processes

MDPI

Article

Market-Informed Multi-Step Forecasting of Intraday Net Load with Operational Risk Screening

Weiping Zhu [1], Zhihan Jiang [2,*], Haiqing Huang [2], Haoran Wu [2], Jiangyu Zhu [2], Jiarui Li [3] and Zhengsen Ji [3]

[1] State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China; zhuwp_jsgrid@163.com
[2] State Grid Zhenjiang Power Supply Company, Zhenjiang 212000, China; huanghq_jsgrid@163.com (H.H.); wuhr_jsgrid@163.com (H.W.); zhujy_jsgrid@163.com (J.Z.)
[3] School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China; ljr787903@163.com (J.L.); jzs0686@sdut.edu.cn (Z.J.)
* Correspondence: jiangzh_jsgrid@163.com

Abstract

The increasing penetration of wind and photovoltaic generation has intensified net-load variability and increased the difficulty of intraday forecasting and operational risk identification. This paper proposes a market-informed multi-step net-load forecasting and risk-screening framework based on Chronos2-Small. Low-rank adaptation (LoRA) is employed for parameter-efficient fine-tuning, while historical net load, meteorological variables, system operating conditions, day-ahead prices, historical real-time prices, and temporal features are jointly incorporated. The model outputs P10, P50, and P90 quantiles for 4, 8, and 12 h forecasting horizons. Experiments on 8544 15 min observations from the Jiangsu power system show that the proposed model achieves nMAE values of 2.72%, 4.22%, and 4.45% for the three horizons, respectively. Across three seeds, MAE is reduced by 37.23%, 33.02%, and 35.46% relative to covariate-extended PatchTST. SHAP analysis shows that the relative importance of day-ahead market signals among the five exogenous groups increases from 32.01% to 65.21% as the forecasting horizon extends. A three-level operational risk mechanism is further constructed using the relative P10–P90 interval width and P50 net-load ramp, with day-ahead prices providing an independent market-sensitive label. The Watch and High states capture 95.41–97.58% of independently defined large-ramp events, with a precision of 23.69–26.41%. Pinball loss, interval score, coverage, and width are used to assess probabilistic performance.

Keywords: net-load forecasting; time-series foundation model; SHAP analysis; operational risk management

Check for updates

Academic Editor: Giancarlo Cravotto

Received: 27 September 2026
Revised: 2 November 2026
Accepted: 5 November 2026
Published: 9 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

1.1. Background and Motivation

The increasing penetration of wind and photovoltaic (PV) generation is reshaping power system operation from conventional load-following toward net-load balancing. Renewable intermittency and forecasting errors increase balancing, ramping, and reserve requirements, while high renewable penetration can produce pronounced peak–valley variations and rapid net-load ramps [1,2]. Net load, commonly defined as system load minus renewable generation, therefore provides a direct representation of the effective demand that must be balanced by conventional generation and other flexible resources.

Previous studies have demonstrated the value of jointly considering load and renewable generation within a net-load forecasting framework. Copula-based and probabilistic ap-

proaches have been developed to characterize aggregated net-load uncertainty [3,4], while recent studies have addressed PV-induced peak–valley variations, distribution-level net-load forecasting, and probabilistic forecasting for renewable-integrated energy systems [5–7]. These studies confirm that accurate net-load forecasting is increasingly important for maintaining operational flexibility under high renewable penetration.

Net-load dynamics are influenced not only by historical power measurements but also by weather conditions, temporal patterns, and market operation. The integration of historical observations, meteorological variables, and calendar information has been shown to improve short-term load forecasting performance [8]. However, deterministic forecasts alone provide limited information about future uncertainty. Probabilistic forecasting, particularly quantile-based forecasting, can provide both a central estimate and an uncertainty range, which is more suitable for risk-aware system operation [9].

Electricity market information may offer an additional source of predictive information. Day-ahead and real-time prices reflect the combined effects of supply–demand balance, renewable output, generation costs, system scarcity, and market behavior. Therefore, incorporating market information together with net-load history and weather variables may improve the representation of short-term system states. For intraday forecasting, day-ahead prices are assumed to be available in advance, while real-time prices observed before the forecasting origin can be used as historical market variables without introducing future information leakage.

1.2. Literature Review and Research Gap

Load and net-load forecasting methods have evolved from conventional statistical approaches toward machine learning, deep learning, and probabilistic forecasting models. Quantile-combination methods can reconstruct predictive distributions from multiple quantile forecasts, while nonparametric approaches improve robustness under complex operating conditions [10,11]. Bayesian mixture density networks have further been used to explicitly represent forecasting uncertainty [12]. In deterministic and multi-step forecasting, recurrent and convolutional neural networks remain widely used because of their ability to extract nonlinear temporal features, while hybrid CNN–GRU models have demonstrated good performance in short-term load forecasting [13]. Nevertheless, these models are generally trained from scratch for a specific dataset and may become sensitive to data quantity and representativeness when only limited regional observations are available. WaveKoop-KAN [14] integrates wavelet decomposition, Koopman dynamics, and gated temporal convolutions through Kolmogorov–Arnold fusion, while BiKoop [15] models asymmetric latent dynamics using smoothly gated Koopman operators. These studies advance univariate forecasting architectures. This study investigates parameter-efficient adaptation for multivariate net-load forecasting with market information and operational risk screening.

The relationship between electricity market conditions and load dynamics has also attracted increasing attention. Price volatility regimes can be associated with changing load patterns and increased non-stationarity [16], while extensive electricity price forecasting research has shown that market prices contain information related to load, renewable generation, supply conditions, and market structure [17]. However, most existing net-load forecasting studies primarily rely on physical variables, including historical net load, weather, and renewable generation. The incremental value of day-ahead prices and historically observed real-time prices for net-load forecasting has received considerably less attention, particularly across different forecasting horizons.

Time-series foundation models provide a new approach to forecasting when task-specific data are limited. Chronos demonstrated that large-scale pretraining on hetero-

geneous time-series datasets can provide transferable representations for zero-shot and probabilistic forecasting [18]. Recent studies on time-series foundation models further indicate that pretraining and downstream adaptation can reduce dependence on task-specific training and improve transferability across datasets [19]. However, the application of such models to regional net-load forecasting remains limited. In particular, parameter-efficient adaptation of pretrained models using only a few months of power-system data has not been sufficiently investigated. This motivates the use of lightweight fine-tuning methods such as low-rank adaptation (LoRA), which can preserve pretrained temporal knowledge while adapting the model to local net-load, weather, and market characteristics.

Interpretability represents another important issue. Although foundation models offer strong nonlinear representation capability, their forecasting mechanisms are less transparent than those of conventional models. SHAP-based methods have been applied to quantify the contributions of temporal and exogenous variables in multi-step load forecasting [20], and recent research has emphasized the importance of translating explainable AI results into information that can support operational decision-making [21]. Nevertheless, existing studies rarely combine market-informed net-load forecasting, parameter-efficient foundation model adaptation, feature-level interpretation, probabilistic uncertainty, and lightweight operational risk assessment within a unified framework.

1.3. Contribution

The main contributions of this paper are as follows:

(1) A multi-step net-load forecasting framework based on the Chronos-2-Small time-series foundation model is developed. LoRA is employed for parameter-efficient adaptation under limited historical data, enabling 16-, 32-, and 48-step forecasting with P10, P50, and P90 quantile outputs.

(2) Electricity market information is incorporated into the forecasting framework together with historical net load, meteorological, and temporal variables. Ablation experiments and SHAP analysis are conducted to quantify the contributions of market and exogenous variables across different forecasting horizons.

(3) A lightweight operational risk management mechanism is constructed using the relative P10–P90 interval width and P50 net-load ramping. Three operational risk levels and an independent market-sensitive label based on day-ahead prices are introduced to translate forecasting results into practical decision-support information.

1.4. Paper Organization

The remainder of this paper is organized as follows. Section 2 presents the proposed methodology, including the formulation of the multi-step intraday net-load forecasting problem, the Chronos2-Small forecasting framework, LoRA-based parameter-efficient fine-tuning, grouped masked SHAP analysis, and the operational risk identification mechanism based on prediction interval width and net-load ramping. Section 3 presents the case study and results using data from the Jiangsu power system, including the dataset and experimental settings, benchmark comparisons, analysis of the contribution of market information, SHAP-based interpretation, and operational risk identification under different forecasting horizons. Section 4 discusses the findings in relation to prior work. Section 5 summarizes the main findings and outlines directions for future research. Figure 1 presents the overall workflow of the proposed framework.

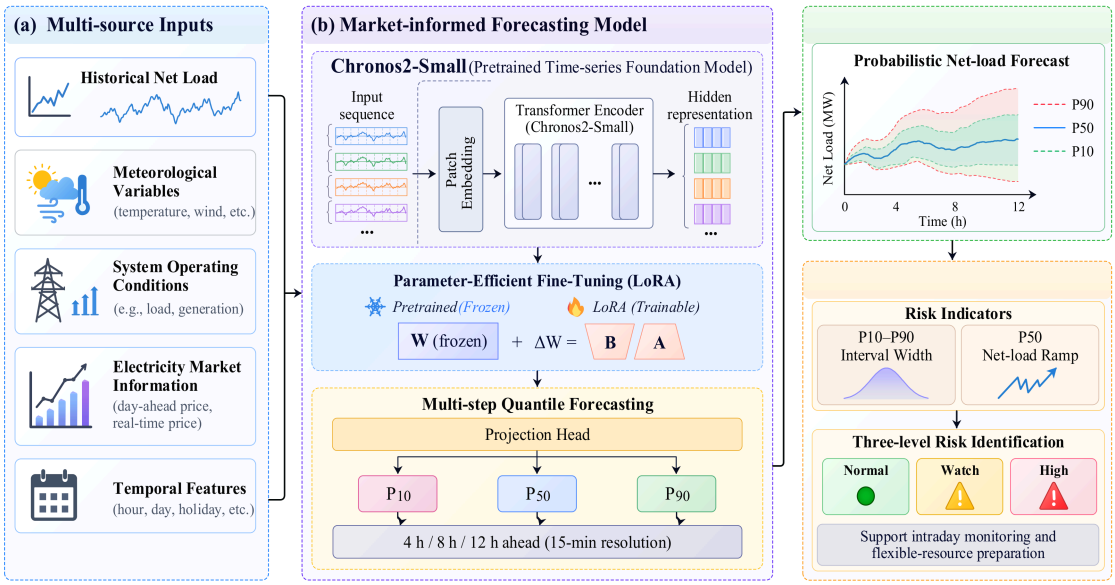


Figure 1. Research flowchart.

2. Methodology

2.1. Problem Formulation

This study investigates multi-step intraday probabilistic net-load forecasting at a 15 min resolution. Let y_t denote the net load at time t. At each forecast origin t, the model uses information from the previous L time steps to forecast the net load over the next H time steps.

The historical inputs include net load, meteorological variables, system operating conditions, time-related variables, and day-ahead and historical real-time prices:

$$X_t^{past} = \left\{ y_{t-L+1:t}, w_{t-L+1:t}, o_{t-L+1:t}, p^{DA}_{t-L+1:t}, p^{RT}_{t-L+1:t}, c_{t-L+1:t} \right\} \tag{1}$$

where $\mathbf{w}$, $\mathbf{c}$, and $\mathbf{p}^{\mathrm{DA}}$ denote the meteorological variables, time-related variables, and day-ahead prices, respectively. Here, o denotes system operating variables and p^{RT} denotes historical real-time prices.

