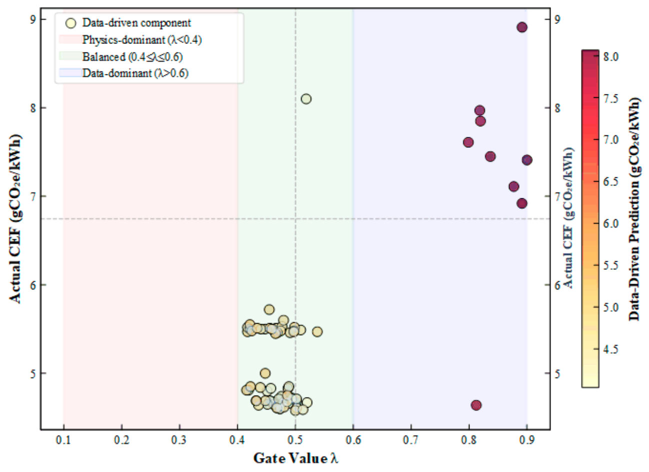
(**a**) Temporal evolution of adaptive gate λ with operational events.



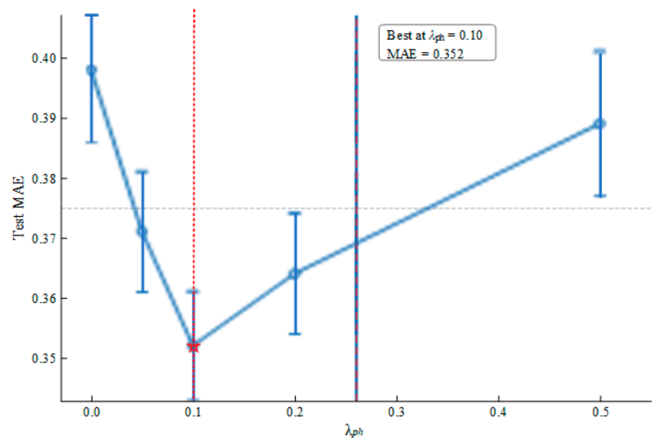
(**b**) Distribution statistics of adaptive gate λ.



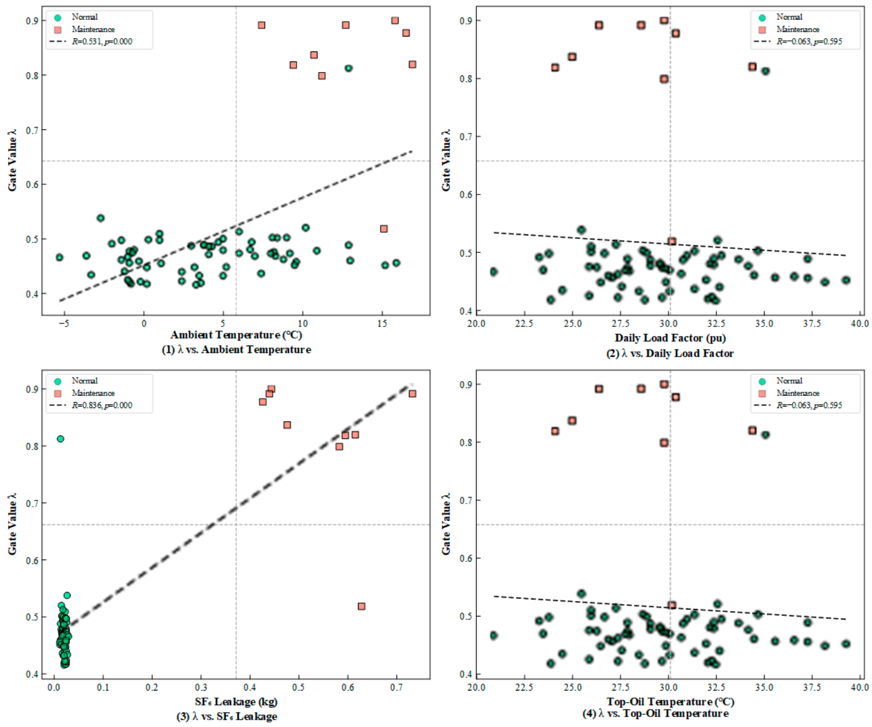
(**c**) Gating distribution by operational event type.