Assuming day-ahead prices are available before the forecast origin, and the sine and cosine encodings of the hour can be calculated in advance, these variables are treated as known future covariates:

$$\mathbf{X}_t^{\mathrm{future}} = \left\{ \mathbf{p}^{\mathrm{DA}}_{t+1:t+H}, \mathbf{c}_{t+1:t+H} \right\} \tag{2}$$

The forecasting task is formulated as follows:

$$\hat{P}_{\tau,t+1:t+H} = f_\theta\left(\mathbf{X}_t^{\mathrm{past}}, \mathbf{X}_t^{\mathrm{future}} \right), \qquad \tau \in \{0.1, 0.5, 0.9\} \tag{3}$$

where $\hat{P}_{\tau,t+1:t+H}$ denotes the predicted net-load quantile over the next H time steps.

2.2. Chronos2-Small Forecasting Framework

Chronos2-Small is a pretrained time-series foundation model with approximately 28 million parameters. It employs a patch-based Transformer encoder to jointly process the target series, historical covariates, and known future covariates [22].

For each variable j, instance normalization is performed as follows:

$$\tilde{x}_t^{(j)} = \operatorname{arcsinh}\left(\frac{x_t^{(j)} - \mu_j}{\sigma_j} \right) \tag{4}$$

where μ_j and σ_j are calculated from the historical window. The normalized sequence is divided into non-overlapping patches of length P = 16. Each patch is combined with its time encoding and observation mask and then mapped into a hidden representation:

$$\mathbf{e}_{j,k} = \phi\left(\left[\tilde{\mathbf{x}}_{j,k}, \mathbf{r}_k, \mathbf{m}_{j,k}\right]\right) \tag{5}$$

The encoder consists of six blocks, each containing temporal self-attention, group self-attention, and a feed-forward network:

$$\begin{cases} \mathbf{H}^T = \mathbf{H} + \mathrm{MSA}_T(\mathrm{LN}(\mathbf{H})) \\ \mathbf{H}^G = \mathbf{H}^T + \mathrm{MSA}_G(\mathrm{LN}(\mathbf{H}^T)) \\ \mathbf{H}^{\mathrm{out}} = \mathbf{H}^G + \mathrm{FFN}(\mathrm{LN}(\mathbf{H}^G)) \end{cases} \tag{6}$$

Temporal self-attention captures dependencies along the time dimension, whereas group self-attention models interactions between net load and its covariates. Day-ahead price, hour sine, and hour cosine are incorporated as known future covariates [23].

Finally, the future hidden representations are mapped to quantile forecasts, of which three are retained for risk screening:

$$\{\hat{\mathbf{P}}_{10}, \hat{\mathbf{P}}_{50}, \hat{\mathbf{P}}_{90}\} = \psi\left(\mathbf{H}^{\mathrm{out}}_{\mathrm{future}}\right) \tag{7}$$

P50 is used as the point forecast, while P10 and P90 define the nominal 80% prediction interval. The native head produces 13 quantiles (0.01, 0.05, 0.10, 0.20, ..., 0.90, 0.95, 0.99), of which nine levels from P10 to P90 are evaluated. The 16-, 32-, and 48-step forecasts are generated directly in a single forward pass.

Context normalization uses historical window means and standard deviations; zero scales are replaced by 10^{-5}. Known future inputs and target labels reuse the context scale. Internal masks indicate unavailable future covariates. The architecture of Chronos2-Small is shown in Figure 2.

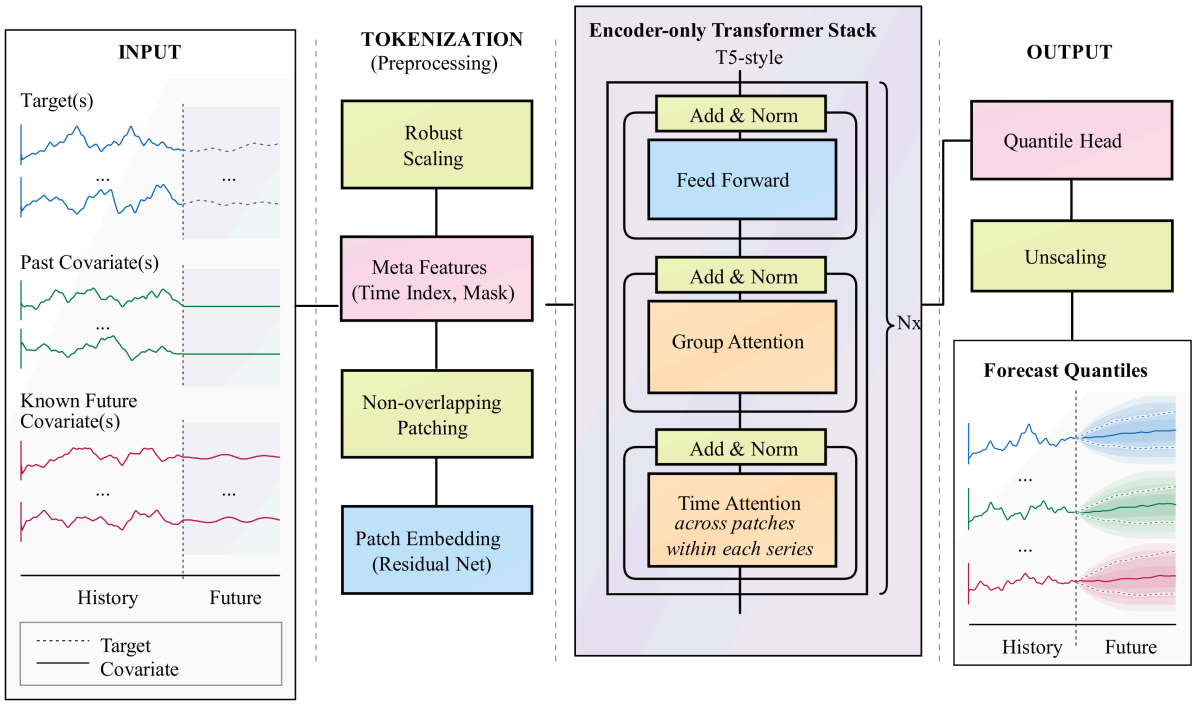


Figure 2. Chronos2-Small model structure.

2.3. LoRA Fine-Tuning

Low-rank adaptation (LoRA) is employed to adapt Chronos2-Small to net-load forecasting without updating all model parameters [24]. For a pretrained weight matrix **W**, LoRA represents its update using two low-rank matrices:

$$\mathbf{W}' = \mathbf{W} + \frac{\alpha}{r}\mathbf{BA} \tag{8}$$

where **W** remains frozen, while **A** and **B** are trainable. The parameters r and α denote the rank and scaling factor, respectively. For column-vector inputs, A has size r × d_in and B has size d_out × r. Each forecast horizon and seed uses an independently fitted adapter.

LoRA is applied to the query, key, value, and output projections—W_Q, W_K, W_V, and W_O—in both the temporal and group self-attention layers [25]. The feed-forward networks and the remaining pretrained parameters are frozen. This allows the model to adapt its temporal and cross-variable representations with a small number of trainable parameters [26].

The model is optimized using the quantile loss:

$$\mathcal{L} = \frac{1}{NH}\sum_{i=1}^{N}\sum_{h=1}^{H}\sum_{\tau\in\mathcal{T}}\rho_\tau\left(y_{i,t+h} - \hat{P}_{\tau,i,t+h}\right) \tag{9}$$

Here, N is the number of training samples, H is the forecast horizon, and $\mathcal{L}$ is the set of quantile levels. After fine-tuning, P10, P50, and P90 are retained for point and interval forecasting. Training uses all 13 native levels. The implementation sums twice the pinball loss over quantiles and averages over forecast steps and channel series, with unavailable labels masked; only net-load labels contribute. The LoRA mechanism is illustrated in Figure 3.

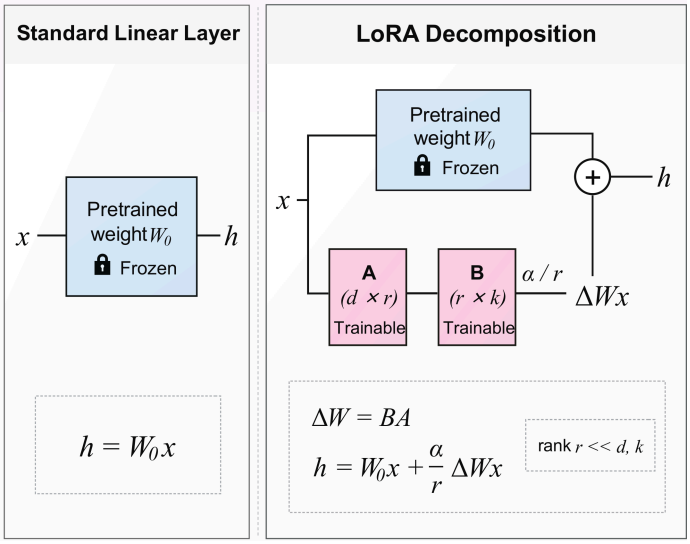


Figure 3. LoRA schematic. Under the column-vector convention of Equation (8), the factor dimensions are A: r × k and B: d × r.

2.4. Grouped Masked SHAP Analysis

A grouped masked SHAP method is developed to quantify the contributions of different covariate types to net-load forecasts. Variables with similar physical meanings or strong correlations are grouped together, which reduces attribution dispersion and redundancy in individual feature explanations.

SHAP provides additive feature attributions based on Shapley values by evaluating the marginal contribution of an input feature across different feature coalitions [27]. For

correlated or physically related variables, feature dependence may distort individual attributions, motivating aggregation or joint treatment of related features [28]. Recent work on explainable time-series Transformers has further shown that grouping and masking exogenous variables can substantially reduce the coalition space while providing meaningful local and global explanations for electricity load and price forecasting [29].

Let G denote the set of covariate groups and (S ⊆ G) a retained subset. Groups excluded from S are replaced by reference values, producing the masked input:

$$\mathbf{X}^{(S)} = \text{Mask}(\mathbf{X}, S, \mathbf{B}) \tag{10}$$

where B is the reference baseline constructed from the training set. Historical net load remains unchanged for all subsets to ensure a consistent forecasting origin. B is the training set mean for the corresponding 15 min time-of-day slot; mean/median sensitivity is evaluated. Clock encodings use their global training mean or median. Historical and known future values are replaced at their respective timestamps; no future values of past-only variables are supplied.

For forecast step h, the model output for subset S is defined as follows:

$$v_h(S) = f_{\theta,h}(\mathbf{X}^{(S)}) \tag{11}$$

where $f_{\theta,h}$ denotes the P50 forecast at step h. The SHAP value of group g is calculated as follows:

$$\phi_{g,h} = \sum_{S \subseteq G \setminus \{g\}} \frac{|S|!(M-|S|-1)!}{M!} [v_h(S \cup \{g\}) - v_h(S)] \tag{12}$$

where M is the number of feature groups. This value represents the average marginal contribution of group g across all possible feature subsets.

Conditional on the fixed historical net load y_hist, the group contributions satisfy the additive relation:

$$\hat{P}_{50,h}(y_{hist}, X) = v_h(\emptyset; y_{hist}) + \sum\nolimits_{g \in G} \varphi_{g,h}(y_{hist}) \tag{13}$$

where $v_h(\emptyset)$ is the baseline forecast with all covariate groups masked. A positive SHAP value increases the predicted net load, whereas a negative value decreases it. Global group importance is obtained by averaging the absolute SHAP values across the evaluated samples. Attributions are aligned to delivery timestamps and pooled over three seeds before normalization within the five exogenous groups. The resulting shares describe model reliance conditional on net-load history, not causal effects or shares of the whole forecast.

2.5. Operational Risk Identification

Rapid net-load ramps constitute an important operational challenge in systems with high renewable penetration, and their frequency and characteristics can differ substantially from those of conventional load ramps [30]. Meanwhile, forecast uncertainty expressed through quantiles or prediction intervals can provide useful information for system risk awareness and has been applied to the estimation and procurement of ramping reserves under renewable uncertainty [31]. Accordingly, this study combines forecast interval width and predicted net-load ramping to construct a lightweight operational risk-screening mechanism.

The quantile forecasts are converted into operational risk information using forecast uncertainty and net load ramping.

Forecast uncertainty is measured using the relative width of the P10–P90 interval:

$$U_{t,h} = \frac{\hat{P}_{90,t+h} - \hat{P}_{10,t+h}}{max\left(\left|\hat{P}_{50,t+h}\right|, \varepsilon\right)} \tag{14}$$

where ε is a small constant. A larger $U_{t,h}$ indicates greater forecast uncertainty. The implementation uses ε = 1 MW.

The absolute predicted net load ramp is defined as follows:

$$R_{t,h} = \begin{cases} \left|\hat{P}_{50,t+1} - y_t\right|, & h = 1, \\ \left|\hat{P}_{50,t+h} - \hat{P}_{50,t+h-1}\right|, & h > 1, \end{cases} \tag{15}$$

where y_t is the latest observed net load before the forecast origin. This definition avoids artificial ramps between consecutive forecast windows.

Let $U_{t,h}$ and $R_{t,h}$ denote the uncertainty and ramp thresholds for forecast horizon H, respectively. Both thresholds are calibrated using the validation set. The operational risk level is defined as follows:

$$L_{t,h} = \mathbb{I}\left(U_{t,h} \geq U_{\text{th}}^{(H)}\right) + \mathbb{I}\left(R_{t,h} \geq R_{\text{th}}^{(H)}\right) \tag{16}$$

The resulting states are:
- $L_{t,h} = 0$: Normal, with low uncertainty and a small ramp;
-$L_{t,h} = 1$: Watch, with either high uncertainty or a large ramp;
-$L_{t,h} = 2$: High, with both high uncertainty and a large ramp.
Day-ahead prices are further used to identify market-sensitive risk periods:

$$M_{t,h} = \mathbb{I}(L_{t,h} \geq 1)\mathbb{I}(p_{t+h}^{\text{DA}} \geq p_{\text{th}}^{\text{DA}}) \tag{17}$$

where $p_{\text{th}}^{\text{DA}}$ is calibrated from the validation set. Day-ahead prices do not alter the physical risk level but indicate the priority of operational attention and flexible resource preparation.

Normal periods may warrant routine monitoring, Watch periods may warrant enhanced monitoring and resource availability checks, and High periods may warrant advance preparation of ramping and reserve resources. Market-sensitive periods receive additional operational priority.

3. Case Study and Results

3.1. Dataset and Experimental Settings

A measured multivariate net load dataset from the Jiangsu power system was used at a 15 min resolution. It covered the period from 1 January to 30 March 2026 and contained 8544 observations. The data were divided chronologically to prevent information leakage. The chronological division of the dataset is summarized in Table 1.

Table 1. Dataset description.

Dataset	Period	Observations	Proportion
Training	1 January–1 March 2026	5760	67.4%
Validation	2–16 March 2026	1440	16.9%
Test	17–30 March 2026	1344	15.7%

Net load was used as the forecasting target. Day-ahead price, hour sine, and hour cosine were treated as known future covariates, while the remaining variables were available only over the historical context. The test set covered 17–30 March 2026, comprising 1344 observations over 14 days.

All variables were aligned to a continuous 15 min grid. No duplicate timestamps, missing values, or non-finite inputs were found, and no imputation or outlier clipping was used. Time encodings used hour + minute/60. The units and descriptive statistics of the variables are presented in Table 2.

Table 2. Input variables, units, and descriptive statistics.

Variable	Unit	Input Type	Mean ± SD	Minimum	Maximum
Net load	MW	Historical/target	75,419.90 ± 20,292.20	12,681.65	121,465.65
Temperature at 2 m	°C	Historical	8.14 ± 4.52	−4.32	23.78
Wind speed at 100 m	m/s	Historical	5.29 ± 1.91	0.99	12.74
Solar radiation	W/m^2	Historical	77.85 ± 145.14	0.00	739.08
Gas-turbine output	MW	Historical	4711.15 ± 2085.27	990.66	14,435.35
Received power	MW	Historical	14,245.86 ± 4676.75	0.00	27,873.10
Day-ahead price	CNY/MWh	Historical + future	295.66 ± 41.59	200.00	702.00
Real-time weighted-average price	CNY/MWh	Historical	297.44 ± 55.97	200.00	985.80
Hour sine	1	Historical + future	0.00 ± 0.71	−1.00	1.00
Hour cosine	1	Historical + future	0.00 ± 0.71	−1.00	1.00

The statistics summarize all 8544 observations from 1 January to 30 March 2026. SD denotes the sample standard deviation. CNY denotes Chinese yuan; unit 1 denotes a dimensionless variable. Historical inputs end before the forecast window, while future inputs are known over that window.

The benchmark includes Persistence, LightGBM, covariate-extended PatchTST, zero-shot Chronos2-Small, and Chronos2-Small LoRA. The trainable models share the same data splits, eligible training origins, test forecast origins, and permitted input variables. Future covariates comprise only day-ahead prices and time encodings. All models and horizons are evaluated using six configurations combining two context lengths (336 and 672) with three parameter settings. Configurations are selected by validation MAE under model-specific early-stopping rules and independently trained with seeds 2026, 2027, and 2028. Selection uses seed 2025; final statistics are mean ± sample SD across the three independent seeds, without ensembling. Tables A1 and A2 give the search space, stopping rules, and selected settings.

LightGBM uses lagged and summary features, while the PatchTST extension also incorporates known future covariates. Training origins start after 672 observations with stride four and labels confined to training; non-overlapping test windows cover all 1344 test timestamps. Zero-shot Chronos2-Small uses the selected LoRA context.

Point forecasting performance is evaluated using R^2, MAE, RMSE, and nMAE. The latter divides MAE by mean observed test net load.

$$\begin{cases} R^2 = 1 - \dfrac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2} \\ \mathrm{MAE} = \dfrac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \\ \mathrm{RMSE} = \sqrt{\dfrac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \\ \mathrm{nMAE} = \dfrac{\mathrm{MAE}}{\frac{1}{N} \sum_{i=1}^{N} |y_i|} \times 100\% \end{cases} \tag{18}$$

Probabilistic performance is evaluated using pinball loss $\varrho\tau(e) = \max(\tau e, (\tau - 1)e)$, averaged over P10–P90, and the 80% interval score $IS_{0.2} = (u - l) + 10 \max(l - y,0) + 10 \max(y - u,0)$, where l = P10 and u = P90 [32]. Coverage and mean width assess calibration and sharpness; lower scores are better.

For empirical calibration, the validation scores $\max(P10 - y, y - P90)$ determine a nonnegative expansion d using their $\min(n, \lceil 0.8(n + 1) \rceil)$-th order statistic (n = 1440). Test intervals become [P10 − d, P90 + d], with interior quantiles unchanged.

For SHAP analysis, the covariates were divided into five groups, and all 32 possible coalitions were evaluated. Every fourth test point was selected, resulting in 336 explained

samples for each forecast horizon. The uncertainty, ramp, and day-ahead price thresholds used for risk identification were determined from the 75th percentiles of the validation set.

Paired 95% intervals used 5000 circular moving-block resamples of aligned, seed-averaged absolute errors, with 24 and 48 h blocks. These intervals described uncertainty conditional on the fitted models and test period (Table A5).

3.2. Benchmark Comparison

Table 3 reports the benchmark results, with the trainable models evaluated across three random seeds. Chronos2-Small LoRA achieved the lowest mean MAE among the evaluated models, reducing MAE relative to covariate-extended PatchTST by 37.23%, 33.02%, and 35.46% at the 4, 8, and 12 h horizons, respectively.

Table 3. Point forecasting performance of benchmark model.

Model	Horizon (Steps)	R^2	MAE (MW)	RMSE (MW)	nMAE (%)
Persistence	16	0.2556	7640.72	12,766.34	9.83
LightGBM	16	0.8265 ± 0.0032	4506.69 ± 85.83	6162.78 ± 57.06	5.80 ± 0.11
PatchTST + covariates	16	0.8857 ± 0.0220	3369.60 ± 283.04	4986.89 ± 471.04	4.34 ± 0.36
Chronos2-Small zero-shot	16	0.9458	2179.16	3445.11	2.80
Chronos2-Small LoRA	16	**0.9502 ± 0.0005**	2115.22 ± 24.01	3301.96 ± 17.74	**2.72 ± 0.03**
Persistence	32	0.2556	7640.72	12,766.34	9.83
LightGBM	32	0.5715 ± 0.0191	6729.18 ± 101.73	9684.59 ± 214.97	8.66 ± 0.13
PatchTST + covariates	32	0.7997 ± 0.0109	4897.43 ± 180.31	6621.37 ± 180.05	6.30 ± 0.23
Chronos2-Small zero-shot	32	0.8854	3311.76	5008.69	4.26
Chronos2-Small LoRA	32	**0.8874 ± 0.0026**	3280.20 ± 49.19	4964.86 ± 56.42	**4.22 ± 0.06**
Persistence	48	0.2556	7640.72	12,766.34	9.83
LightGBM	48	0.5780 ± 0.0151	6920.86 ± 59.33	9610.51 ± 171.68	8.91 ± 0.08
PatchTST + covariates	48	0.7138 ± 0.0206	5359.71 ± 242.65	7912.84 ± 285.99	6.90 ± 0.31
Chronos2-Small zero-shot	48	0.8525	3623.75	5682.87	4.66
Chronos2-Small LoRA	48	**0.8758 ± 0.0058**	3459.42 ± 19.82	5214.42 ± 122.69	**4.45 ± 0.03**

Compared to covariate-extended PatchTST, Chronos2-Small LoRA reduced MAE by 37.23%, 33.02%, and 35.46% at the three horizons. The corresponding reductions relative to LightGBM were 53.06%, 51.25%, and 50.01%. Persistence produced the largest errors, showing that a fixed daily-lag rule cannot adequately capture net load variations caused by changing weather, operating, and market conditions.