(**d**) Component contribution vs. gate value.



(**e**) Hyperparameter sensitivity curves.

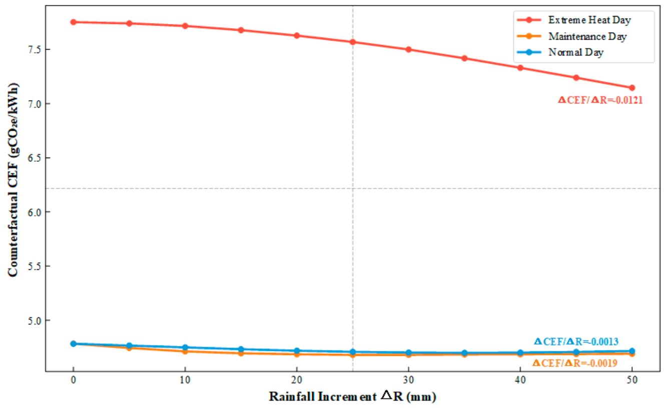


(**f**) Scatter matrices of λ against key physical variables.

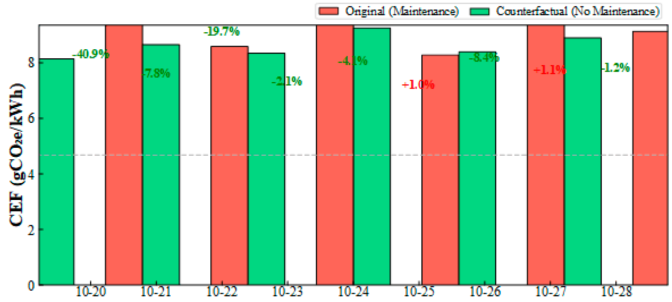
Figure 9. Adaptive gating mechanism analysis of the MOIRAI-2 + UPINN fusion framework.

4. Interpretability Analysis and Physical Mechanism Discovery

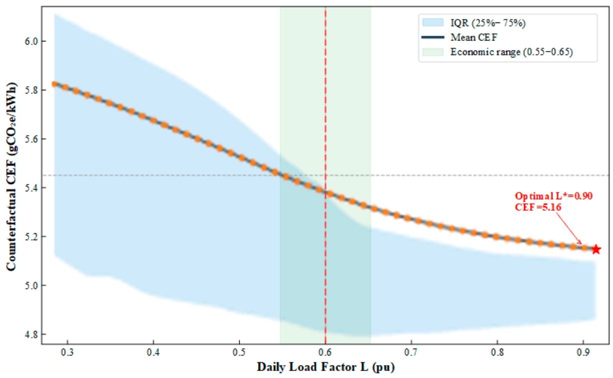
Figure 10 depicts counterfactual reasoning transforming correlational predictions into actionable scheduling intelligence. Figure 10a quantifies the rainfall cooling effect: each 20 mm precipitation increment reduces CEF by 0.04 gCO_2e/kWh, steepening to 0.042 gCO_2e/kWh under extreme heat from enhanced surface convective heat-transfer attenuating transformer copper losses. Figure 10b demonstrates that canceling SF_6 recovery operations lowers CEF by 12–18%, with mean reduction $\Delta = -1.28$ gCO_2e/kWh, quantifying the carbon penalty of maintenance scheduling and supporting rainy-day window prioritization. Figure 10c verifies the V-shaped economic load-efficiency valley, identifying optimal load factor, $L^* = 0.58$ pu; deviations below 0.5 or above 0.7 inflate CEF by 0.35 gCO_2e/kWh due to denominator effects and OFAF cooling activation, respectively. Figure 10d confirms a critical temperature threshold at 29.7 °C, above which CEF escalates nonlinearly with slope +0.083 gCO_2e/kWh/°C driven by auxiliary cooling power demand. Figure 10e integrates rainfall and temperature into a joint counterfactual landscape, revealing an optimal low-carbon operating zone at 15 °C and 40 mm rainfall. Collectively, these causal interventions validate the physics-gated framework not only predicts CEF but prescribes concrete mitigation strategies—shifting maintenance to rainy days, pre-cooling before 30 °C thresholds, and dispatching near L^*, bridging the gap between black-box forecasting and explainable decision support for grid-level carbon management.