Chronos2-Small also showed strong zero-shot forecasting capability. LoRA fine-tuning reduced its mean MAE by 2.93%, 0.95%, and 4.53%, respectively. The paired intervals include zero at all horizons, so the adaptation gain is not statistically established on this test period (Table A5).

Figure 4 compares point forecasting metrics across horizons. The prediction curves in Figure A1 illustrate the corresponding forecasts over 21–23 March 2026 using seed 2026.

Persistence metrics remained unchanged across the three horizons because each test point was predicted using the net load observed at the same time on the previous day. Changing the forecast horizon only altered the division of forecast windows and did not change the pointwise predictions.

Removing the day-ahead price reduced 12 h forecasting performance, with MAE increasing by 50.51%. Five three-seed ablations additionally removed real-time prices, weather, system operating variables, or both price series under fixed full-model settings (Table A4). Removing both prices increased 12 h MAE by 53.37%, whereas removing real-time prices decreased 8 h MAE by 5.67%; the benefits therefore depend on the variable group and horizon.

Measured median inference times were 21.16–28.14 ms for LoRA, 4.49–6.15 ms for the PatchTST extension, and 1.00–1.18 ms for LightGBM. Training costs and timing boundaries are reported in Table A9.

Table A3 compares native and validation-expanded intervals. LoRA coverage increased from 67.24%, 66.39%, and 75.99% to 83.09%, 81.47%, and 85.42%, respectively, at the cost of wider intervals. Interval scores improved at 4 and 8 h but slightly worsened at 12 h. Native undercoverage may reflect finite-sample estimation and temporal distribution shift.

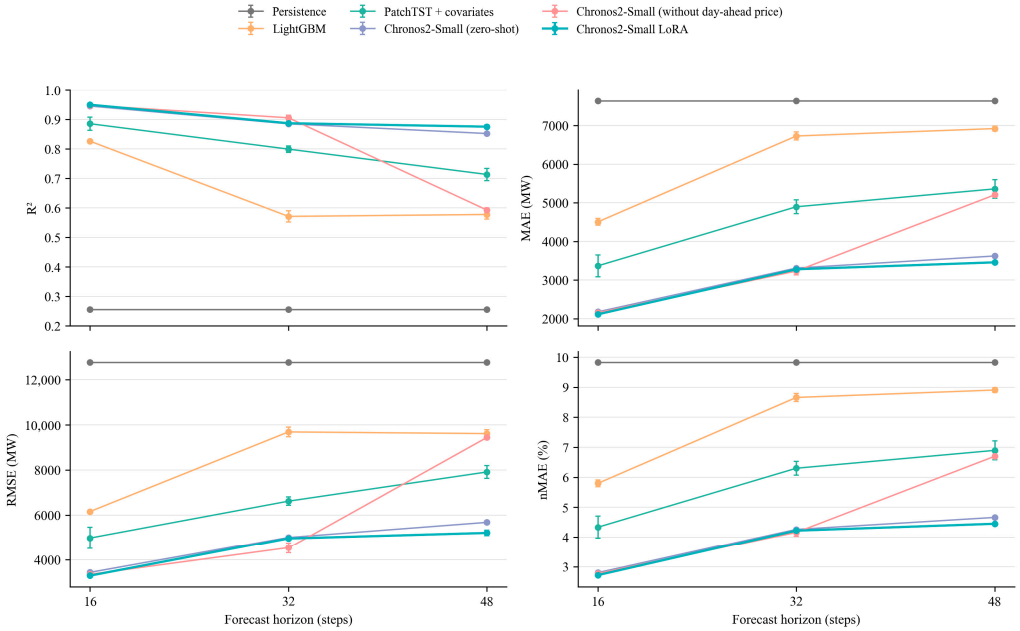


Figure 4. Prediction metrics for each model.

3.3. SHAP Analysis

To investigate the effects of different covariates on net-load forecasting, grouped masked SHAP analysis was applied to the Chronos2-Small LoRA model. The covariates were divided into five groups according to their physical meaning and temporal availability: weather and renewable conditions, system operating conditions, day-ahead market signals, historical real-time market conditions, and intraday cycles. Historical net load was retained in all feature coalitions and was neither masked nor assigned a SHAP contribution. To reduce the computational cost, one sample was selected at every fourth time point in the test set, and SHAP values were calculated separately for the 16-, 32-, and 48-step forecasts. This gave 336 common timestamps per horizon and seed. The analysis was repeated with three seeds and training-slot mean/median references. The feature groups used in the SHAP analysis are summarized in Table 4.

Figure 5 presents the grouped SHAP summary plots for the three forecast horizons. The bars represent the mean absolute SHAP values of the feature groups, while the scatter points show the direction and relative magnitude of their contributions across different samples. The repeated-seed display pools all 1008 attributions per horizon; scatter colors indicate the sign of the contribution, and each group is scaled by its SHAP standard deviation.

For the 16-step forecast, weather and renewable conditions and the day-ahead market signal were the two most influential groups, accounting for 27.73% and 32.01% of the exogenous-group importance, respectively. System operating conditions, the intraday cycle, and historical real-time market conditions contributed 19.44%, 13.43%, and 7.38%, respectively. These results indicate that short-horizon forecasts depend jointly on physical operating conditions and market information.

Table 4. SHAP feature grouping.

Feature Group	Variables	Input Availability	Interpretation
Weather and renewable conditions (G1)	temperature; wind speed; solar radiation	Historical inputs	Weather conditions and renewable generation potential
System operating conditions (G2)	Gas turbine output; received power	Historical inputs	System regulation status and external power support
Day-ahead market signal (G3)	Day-ahead price	Historical and known future input	Market expectation of future supply-demand conditions
Historical real-time market conditions (G4)	Real-time market weighted-average price	Historical input	Recent real-time market conditions
Intraday cycle (G5)	Hour sine; hour cosine	Historical and known future inputs	Intraday periodicity

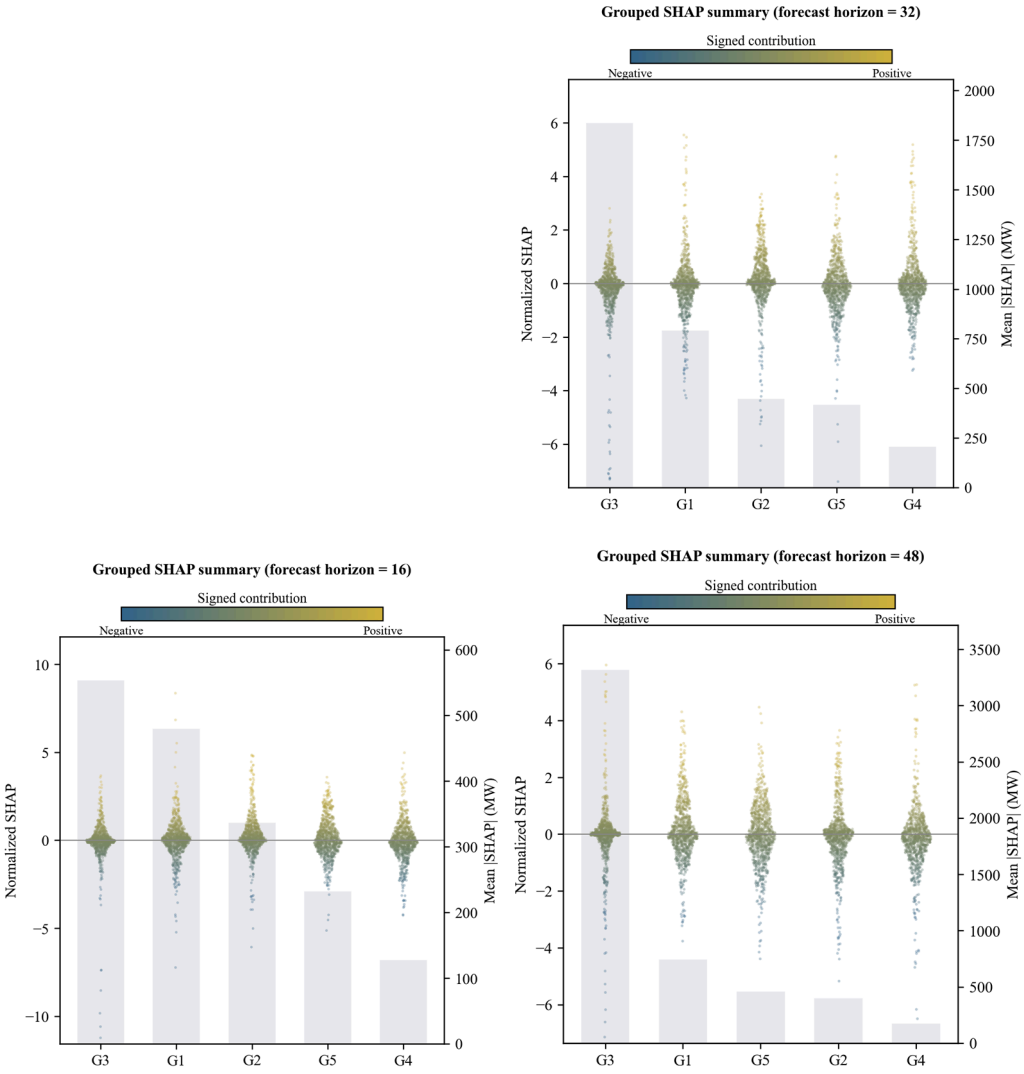


Figure 5. Grouped SHAP summary plots under different forecast horizons.

As the forecast horizon increased, the importance of the day-ahead market signal increased markedly. Its contribution rose from 32.01% for the 16-step forecast to 49.67% for the 32-step forecast and reached 65.21% for the 48-step forecast. By contrast, the contribution of weather and renewable conditions decreased from 27.73% to 14.60%, while that of historical real-time market conditions declined from 7.38% to 3.41%. The increasing day-ahead price share may reflect both its predictive information and its availability throughout the forecast window, as discussed in Section 4.

The SHAP values are distributed on both sides of zero, indicating that the same feature group may increase or decrease the predicted net load under different operating conditions. For some feature groups, the contribution direction varied across samples, reflecting nonlinear relationships among net load, weather conditions, system operation, and market variables.

Figure 6 presents grouped SHAP waterfalls at the common maximum and minimum observed net-load timestamps among the 336 explained samples: 18 March 2026 at 10:00 (99,504.80 MW) and 29 March 2026 at 12:00 (35,266.40 MW). The three horizons use these same target timestamps, seed 2026, and the training-slot mean reference; their forecast windows and historical contexts may differ.

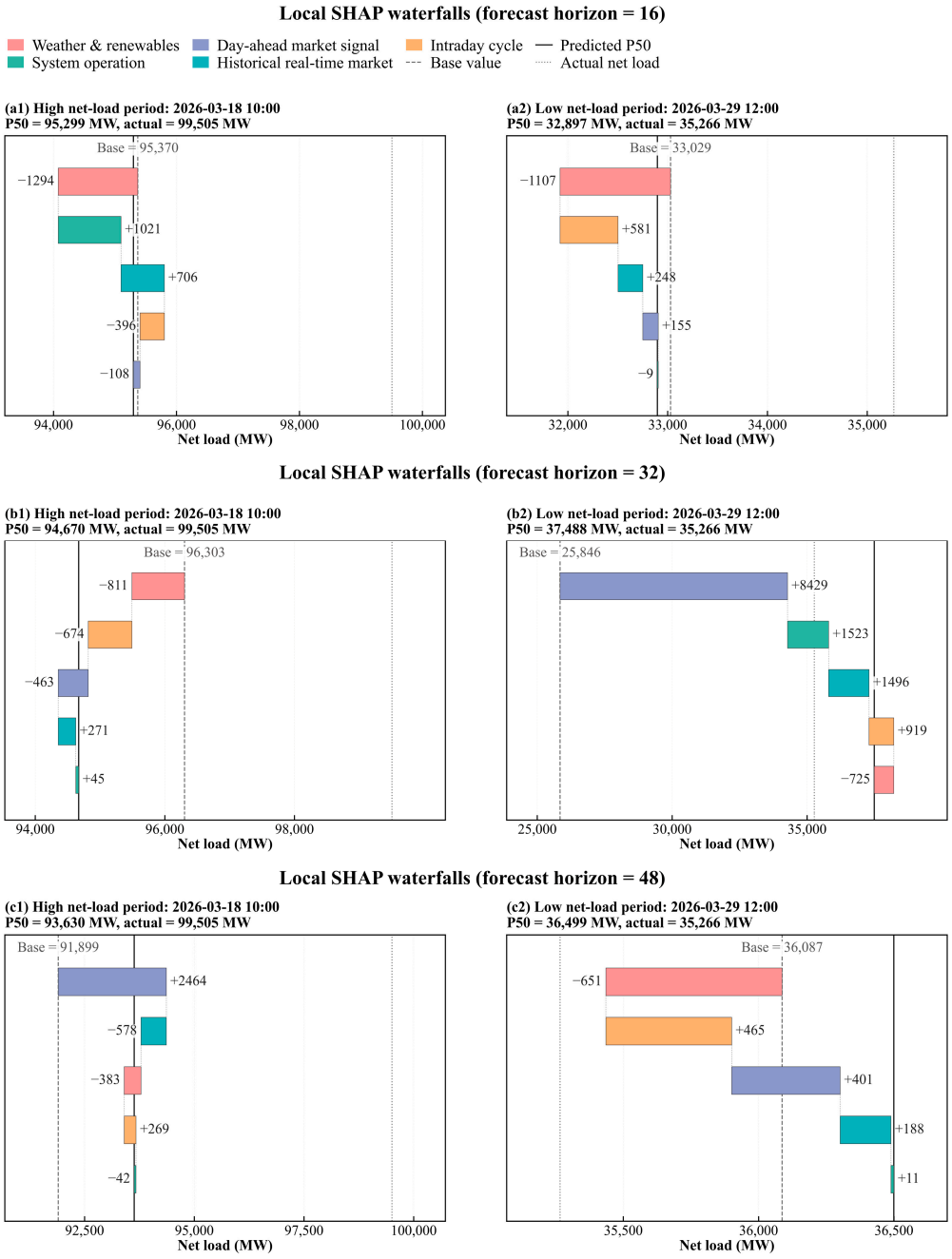


Figure 6. Local grouped SHAP contributions at the common high net-load (18 March 2026, 10:00) and low net-load (29 March 2026, 12:00) target timestamps for the 4, 8, and 12 h forecasts.

In Figure 6, the colored bars represent the positive and negative contributions of the five feature groups relative to the baseline value. The sum of these contributions and the baseline value forms the P50 forecast. At the high net-load timestamp, the day-ahead contribution changed from −108 and −463 MW at 4 and 8 h to +2464 MW at 12 h. At the low net-load timestamp, the 8 h forecast had a day-ahead contribution of +8429 MW. These differences illustrate how local covariate contributions vary with the forecast window and operating conditions.

Overall, the SHAP results demonstrate that weather and renewable conditions are important contributors to short-horizon forecasts, whereas day-ahead market information provides greater explanatory value for longer forecast horizons. Historical real-time market information has a relatively limited marginal contribution.

Day-ahead prices ranked first in all 18 horizon–seed–reference combinations; mean/median rank correlations were 0.9–1.0 and the largest share difference was 2.53 percentage points. Table A8 reports 48 h block uncertainty. All 54 sampling-offset comparisons preserved the ranking (maximum share change: 1.48 percentage points).

3.4. Operational Risk Identification Results

Based on the P10, P50, and P90 outputs of Chronos2-Small LoRA, operational risks were identified using the relative prediction interval width and the P50 ramp. The corresponding thresholds were determined from the 75th percentiles of the validation set and then fixed for the test set. The operating states were classified as Normal, Watch, or High according to whether uncertainty and the predicted ramp exceeded their respective thresholds. Periods with high day-ahead prices and a risk level of Watch or High were additionally labeled as market-sensitive. Native intervals are used throughout screening. P75 defines an upper-tail reference, and P65/P85 are examined for sensitivity. An independent event is an actual 15 min ramp of at least 2530.057 MW, the validation P90 cutoff; 138 test timestamps meet it.

Table 5 summarizes the risk-state distribution and identification performance for the three forecast horizons.

Table 5. Operational risk identification results for different forecast horizons.

Forecast Horizon	Lead Time	Normal (%)	Watch (%)	High (%)	Market-Sensitive Periods (%)	Observed-Ramp Recall (%)	High/Normal MAE Ratio	High/Normal Actual-Ramp Ratio
16	4 h	58.41	28.89	12.70	11.66	97.58	2.96	3.61
32	8 h	58.09	30.80	11.11	10.29	96.62	3.06	3.39
48	12 h	62.85	23.49	13.67	10.22	95.41	3.42	3.38

Across the three forecast horizons, Normal periods accounted for 58.09–62.85% of the test set, Watch periods for 23.49–30.80%, and High periods for 11.11–13.67%. The combined alert fraction was 37.15–41.91%.

MAE during High periods was 2.96–3.42 times higher than that during Normal periods. Similarly, the mean actual net-load ramp during High periods was 3.38–3.61 times higher than that during Normal periods. Therefore, the periods identified as High were associated with larger forecast errors and more severe actual ramps.

The Spearman correlations between relative uncertainty and absolute forecast error were 0.552, 0.628, and 0.642; the corresponding predicted-versus-actual ramp correlations were 0.650, 0.636, and 0.638. The combined Watch and High states captured 95.41–97.58% of independently defined large-ramp events, with a precision of 23.69–26.41% and an F1 of 38.04–41.36% (Tables A6 and A7).

Figure 7 compares the risk-state distribution, outcome amplification, and high-ramp capture performance across different lead times.

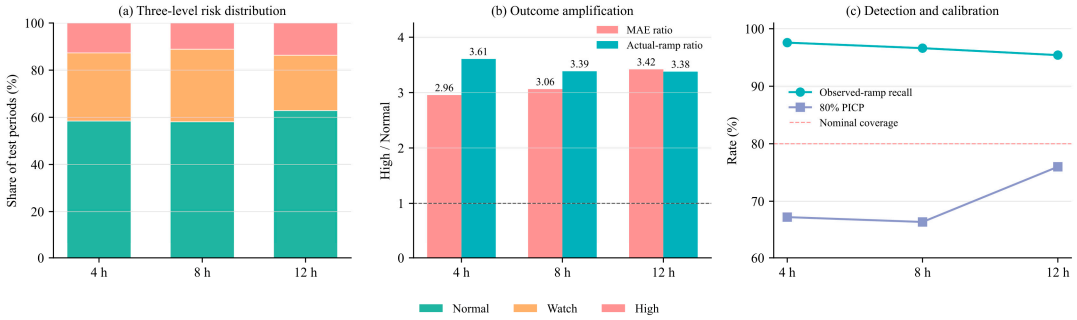


Figure 7. Operational risk identification results under different forecast horizons.

The native 80% prediction intervals had coverage rates of 67.24%, 66.39%, and 75.99% for the 16-, 32-, and 48-step forecasts, respectively. Table A3 reports calibration alongside interval width and proper scores, including validation-based expansion.

The day-ahead price was used only to assign the market-sensitive label and did not directly alter the three basic risk levels. Market-sensitive periods accounted for 10.22–11.66% of the test set, indicating that high market prices and elevated operating risks occurred simultaneously only during certain periods. This distinction prevents high prices from being directly interpreted as high operating risks while identifying risk periods with stronger market relevance.

Higher screening percentiles reduce alerts and generally trade recall for precision (Table A6). Compared with ramp-only screening, the joint rule has higher recall but lower F1 (Table A7).

To illustrate the temporal variation in risk levels, Figure 8 presents the first 48 h of the test period (seed 2026) from the 48-step forecast, including the actual net load, P50 forecast, 80% prediction interval, identified risk level, and day-ahead price. The green, orange, and red background areas indicate Normal, Watch, and High risk levels, respectively.

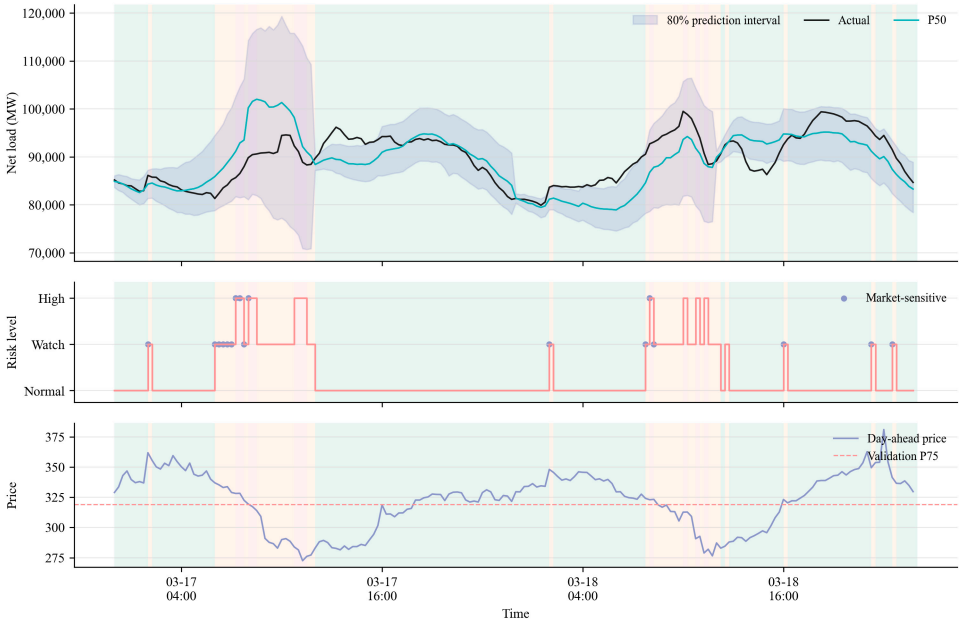


Figure 8. Operational risk identification during a representative period for the 48-step forecast.

When either the prediction interval widened or the P50 ramp exceeded its threshold, the operating state changed from Normal to Watch. When both indicators exceeded their thresholds, the state was upgraded to High. High periods mainly occurred when the prediction interval expanded substantially or the net load changed rapidly, and these

periods were generally associated with larger forecast errors and actual ramps. The market-sensitive markers further identified Watch or High periods accompanied by elevated day-ahead prices.