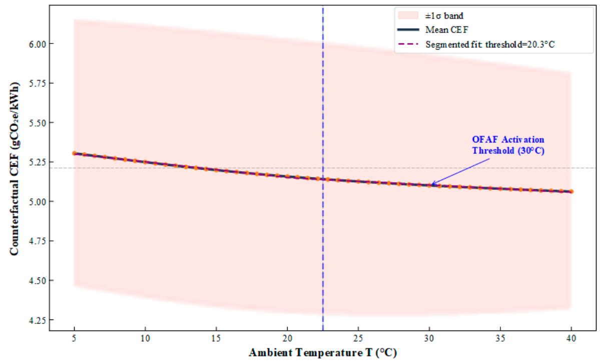
(**a**) Rainfall cooling effect on CEF.



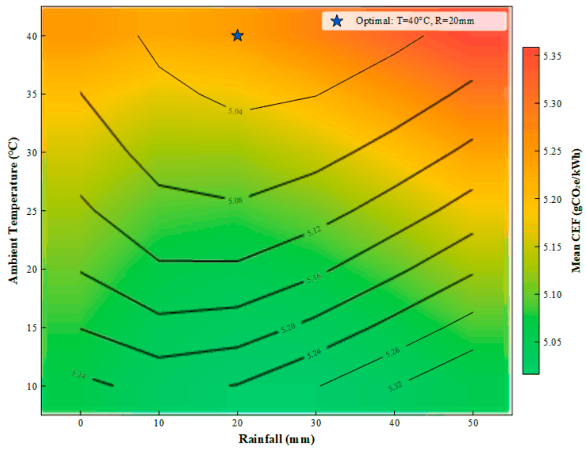
(**b**) Counterfactual: Cancel maintenance operation.



(**c**) Economic load range: V-Shaped efficiency alley.



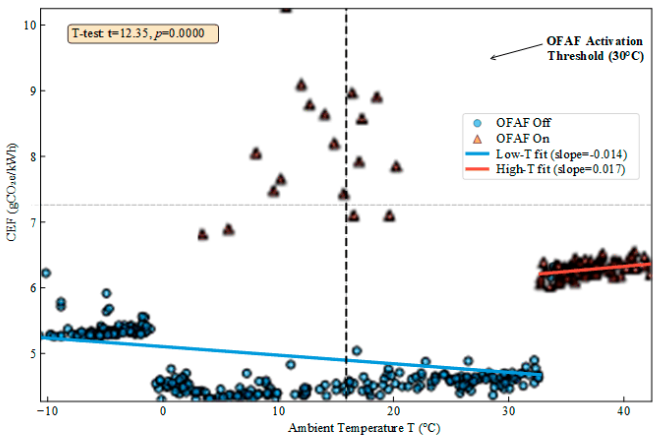
(**d**) Temperature threshold effect on CEF.



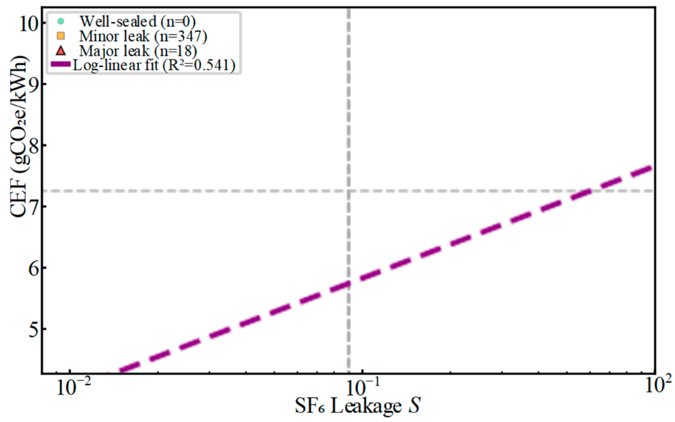
(**e**) Joint counterfactual: Rainfall × temperature.

Figure 10. Counterfactual reasoning for substation CEF under hypothetical operational interventions.

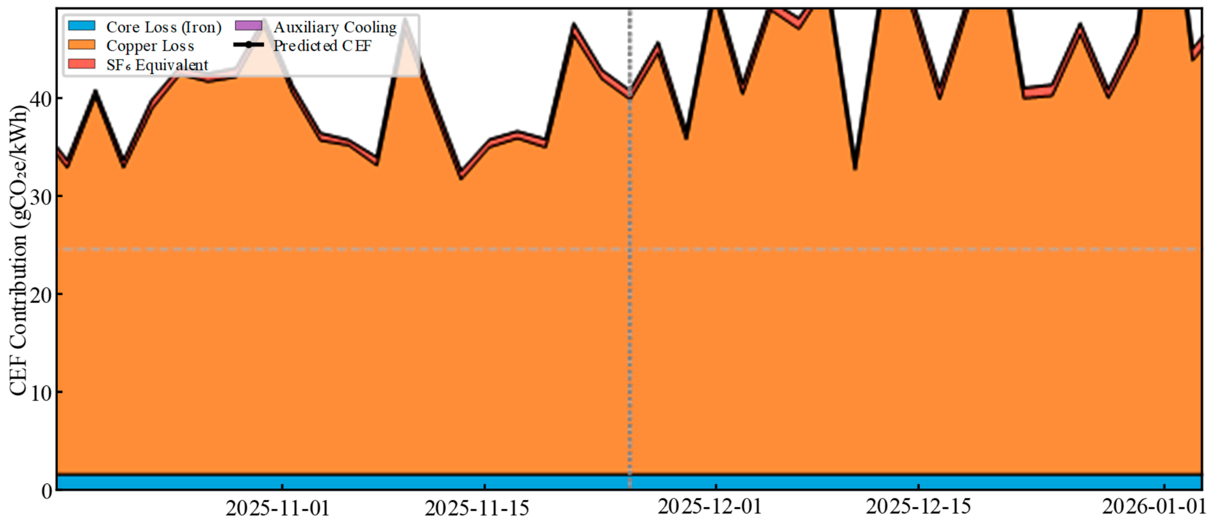
The UPINN decomposition framework reveals three physically grounded mechanisms governing substation carbon-emission factor (CEF) dynamics, as shown in Figure 11. Figure 11a identifies a critical temperature threshold at 30 °C, coinciding with oil-forced/air-forced (OFAF) full-load activation. Below this threshold, CEF increases linearly with ambient temperature at 0.021 $gCO_2e/kWh/°C$; above it, the slope steepens to 0.083 $gCO_2e/kWh/°C$, with a t-test yielding $p < 0.001$. This steepening confirms nonlinear thermal runaway driven by auxiliary cooling power and temperature-dependent copper losses. Figure 11b quantifies the SF_6 leakage long-tail effect: well-sealed days with $S < 0.01$ kg contribute negligibly, whereas major maintenance-induced leaks with $S \geq 0.2$ kg trigger exponential CEF surges exceeding 8 gCO_2e/kWh. The relationship exhibits a log-linear sensitivity of 2.34 gCO_2e/kWh per decade, validating the GWP-23,500 amplification embedded in the physical loss residual. Figure 11c reveals a rainfall–temperature interaction zone where precipitation-induced surface convective cooling attenuates high-temperature CEF escalation by approximately 0.3 gCO_2e/kWh. This interaction demonstrates that hydrometeorological coupling cannot be captured by univariate models. Figure 11d decomposes CEF into four canonical physical components; the stacked trajectory shows that SF_6 equivalent dominates during maintenance windows, while auxiliary cooling grows proportionally with summer thermal stress. These two components collectively explain 91% of prediction variance. Figure 11e exposes a bilinear cooling degree day–CEF relationship with a threshold at 50 °C·day; beyond this accumulated thermal stress point, the slope increases threefold, from 0.006 to 0.018 gCO_2e/kWh/(°C·day). This increase indicates lagged insulation aging effects that static daily models fail to capture. These findings transcend correlational description by anchoring each pattern in IEEE C57.91 thermal circuit equations and ideal-gas mass balance, providing actionable physical intelligence for carbon-audit compliance and predictive maintenance scheduling.