Overall, the proposed three-level mechanism converted the model outputs into clear operational risk states and flagged periods with large ramps or elevated forecast errors. These results demonstrate its potential to support rolling monitoring, flexible-resource checks, and the preparation of ramping and reserve resources.

4. Discussion

Hybrid load forecasting combines feature extraction and temporal modeling, as illustrated by the CNN–GRU approach [13], while BiKoop [15] represents temporal dynamics through gated Koopman operators. Chronos-2 [22] provides pretrained representations for multivariate and covariate-informed forecasting. The present study adapts this framework to regional net-load forecasting using LoRA and market covariates. MAE reductions of 33.02–37.23% relative to covariate-extended PatchTST, together with strong zero-shot performance, support the use of pretrained representations in this application. LoRA updates only 0.699–2.738% of the augmented model parameters; its small mean improvements over zero-shot Chronos2-Small are not statistically significant.

The interpretation of market inputs depends on their temporal availability. G3 (day-ahead prices) and G5 (intraday cycles) are known throughout the forecast window, whereas G1, G2, and G4 provide historical observations. As the horizon extends, the larger G3 SHAP share may therefore reflect both the predictive content of prices and their advance availability. These two effects are not separated by the current attribution analysis. Removing day-ahead prices increases 12 h MAE by 50.51%, supporting their value in the implemented forecasting setup. Conversely, removing historical real-time prices decreases 8 h MAE by 5.67%, showing that the benefit varies across market variables and horizons.

Probabilistic load forecasting research [9,12] evaluates predictive uncertainty alongside central estimates. In this study, validation-based interval expansion improves coverage at all three horizons, while interval scores improve at 4 and 8 h and slightly worsen at 12 h. This result highlights the balance between calibration and interval width. Linking these forecasts to observed-ramp events further provides an operational interpretation: joint screening achieves 95.41–97.58% recall, with precision of 23.69–26.41% and lower F1 than ramp-only screening. The joint rule favors broad event capture, whereas ramp-only screening produces a more selective set of alerts.

5. Conclusions

Under increasing renewable energy penetration, accurate intraday net-load forecasting and effective identification of operationally difficult periods are important for power-system operation. This paper develops a market-informed multi-step net-load forecasting framework based on Chronos2-Small with LoRA adaptation, together with SHAP-based interpretation and a lightweight operational risk-screening mechanism. The main conclusions are as follows:

(1) In the Jiangsu case study, Chronos2-Small LoRA achieved the lowest mean MAE among the evaluated models at the 4, 8, and 12 h forecasting horizons, with R^2 values of 0.9502, 0.8874, and 0.8758 and nMAE values of 2.72%, 4.22%, and 4.45%, respectively. Across the three seeds, compared to covariate-extended PatchTST, MAE was reduced by 37.23%, 33.02%, and 35.46%. The gains over zero-shot Chronos2-Small were not statistically significant.

(2) The SHAP results show a horizon-dependent contribution of market information. The day-ahead price share of exogenous-group SHAP importance increased from 32.01%

at 4 h to 49.67% at 8 h and 65.21% at 12 h, indicating greater model reliance on day-ahead market information at longer forecasting horizons.

(3) The proposed three-level operational risk mechanism identifies periods with elevated observed ramps. The combined Watch and High states captured 95.41–97.58% of independently defined large-ramp events, with a precision of 23.69–26.41%.

This study uses three months of observations from Jiangsu with a 14-day test period. Further work will examine seasonal and cross-regional performance, verify covariate availability at forecast time, and assess operational decision costs.

Author Contributions: Conceptualization, W.Z.; Methodology, Z.J. (Zhihan Jiang); Software, H.H.; Validation, H.W.; Formal analysis, J.Z.; Investigation, J.L.; Visualization, Z.J. (Zhengsen Ji). All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by State Grid Jiangsu Electric Power Company (Project No. SHJY-SR-A-2026092).

Data Availability Statement: The measured data are not publicly available because of privacy and confidentiality restrictions. The authors have permission to use the data for this study.

Conflicts of Interest: Author Weiping Zhu was employed by the company State Grid Jiangsu Electric Power Co., Ltd. Authors Zhihan Jiang, Haiqing Huang, Haoran Wu, and Jiangyu Zhu were employed by the company State Grid Zhenjiang Power Supply Company. This research was supported by State Grid Jiangsu Electric Power Company.

Nomenclature

Symbol	Definition
TSFM; LoRA; SHAP	Time-series foundation model; low-rank adaptation; Shapley additive explanations
DA; RT; PV	Day-ahead; real-time; photovoltaic
FFN; MSA; LN	Feed-forward network; multi-head self-attention; layer normalization
MAE; RMSE; R^2; nMAE	Mean absolute error; root mean square error; coefficient of determination; normalized MAE
PICP; IS; CRPS	Prediction interval coverage probability; interval score; continuous ranked probability score
SD; CI; FPR; ROC-AUC	Sample standard deviation; confidence interval; false-positive rate; area under the receiver-operating-characteristic curve
TP; FP; FN; TN	True positives; false positives; false negatives; true negatives
t; h; H; L	Last observed time; forecast-step index; horizon length; historical context length
i; j; k	Training sample index; variable/channel index; patch index
y_t; $\bar{y}$	Observed net load; mean test net load used to normalize MAE
y_hist	Net-load history fixed during coalition evaluation
w_t; o_t	Weather vector (temperature, wind, radiation); operation vector (gas-turbine output, received power)
p^{DA}; p^{RT}	Day-ahead price; historically observed real-time price
c_t	Sine/cosine of 2π(hour + minute/60)/24
X; X_past; X_future	Complete model inputs; historical inputs; permitted known future covariates
$f\theta$; θ	Forecast mapping; model parameters
τ; $\mathcal{T}$	Quantile level; set of quantile levels (13 native levels, nine evaluated levels from 0.1 to 0.9)
$P\tau$; P10; P50; P90	Predicted τ-quantile; the 0.1, 0.5 and 0.9 quantiles
$x^{(j)}$; $\tilde{x}^{(j)}$; μ_j; σ_j	Raw and transformed channel values; context mean; standard deviation
P; $\tilde{x}_{jk}$; e_{jk}	Patch length 16; normalized patch for channel j and patch k; patch embedding
r_k; m_{jk}; φ	Patch time encoding; observation mask; patch embedding mapping

H; H^{T}; H^{G}; H_out	Hidden state; state after time attention; after group attention; after FFN
MSA_T; MSA_G; ψ	Temporal attention; group attention; quantile output mapping
W; W'; ΔW	Frozen weight, adapted weight, and update; dimensions d_out × d_in
A; B (LoRA); r; α	LoRA factors r × d_in and d_out × r; rank; scale; $\Delta W = BA$; $W' = W + (\alpha/r)\Delta W$
W_Q; W_K; W_V; W_O	Query, key, value, and output projection matrices
$\mathcal{L}$; $\varrho\tau$; e; N	Training loss; pinball loss; observation-minus-quantile residual; number of training windows
G; g; S; M	Set of five exogenous groups; group; retained coalition; number of groups $M = 5$

Appendix A. Additional Experimental Details

The supplementary results use the same chronological split and three final seeds as Table 3.

Table A1. Hyperparameter search space.

Model	Three Parameter Settings	Parameter Order and Stopping
LightGBM	(0.03, 63, 30, 1); (0.05, 31, 20, 1); (0.03, 127, 50, 5)	Learning rate, leaves, min child samples, L2; ≤3000 trees; patience 100
PatchTST extension	(0.0005, 64, 3, 0.1); (0.001, 64, 2, 0.2); (0.0003, 128, 3, 0.1)	Learning rate, width, layers, dropout; ≤100 epochs; patience 15
Chronos2 LoRA	(0.00001, 8); (0.00005, 4); (0.0001, 16)	Learning rate and rank; $\alpha = 2r$; ≤1000 steps; validation every 100 steps; patience four evaluations

Each setting is crossed with contexts of 336 and 672. Selection uses seed 2025 and validation MAE; final fits use seeds 2026–2028.

Table A2. Selected configurations.

Model	h	Context	Selected parameters
LightGBM	4	336	lr = 0.03; leaves = 63; min_child = 30; L2 = 1.0
LightGBM	8	672	lr = 0.05; leaves = 31; min_child = 20; L2 = 1.0
LightGBM	12	672	lr = 0.05; leaves = 31; min_child = 20; L2 = 1.0
Covariate-extended PatchTST	4	336	lr = 0.0005; width = 64; layers = 3; dropout = 0.1
Covariate-extended PatchTST	8	336	lr = 0.0005; width = 64; layers = 3; dropout = 0.1
Covariate-extended PatchTST	12	336	lr = 0.001; width = 64; layers = 2; dropout = 0.2
Chronos2-Small LoRA	4	672	lr = 0.0001; rank = 16; $\alpha = 32$; dropout = 0.05
Chronos2-Small LoRA	8	672	lr = 0.0001; rank = 16; $\alpha = 32$; dropout = 0.05
Chronos2-Small LoRA	12	672	lr = 5×10^{-5}; rank = 4; $\alpha = 8$; dropout = 0.05

Experiments used autogluon/chronos-2-small, with 27,934,624 frozen backbone parameters. The model has six blocks, a hidden width of 512, eight attention heads, an FFN width of 2048, and a patch length of 16. Attention Q/K/V/O LoRA uses $\alpha = 2r$ and dropout 0.05; FFNs remain frozen. Rank 16 trains 786,432/28,721,056 parameters (2.738%); rank 4 trains 196,608/28,131,232 (0.699%). Independent adapters are fitted per horizon and seed. Native batch size 64 counts channel series and yields seven ten-channel windows. Training uses fused AdamW, a linear schedule, no warm-up, and the best validation checkpoint.

LightGBM uses lagged values and rolling mean/SD/min/max features from permitted inputs, a squared-error objective, a subsample of 0.9, a column subsample of 0.8, and eight threads. The PatchTST extension uses a patch length of 16, a stride of 8, four heads, an FFN width of 2d, and training-only channel scaling; its head combines historical representations with per-step and whole-trajectory known-future projections. Training uses MSE, AdamW, weight decay 10^{-4}, batch size 64 windows, clipping 1.0, and learning-rate halving after five non-improving epochs. Chronos retains native context scaling; LightGBM uses un-

scaled features. Invalid or duplicate timestamps, irregular intervals, and missing/non-finite modeled values are rejected.

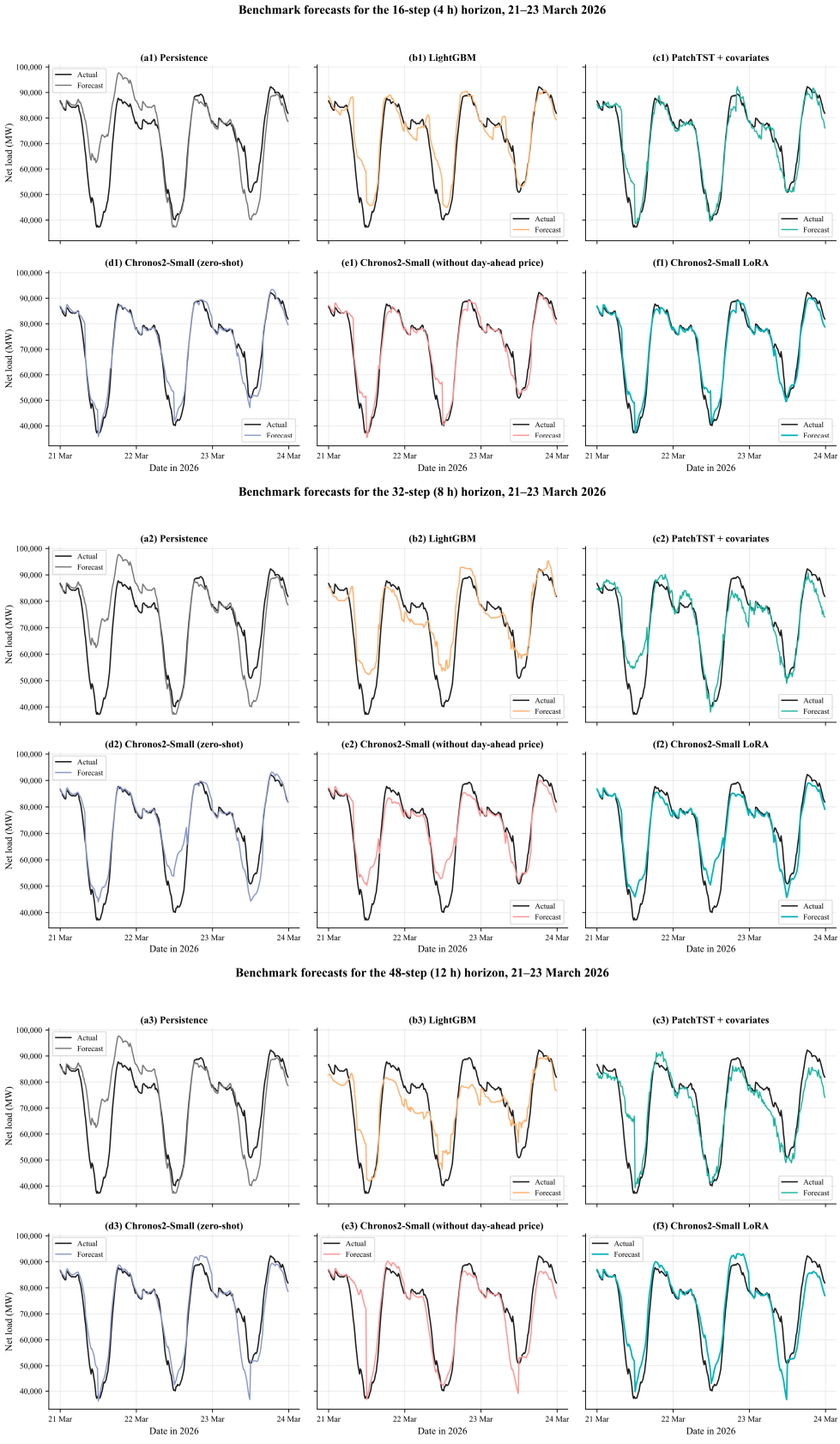


Figure A1. Prediction curve comparison chart. Seed 2026; the same 21–23 March period is shown for all methods.

Software: Python 3.11.15, PyTorch 2.11.0 + cu128, chronos-forecasting 2.3.1, Transformers 4.48.3, PEFT 0.20.0, LightGBM 4.7.0, NumPy 2.4.3, pandas 3.0.2, and SciPy 1.17.1. Hardware: NVIDIA RTX 5060 (8 GB), AMD Ryzen 5 7500F, and 16 GB RAM. Neural training uses float32 and deterministic algorithms with TF32 disabled.

Table A3. Probabilistic performance before and after validation-based interval expansion.

Model	h	Interval	Mean Pinball (MW)	Coverage (%)	Mean Width (MW)	Interval Score (MW)
Zero-shot	4	Native	877.46	74.48	5647.95	11,124.72
Zero-shot	4	Expanded	878.38	85.34	7365.87	11,207.35
LoRA	4	Native	848.90 ± 12.94	67.24 ± 1.45	4682.53 ± 134.76	10,719.55 ± 270.97
LoRA	4	Expanded	845.49 ± 11.63	83.09 ± 0.96	6304.79 ± 83.71	10,412.42 ± 160.54
Zero-shot	8	Native	1317.84	71.88	8097.49	16,390.45
Zero-shot	8	Expanded	1313.74	81.03	10,583.91	16,021.53
LoRA	8	Native	1301.86 ± 6.79	66.39 ± 1.73	6991.70 ± 225.53	16,025.97 ± 296.06
LoRA	8	Expanded	1293.70 ± 8.02	81.47 ± 0.39	9850.24 ± 105.93	15,291.36 ± 204.31
Zero-shot	12	Native	1442.76	75.00	8712.29	18,146.84
Zero-shot	12	Expanded	1444.24	85.19	11,956.40	18,280.56
LoRA	12	Native	1348.76 ± 12.62	75.99 ± 2.91	9117.70 ± 869.46	15,795.37 ± 437.09
LoRA	12	Expanded	1350.01 ± 10.69	85.42 ± 1.08	11,512.16 ± 390.20	15,907.34 ± 308.63

Pinball loss averages nine quantile levels. LoRA: mean ± sample SD over three seeds; zero-shot: fixed reference. Lower pinball loss and interval score indicate better performance.

Table A4. Controlled covariate ablation.

Removed Input	h	MAE (MW)	Change (%)	95% CI of ΔMAE (MW)
Day-ahead price	4	2154.01 ± 57.12	1.83	[−88.08, 155.53]
Day-ahead price	8	3234.26 ± 100.06	−1.40	[−731.79, 503.47]
Day-ahead price	12	5206.68 ± 46.46	50.51	[733.65, 2907.60]
Real-time price	4	2081.48 ± 62.55	−1.60	[−140.71, 73.49]
Real-time price	8	3094.07 ± 125.40	−5.67	[−315.33, −68.86]
Real-time price	12	3463.00 ± 70.12	0.10	[−99.22, 152.96]
Weather	4	2075.19 ± 22.63	−1.89	[−164.33, 85.43]
Weather	8	3085.25 ± 100.27	−5.94	[−503.04, 99.57]
Weather	12	3371.73 ± 59.35	−2.53	[−232.91, 60.61]
System operation	4	2146.97 ± 83.62	1.50	[−107.89, 161.27]
System operation	8	3309.16 ± 42.20	0.88	[−152.40, 203.54]
System operation	12	3568.47 ± 0.60	3.15	[−112.76, 348.08]
Both prices	4	2128.35 ± 74.66	0.62	[−107.39, 126.61]
Both prices	8	3353.07 ± 101.06	2.22	[−643.68, 636.89]
Both prices	12	5305.78 ± 81.75	53.37	[868.72, 2964.53]

Mean ± sample SD; change is relative to the full model. Intervals use 2000 circular 48 h block resamples. Full-model settings and seven windows per batch are retained; loss is multiplied by retained channels/10 to preserve target-loss scaling.

Table A5. Paired MAE reductions of LoRA relative to each comparator.

h	Comparator	MAE Reduction (MW)	24-h Block 95% CI (MW)	48-h Block 95% CI (MW)
4	LightGBM	2391.47	[1645.54, 3161.26]	[1594.91, 3289.81]
4	Covariate-extended PatchTST	1254.37	[865.58, 1651.32]	[847.55, 1657.29]
4	Chronos2-Small zero-shot	63.94	[−134.87, 248.03]	[−112.45, 263.07]
8	LightGBM	3448.98	[1679.59, 5764.54]	[1569.76, 6186.09]
8	Covariate-extended PatchTST	1617.23	[639.72, 2626.97]	[574.42, 2729.54]
8	Chronos2-Small zero-shot	31.56	[−416.79, 445.55]	[−381.26, 457.01]

Table A5. *Cont.*

h	Comparator	MAE Reduction (MW)	24-h Block 95% CI (MW)	48-h Block 95% CI (MW)
12	LightGBM	3461.44	[2035.83, 5375.81]	[1938.30, 5515.53]
12	Covariate-extended PatchTST	1900.29	[955.52, 2806.99]	[824.94, 2793.69]
12	Chronos2-Small zero-shot	164.33	[−308.20, 653.28]	[−389.29, 690.22]

Positive values favor LoRA. Paired percentile intervals use 5000 circular block resamples of seed-averaged absolute errors; model-selection uncertainty and multiplicity are not incorporated.

Table A6. Sensitivity of joint screening to the validation percentile.

h	Cutoff	Precision (%)	Recall (%)	F1 (%)	FPR (%)	Alert Points
4	P65	19.99 ± 0.45	98.55 ± 0.72	33.23 ± 0.67	45.16 ± 0.97	680.7 ± 10.8
4	P75	24.10 ± 0.46	97.58 ± 1.11	38.65 ± 0.51	35.19 ± 1.26	559.0 ± 16.6
4	P85	31.58 ± 0.61	93.24 ± 2.74	47.17 ± 0.43	23.13 ± 1.30	407.7 ± 19.4
8	P65	19.58 ± 0.69	98.07 ± 0.42	32.64 ± 0.96	46.13 ± 2.05	691.7 ± 24.8
8	P75	23.69 ± 0.84	96.62 ± 1.11	38.04 ± 1.13	35.66 ± 1.49	563.3 ± 17.6
8	P85	30.52 ± 2.50	88.41 ± 0.72	45.33 ± 2.75	23.19 ± 2.86	401.7 ± 34.8
12	P65	20.40 ± 0.32	97.58 ± 0.42	33.74 ± 0.46	43.59 ± 0.68	660.3 ± 7.6
12	P75	26.41 ± 1.31	95.41 ± 1.11	41.36 ± 1.64	30.49 ± 2.07	499.3 ± 25.0
12	P85	33.71 ± 0.84	89.13 ± 0.72	48.92 ± 0.93	20.07 ± 0.72	365.0 ± 8.7

Mean ± sample SD over three seeds; the observed-event cutoff remains fixed at validation P90.