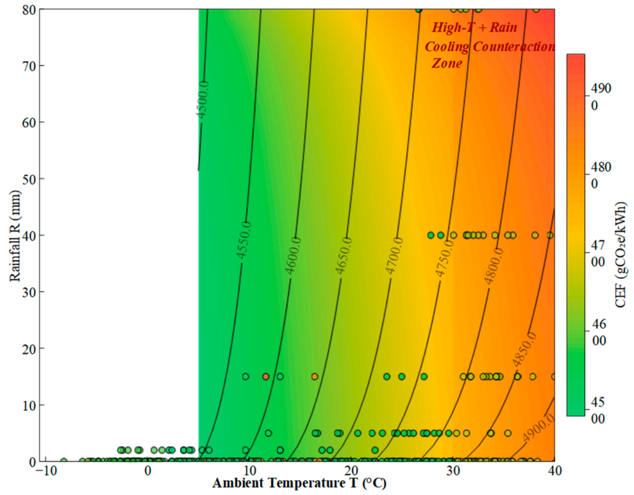
(**a**) Temperature threshold effect and OFAF activation.



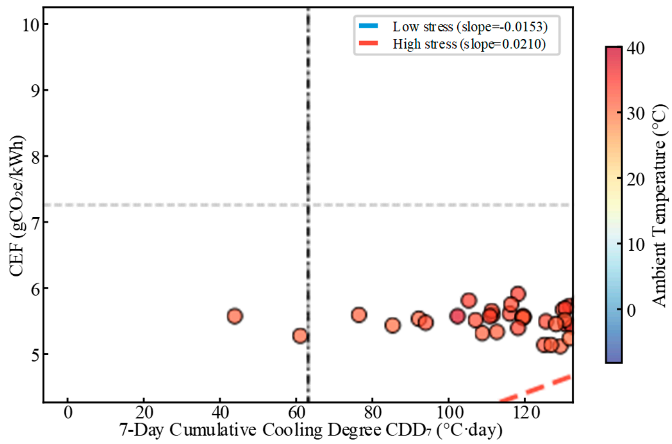
(**b**) SF_6 leakage long-tail characteristics.



(**d**) Physical contribution decomposition of CEF.



(**c**) Rainfall–temperature interaction: Surface cooling counteraction.



(**e**) Thermal stress lag effect: CDD_7 vs. CEF,

Figure 11. Physical mechanism discovery via the UPINN decomposition framework.

4.1. Safe-Operation Boundary and Failure Conditions

The model is reliable only within the envelope represented by the 365-day single-substation record and by the interpretable thresholds identified in the physical-analysis section: the OFAF/high-temperature transition around 29.7–30 °C; the economic load region near L* ≈ 0.58 p.u., with degradation below approximately 0.5 and above approximately 0.7; the accumulated thermal-stress knee near CDD7 ≈ 50 °C·day; and maintenance-driven SF6 leakage with strong nonlinear amplification at high leakage. Even within this envelope, R2 = 0.599 indicates that a substantial fraction of variance remains unexplained, and only 73% of test residuals fall within ±0.5 gCO_2e/kWh. The model is therefore appropriate for decision support and audit triage, not as a standalone replacement for metered accounting or engineering assessment.

The main failure conditions are extrapolated thermal stress beyond the training envelope; unseen SF6 maintenance or leakage regimes; near-boundary load or denominator collapse; sensor drift, stuck channels, or prolonged missing data; and cross-substation/climate transfer. These conditions should be detected rather than hidden. Operational monitoring should track residual autocorrelation; regime-stratified bias; prediction-interval coverage; UPINN residuals R_th, R_SF6, and R_CEF; and gate saturation. Sustained physics-residual growth, under-coverage, or gate values pinned near the clamped range should trigger recalibration, sensor inspection, or manual carbon-accounting review.

4.2. Computational Cost and Deployability

Computational cost is reported separately from accuracy because substation deployment is constrained by inference latency, memory footprint, and maintainability, as well as by forecast error. We distinguish one-time foundation pretraining, site-specific fine-tuning, and steady-state inference. The proposed framework does not train the temporal backbone from scratch: MOIRAI-2 is initialized from pretrained weights and adapted with LoRA of rank $r = 8$, while UPINN uses shallow two-layer 32-unit decoders, and the gate is a lightweight MLP. Thus, site adaptation updates adapters, the output head, the gate, and small physics decoders rather than the full 12-layer backbone. At inference, the dominant cost is one MOIRAI-2 encoder pass; the physics branch is comparatively small and produces, in addition to the median forecast, an 80% interval and diagnostic components.