Table A7. Binary confusion counts and ROC-AUC at P75.

h	Rule	TP	FP	FN	TN	ROC-AUC
4	Joint	134.7 ± 1.5	424.3 ± 15.1	3.3 ± 1.5	781.7 ± 15.1	0.923 ± 0.009
4	High only	57.3 ± 2.1	113.3 ± 6.8	80.7 ± 2.1	1092.7 ± 6.8	0.803 ± 0.005
4	Market-sensitive	28.0 ± 0.0	128.7 ± 10.4	110.0 ± 0.0	1077.3 ± 10.4	0.767 ± 0.011
4	Ramp only	120.7 ± 2.1	248.7 ± 3.5	17.3 ± 2.1	957.3 ± 3.5	0.919 ± 0.004
4	Uncertainty only	71.3 ± 2.1	289.0 ± 16.5	66.7 ± 2.1	917.0 ± 16.5	0.703 ± 0.006
8	Joint	133.3 ± 1.5	430.0 ± 18.0	4.7 ± 1.5	776.0 ± 18.0	0.889 ± 0.006
8	High only	50.0 ± 3.6	99.3 ± 12.5	88.0 ± 3.6	1106.7 ± 12.5	0.713 ± 0.014
8	Market-sensitive	28.0 ± 1.0	110.3 ± 14.0	110.0 ± 1.0	1095.7 ± 14.0	0.776 ± 0.016
8	Ramp only	124.0 ± 3.6	237.0 ± 10.4	14.0 ± 3.6	969.0 ± 10.4	0.917 ± 0.004
8	Uncertainty only	59.3 ± 1.5	292.3 ± 21.2	78.7 ± 1.5	913.7 ± 21.2	0.576 ± 0.010
12	Joint	131.7 ± 1.5	367.7 ± 24.9	6.3 ± 1.5	838.3 ± 24.9	0.900 ± 0.002
12	High only	63.0 ± 2.6	120.7 ± 3.8	75.0 ± 2.6	1085.3 ± 3.8	0.800 ± 0.004
12	Market-sensitive	28.0 ± 0.0	109.3 ± 12.5	110.0 ± 0.0	1096.7 ± 12.5	0.762 ± 0.014
12	Ramp only	124.7 ± 1.2	244.0 ± 14.7	13.3 ± 1.2	962.0 ± 14.7	0.917 ± 0.007
12	Uncertainty only	70.0 ± 3.6	244.3 ± 17.9	68.0 ± 3.6	961.7 ± 17.9	0.692 ± 0.008

Counts are means ± sample SD and need not be integers. Continuous scores use max(u,a) for Joint, min(u,a) for High, min(max(u,a),z) for Market-sensitive, a for Ramp only, and u for Uncertainty only; u = U/U_th, a = ramp/ramp_th, z = 1 + (price − price_th)/max(|price_th|,1).

For ramp-only screening, recall is 87.44%, 89.86%, 90.34% and F1 is 47.57%, 49.71%, 49.24% at 4, 8, and 12 h, respectively. Joint-rule false-positive rates are 35.19%, 35.66%, 30.49%, respectively.

Table A8. Stability of the day-ahead attribution share.

h	Reference	DA Share (%)	48-h Block 95% CI (%)	Top-Rank Frequency (%)
4	Mean	32.01	[27.15, 37.18]	89.15
4	Median	32.42	[27.60, 37.56]	92.25

Table A8. *Cont.*

h	Reference	DA Share (%)	48-h Block 95% CI (%)	Top-Rank Frequency (%)
8	Mean	49.67	[42.35, 56.12]	100.00
8	Median	49.17	[42.03, 55.41]	100.00
12	Mean	65.21	[57.34, 70.98]	100.00
12	Median	65.01	[57.25, 70.59]	100.00

Shares normalize mean absolute attributions within five exogenous groups. Intervals and top-rank frequencies use 2000 circular 48 h block resamples of 336 common hourly timestamps; three seeds are pooled. Other 15-min sampling offsets are also checked.

Table A9. Training and inference costs.

Model	h	Fit Time (s)	Six Trials (s)	Median (ms)	P95 (ms)
LightGBM	4	4.07 ± 0.72	27.47	1.00	1.36
LightGBM	8	4.46 ± 0.96	32.50	1.14	1.17
LightGBM	12	5.33 ± 0.81	49.40	1.18	1.23
Covariate-extended PatchTST	4	18.55 ± 5.19	213.55	5.63	5.80
Covariate-extended PatchTST	8	16.67 ± 2.43	212.23	6.15	6.52
Covariate-extended PatchTST	12	11.45 ± 1.21	146.34	4.49	5.12
Chronos2-Small LoRA	4	94.03 ± 10.70	371.34	24.22	26.70
Chronos2-Small LoRA	8	67.77 ± 11.66	434.36	28.14	30.03
Chronos2-Small LoRA	12	86.39 ± 22.75	439.41	21.16	22.79

Training measures fit-call wall time. Inference uses seed 2026, three warm-ups and ten windows, from in-memory raw observations to CPU forecasts; preprocessing and transfers are included, while loading, disk I/O, and SHAP are excluded. Neural models use the same GPU and LightGBM uses eight CPU threads.

References

1. Makarov, Y.V.; Etingov, P.V.; Ma, J.; Huang, Z.; Subbarao, K. Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures. *IEEE Trans. Sustain. Energy* **2011**, *2*, 433–442. [CrossRef]
2. Huber, M.; Dimkova, D.; Hamacher, T. Integration of wind and solar power in Europe: Assessment of flexibility requirements. *Energy* **2014**, *69*, 236–246. [CrossRef]
3. Sreekumar, S.; Sharma, K.C.; Bhakar, R. Gumbel copula based aggregated net load forecasting for modern power systems. *IET Gener. Transm. Distrib.* **2018**, *12*, 4348–4358. [CrossRef]
4. Sreekumar, S.; Khan, N.U.; Rana, A.S.; Sajjadi, M.; Kothari, D.P. Aggregated net-load forecasting using Markov-Chain Monte-Carlo regression and C-vine copula. *Appl. Energy* **2022**, *328*, 120171. [CrossRef]
5. Rubasinghe, O.; Zhang, T.; Zhang, X.; Choi, S.S.; Chau, T.K.; Chow, Y.; Fernando, T.; Iu, H.H.-C. Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high PV penetration. *Appl. Energy* **2023**, *333*, 120641. [CrossRef]
6. Tziolis, G.; Spanias, C.; Theodoride, M.; Theocharides, S.; Lopez-Lorente, J.; Livera, A.; Makrides, G.; Georghiou, G.E. Short-term electric net load forecasting for solar-integrated distribution systems based on Bayesian neural networks and statistical post-processing. *Energy* **2023**, *271*, 127018. [CrossRef]
7. Telle, J.-S.; Upadhaya, A.; Schönfeldt, P.; Steens, T.; Hanke, B.; von Maydell, K. Probabilistic net load forecasting framework for application in distributed integrated renewable energy systems. *Energy Rep.* **2024**, *11*, 2535–2553. [CrossRef]
8. Pinheiro, M.G.; Madeira, S.C.; Francisco, A.P. Short-term electricity load forecasting—A systematic approach from system level to secondary substations. *Appl. Energy* **2023**, *332*, 120493. [CrossRef]
9. Hong, T.; Fan, S. Probabilistic electric load forecasting: A tutorial review. *Int. J. Forecast.* **2016**, *32*, 914–938. [CrossRef]
10. Zhang, S.; Wang, Y.; Zhang, Y.; Wang, D.; Zhang, N. Load probability density forecasting by transforming and combining quantile forecasts. *Appl. Energy* **2020**, *277*, 115600. [CrossRef]
11. He, Y.; Cao, C.; Wang, S.; Fu, H. Nonparametric probabilistic load forecasting based on quantile combination in electrical power systems. *Appl. Energy* **2022**, *322*, 119507. [CrossRef]
12. Brusaferri, A.; Matteucci, M.; Spinelli, S.; Vitali, A. Probabilistic electric load forecasting through Bayesian mixture density networks. *Appl. Energy* **2022**, *309*, 118341. [CrossRef]

13. Cui, J.; Kuang, W.; Geng, K.; Bi, A.; Bi, F.; Zheng, X.; Lin, C. Advanced short-term load forecasting with XGBoost-RF feature selection and CNN-GRU. *Processes* **2024**, *12*, 2466. [CrossRef]
14. Phien, N.N. WaveKoop-KAN: A wavelet-decomposed Koopman–gated temporal convolution hybrid with Kolmogorov–Arnold fusion for univariate time series forecasting. *IEEE Access* **2026**, *14*, 96322–96342. [CrossRef]
15. Phien, N.N.; Khiem, H.A.; Phuoc, T.T. BiKoop: Bidirectional Koopman operators with smooth regime gating for univariate time series forecasting. *IEEE Access* **2026**, *14*, 136820–136847. [CrossRef]
16. Dong, X.; Yu, Y.; Jin, H.; Hu, Z.; Bao, J. Short-term load forecasting in price-volatile markets: A pattern-clustering and adaptive modeling approach. *Processes* **2026**, *14*, 5. [CrossRef]
17. Lago, J.; Marcjasz, G.; De Schutter, B.; Weron, R. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Appl. Energy* **2021**, *293*, 116983. [CrossRef]
18. Ansari, A.F.; Stella, L.; Turkmen, C.; Zhang, X.; Mercado, P.; Shen, H.; Shchur, O.; Rangapuram, S.S.; Pineda Arango, S.; Kapoor, S.; et al. Chronos: Learning the language of time series. *Trans. Mach. Learn. Res.* **2024**. [CrossRef]
19. Ye, J.; Yu, Y.; Zhang, W.; Wang, L.; Li, J.; Tsung, F. Empowering time series analysis with foundation models: A comprehensive survey. *Inf. Fusion* **2027**, *137*, 104601. [CrossRef]
20. Neubauer, A.; Brandt, S.; Kriegel, M. Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability. *Energy AI* **2025**, *20*, 100480. [CrossRef]
21. Ramos, D.; Faria, P.; Vale, Z. Linking short-term electricity demand forecasting and explainable AI: A review for building energy applications. *Appl. Energy* **2026**, *412*, 127658. [CrossRef]
22. Ansari, A.F.; Shchur, O.; Küken, J.; Auer, A.; Han, B.; Mercado, P.; Rangapuram, S.S.; Shen, H.; Stella, L.; Zhang, X.; et al. Chronos-2: From univariate to universal forecasting. *arXiv* **2025**, arXiv:2510.15821.
23. Buonanno, A.; Fabozzi, S.; Valenti, M.; Graditi, G. The impact of covariates on zero-shot building energy forecasting using Chronos-2 foundation model. *Electronics* **2026**, *15*, 3474. [CrossRef]
24. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-rank adaptation of large language models. In Proceedings of the Tenth International Conference on Learning Representations (ICLR 2022), Virtual, 25–29 April 2022.
25. Li, Y.; Zhu, W. TRACE: Time series parameter efficient fine-tuning. *Neurocomputing* **2026**, *664*, 132098. [CrossRef]
26. Sheng, L.; Xu, Y.; Zhang, X.; Wei, M. Parameter-efficient adaptation of GPT-style foundation models for power grid load forecasting. *Electron. Lett.* **2026**, *62*, e70612. [CrossRef]
27. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.
28. Aas, K.; Jullum, M.; Løland, A. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. *Artif. Intell.* **2021**, *298*, 103502. [CrossRef]
29. Hertel, M.; Pütz, S.; Mikut, R.; Hagenmeyer, V.; Schäfer, B. Explainable time-series forecasting with sampling-free SHAP for Transformers. *Nat. Commun.* **2026**, *17*, 4723. [CrossRef]
30. Cui, M.; Zhang, J.; Feng, C.; Florita, A.R.; Sun, Y.; Hodge, B.-M. Characterizing and analyzing ramping events in wind power, solar power, load, and netload. *Renew. Energy* **2017**, *111*, 227–244. [CrossRef]
31. Li, B.; Feng, C.; Siebenschuh, C.; Zhang, R.; Spyrou, E.; Krishnan, V.; Hobbs, B.F.; Zhang, J. Sizing ramping reserve using probabilistic solar forecasts: A data-driven method. *Appl. Energy* **2022**, *313*, 118812. [CrossRef]
32. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.