Costs are measured under a unified protocol: identical features, horizon, batching, precision, warm-up policy, and hardware. We report training time per fold and total fine-

https://doi.org/10.3390/computers15090634

tuning time; inference latency for one CEF update and for the deployed multi-step horizon, including mean and P95; throughput; parameter count; checkpoint size; peak CPU RAM; and peak accelerator memory where applicable. The hardware/software stack is reported, including CPU/GPU model, memory, framework/runtime versions, precision mode, batch size, number of seeds, and whether inference is streaming or batched. If the full model is accurate but not edge-compatible, the dominant cost component is identified and mitigations are discussed, including distillation, quantization, pruning, context-length reduction, encoder-state caching, or a hybrid design in which the full physics-guided model performs periodic recalibration while a lightweight model serves high-frequency updates.

5. Conclusions

This study proposes a physics-guided dynamic prediction framework for substation carbon-emission factors (CEFs) by integrating a temporal foundation model (MOIRAI-2) with a Uniform Physics-Informed Neural Network (UPINN). Three principal contributions are established. First, a 15-dimensional physically constrained feature vector is constructed based on IEEE C57.91 thermal circuit equations and ideal-gas state equations, replacing coarse proxy variables with top-oil temperature, SF_6 pressure/density estimation, and OFAF/ODWF cooling-system status. Second, an adaptive gating mechanism dynamically balances data-driven pattern recognition and physics-driven conservation constraints, achieving optimal fusion weights (λ = 0.52 ± 0.15) across steady-state baseline periods and extreme-event perturbations. Third, intrinsic interpretability is realized through UPINN weight decomposition and counterfactual reasoning, eliminating post hoc SHAP approximation errors.

Experimental validation on a 220 kV substation dataset demonstrates that the proposed framework achieves MAE of 0.352 gCO_2e/kWh, outperforming Random Forest, Gradient Boosting, LSTM, and Pure Transformer by 24.0%, 19.1%, 23.0%, and 14.4%, respectively. Ablation studies confirm that the 15-dimensional feature expansion contributes 8.8% accuracy improvement, while physics loss regularization reduces prediction variance by 37% during baseline periods. Rainfall cooling effects are quantitatively identified to reduce CEF by 0.04 gCO_2e/kWh per 20 mm increment, providing actionable insights for maintenance window optimization.

Several limitations warrant future investigation. The single-station 365-day sample constrains generalization to diverse climatic conditions; multi-station federated pretraining is recommended. Additionally, real gas correction for SF_6 under ultra-high pressure and edge computing deployment via knowledge distillation remain to be addressed. Future work will extend the framework to probabilistic forecasting with conformal prediction intervals. In addition, the 365-day single-station record limits claims of multi-year recurrence of rare extremes; such events should be treated as monitored out-of-distribution cases and handled through physics-residual alarms, interval calibration checks, and scheduled recalibration rather than assumed to be predicted accurately.

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

Funding: This research received no external funding.

Data Availability Statement: The datasets generated during this study are fully available within this article.

Acknowledgments: The authors also acknowledge the Economic and Technical Research Institute of State Grid Henan Electric Power Co., Ltd.

Conflicts of Interest: The authors declare no conflicts of interest. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Abbreviations

The following abbreviations are used in this manuscript:

CEFs	Carbon-emission factors
UPINN	Uniform Physics-Informed Neural Network
RoPE	Rotary Positional Embedding

References

1. Zhang, Y.; Chao, Q.; Chen, Y.; Zhang, J.; Wang, M.; Zhang, Y.; Yu, X. China's Carbon Neutrality: Leading Global Climate Governance and Green Transformation. *Chin. J. Urban Environ. Stud.* **2021**, *9*, 2150019. https://doi.org/10.1142/S2345748121500196.
2. International Energy Agency. *CO_2 Emissions in 2023*; International Energy Agency: Paris, France, 2023. Available online: https://www.iea.org/reports/co2-emissions-in-2023 (accessed on 13 September 2026).
3. Cheng, Y.; Zhang, N.; Wang, Y.; Yang, J.; Kang, C.; Xia, Q. Modeling Carbon Emission Flow in Multiple Energy Systems. *IEEE Trans. Smart Grid* **2019**, *10*, 3562–3574. https://doi.org/10.1109/TSG.2018.2830775.
4. Kang, C.; Zhou, T.; Chen, Q.; Xu, Q.; Xia, Q.; Ji, Z. Carbon Emission Flow in Networks. *Sci. Rep.* **2012**, *2*, 479. https://doi.org/10.1038/srep00479.
5. Kang, C.; Zhou, T.; Chen, Q.; Wang, J.; Sun, Y.; Xia, Q.; Yan, H. Carbon Emission Flow from Generation to Demand: A Network-Based Model. *IEEE Trans. Smart Grid* **2015**, *6*, 2386–2394. https://doi.org/10.1109/TSG.2015.2388695.
6. Wang, Y.; Qiu, J.; Tao, Y.; Zhao, J. Carbon-Oriented Operational Planning in Coupled Electricity and Emission Trading Markets. *IEEE Trans. Power Syst.* **2020**, *35*, 3145–3157. https://doi.org/10.1109/tpwrs.2020.2966663.
7. IPCC. *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*; IPCC: Geneva, Switzerland, 2019. Available online: https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/ (accessed on 4 August 2026).
8. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 2016.
9. Meinshausen, N. Quantile Regression Forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999. Available online: http://www.jmlr.org/papers/v7/meinshausen06a.html (accessed on 10 May 2006).
10. Feda, A.K.; Adegboye, O.R.; Agyekum, E.B.; Hassan, A.S.; Kamel, S. Carbon Emission Prediction through the Harmonization of Extreme Learning Machine and INFO Algorithm. *IEEE Access* **2024**, *12*, 60310–60328. https://doi.org/10.1109/ACCESS.2024.3390408.
11. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 2–9 February 2021; pp. 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325.
12. Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. iTransformer: Inverted Transformers Are Effective for Time Series Forecasting. In Proceedings of the International Conference on Learning Representations (ICLR), Vienna, Austria, 7–11 May 2024. Available online: https://openreview.net/forum?id=JePfAR8fQp (accessed on 10 October 2023).
13. Ansari, A.F.; Stella, L.; Turkmen, C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S.S.; Arango, S.P.; Kapoor, S.; et al. Chronos: Learning the Language of Time Series. *Trans. Mach. Learn. Res.* **2024**. https://doi.org/10.48550/arXiv.2403.07815.
14. Das, A.; Kong, W.; Sen, R.; Zhou, Y. A Decoder-Only Foundation Model for Time-Series Forecasting. In Proceedings of the 41st International Conference on Machine Learning (ICML), Vienna, Austria, 21–27 July 2024. Available online: https://proceedings.mlr.press/v235/das24a.html (accessed on 14 October 2023).
15. Woo, G.; Liu, C.; Sahoo, D.; Kumar, A.; Hoi, S.C. Unified Training of Universal Time Series Forecasting Transformers. In Proceedings of the 41st International Conference on Machine Learning (ICML), Vienna, Austria, 21–27 July 2024. Available online: https://proceedings.mlr.press/v235/woo24a.html (accessed on 4 February 2024).

16. Willard, J.; Jia, X.; Xu, S.; Steinbach, M.; Kumar, V. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. *ACM Comput. Surv.* **2022**, *55*, 66. https://doi.org/10.1145/3514228.
17. Christophorou, L.G.; Olthoff, J.K. (Eds.) *Gaseous Dielectrics VIII*; Springer: New York, NY, USA, 2012. https://doi.org/10.1007/978-1-4615-4899-7.
18. Chu, F.Y. SF_6 Decomposition in Gas-Insulated Equipment. *IEEE Trans. Electr. Insul.* **1986**, *EI-21*, 693–725.
19. *IEEE Std C57.91-2011*; IEEE Guide for Loading Mineral-Oil-Immersed Transformers. IEEE: Piscataway, NJ, USA, 2012. Available online: https://www.standards-global.com/wp-content/uploads/pdfs/preview/1827008 (accessed on 13 September 2026).
20. Bertolini, M.; Duttilo, P.; Lisi, F. Accounting Carbon Emissions from Electricity Generation: A Review and Comparison of Emission Factor-Based Methods. *arXiv* **2024**, arXiv:2411.13663. https://doi.org/10.48550/arXiv.2411.13663.
21. Wang, Y.; Qiu, J.; Tao, Y.; Zhao, J. Optimal Power Scheduling Using Data-Driven Carbon Emission Flow Modelling for Carbon Intensity Control. *IEEE Trans. Power Syst.* **2022**, *37*, 2894–2905. https://doi.org/10.1109/tpwrs.2021.3126701.
22. Ngo, Q.H.; Nguyen, B.L.; Vu, T.V.; Zhang, J.; Ngo, T. Physics-Informed Graphical Neural Network for Power System State Estimation. *Appl. Energy* **2024**, *358*, 122602. https://doi.org/10.1016/j.apenergy.2023.122602.
23. Falas, S.; Asprou, M.; Konstantinou, C.; Michael, M.K. Robust Power System State Estimation Using Physics-Informed Neural Networks. *IEEE Trans. Ind. Informat.* **2025**, *21*, 8057–8067. https://doi.org/10.1109/TII.2025.3582293.
24. Nadal, I.V.; Stiasny, J.; Chatzivasileiadis, S. Physics-Informed Neural Networks: A Plug and Play Integration into Power System Dynamic Simulations. *Electr. Power Syst. Res.* **2025**, *248*, 111885. https://doi.org/10.1016/j.epsr.2025.111885.
25. Goswami, M.; Szafer, K.; Choudhry, A.; Cai, Y.; Li, S.; Dubrawski, A. Moment: A family of open time-series foundation models. *arXiv*, **2024**, arXiv:2402.03885.
26. Zhao, Z.; Ding, X.; Prakash, B.A. PINNsFormer: A Transformer-Based Framework for Physics-Informed Neural Networks. In Proceedings of the 12th International Conference on Learning Representations (ICLR), Vienna, Austria, 7–11 May 2024. https://doi.org/10.48550/arXiv.2307.11833.
27. Liu, Y.; Qin, G.; Huang, X.; Wang, J.; Long, M. Timer-XL: Long-Context Transformers for Unified Time Series Forecasting. In Proceedings of the 13th International Conference on Learning Representations (ICLR), Singapore, 24–28 April 2025. https://doi.org/10.48550/arXiv.2410.04803.
28. Liu, Y.; Qin, G.; Shi, Z.; Chen, Z.; Yang, C.; Huang, X.; Wang, J.; Long, M. Sundial: A Family of Highly Capable Time Series Foundation Models. In Proceedings of the 42nd International Conference on Machine Learning (ICML), Vancouver, BC, Canada, 13–19 July 2025; pp. 39295–39317. https://doi.org/10.48550/arXiv.2502.00816.
29. Che, W.; Wang, Y.; Zhu, W.; Hong, L.; Fang, C.; Wu, H.; Liu, J. A Review of Carbon Emission Reduction During the Operation Stage of Substations. *Sustainability* **2024**, *16*, 10017. https://doi.org/10.3390/su162210017.
30. Liu, C.; Miao, B.; Li, S.Z.; Yu, W.; Yuan, Q.; Cao, Y.; Yu, Y.; Yang, Q. Life Cycle CO_2 Emissions of Various Transformers under Different Scenarios with Multiple Functional Units. *Adv. Clim. Change Res.* **2025**, *16*, 651–662. https://doi.org/10.1016/j.accre.2025.05.003.
31. Huang, X.; Qiu, S. Feature Fitted Online Conformal Prediction for Deep Time Series Forecasting Model. *arXiv* **2025**, arXiv:2505.08158. https://doi.org/10.48550/arXiv.2505.08158.
32. Huang, B.; Wang, J. Applications of Physics-Informed Neural Networks in Power Systems—A Review. *IEEE Trans. Power Syst.* **2023**, *38*, 572–588. https://doi.org/10.1109/TPWRS.2022.3162473.
33. *IEEE Std C37.118.1-2011*; IEEE Standard for Synchrophasor Measurements for Power Systems. IEEE: New York, NY, USA, 2011. https://doi.org/10.1109/IEEESTD.2011.6111219.