

ORIGINAL RESEARCH

Open access

Hybrid Forecasting of Abnormal Load Variations in Urban Power Systems Using TimesNet and Sample Augmentation

Charlotte Dupuis^{1*}, Hugo Perrin¹

Abstract

Amid the ongoing acceleration of urban development, tertiary sector consumption and household electricity demand have emerged as the dominant components of overall municipal power consumption. Such loads are highly sensitive to unpredictable factors such as occupant behavior and meteorological conditions, which often lead to irregular deviations from expected trends and, consequently, considerable prediction inaccuracies. The present study introduces an integrated prediction approach to improve forecast precision for city-wide electricity demand, particularly across intervals characterized by anomalous consumption shifts. Initially, a systematic metric is developed to detect and quantify these irregular load excursions by analyzing the residual series obtained through decomposition of the raw load sequence. Next, a data-expansion strategy grounded in Generative Adversarial Networks is developed to synthetically augment the otherwise sparse set of anomalous instances, thereby enriching the dataset available for model training. Lastly, the sophisticated TimesNet architecture is adopted to effectively model the intricate, nonlinear dynamics observed in load behavior throughout periods of unusual variation. Validation experiments performed on real-world consumption records from Chongqing, China, substantiate the performance gains achieved by the suggested framework.

Keywords Abnormal load variation forecasting, Load decomposition, Sample augmentation, Generative adversarial nets, TimesNet

*Correspondence:

Charlotte Dupuis

charlotte.dupuis@univ-lille.fr

¹ Department of Smart Energy Networks, Faculty of Engineering, University of Lille, Lille, France

Introduction

Background and motivation

Accurate short-term electricity demand prediction remains an essential foundation for reliable power grid management and has attracted sustained scholarly interest over recent decades. Progressive urban expansion, combined with changes in economic composition, has led to tertiary-sector use and domestic consumption accounting for the largest share of city-level electricity demand. As one example, official statistics published by the Shanghai administration in 2024 reveal that air-conditioning-related demand accounted for up to 50% of the peak summer load and 43% of the winter load in that city. In consequence, contemporary urban load patterns are increasingly governed by human routines and atmospheric variables—both probabilistic in nature—leading to heightened volatility in consumption levels. Simultaneously, climate change has driven a noticeable rise in the frequency of severe weather episodes, including prolonged heatwaves and intense storm systems. **Figure 1** illustrates this phenomenon by presenting the daily temperature record alongside the combined electricity usage of over 1,000 households in Austin, Texas, throughout August 2017 [1, 2]. The graph clearly shows distinct atypical consumption patterns coinciding with the two major extreme-

weather intervals. Taken together, these trends produce more frequent occurrences of non-standard load behavior in urban settings, which, in turn, generate substantial discrepancies in forecasting outcomes. For this reason, the reliable detection and representation of such aberrant load movements (occasionally termed inflection points or turning points elsewhere in the literature) now represent a central research and engineering objective for advancing the reliability of urban electricity demand forecasts.

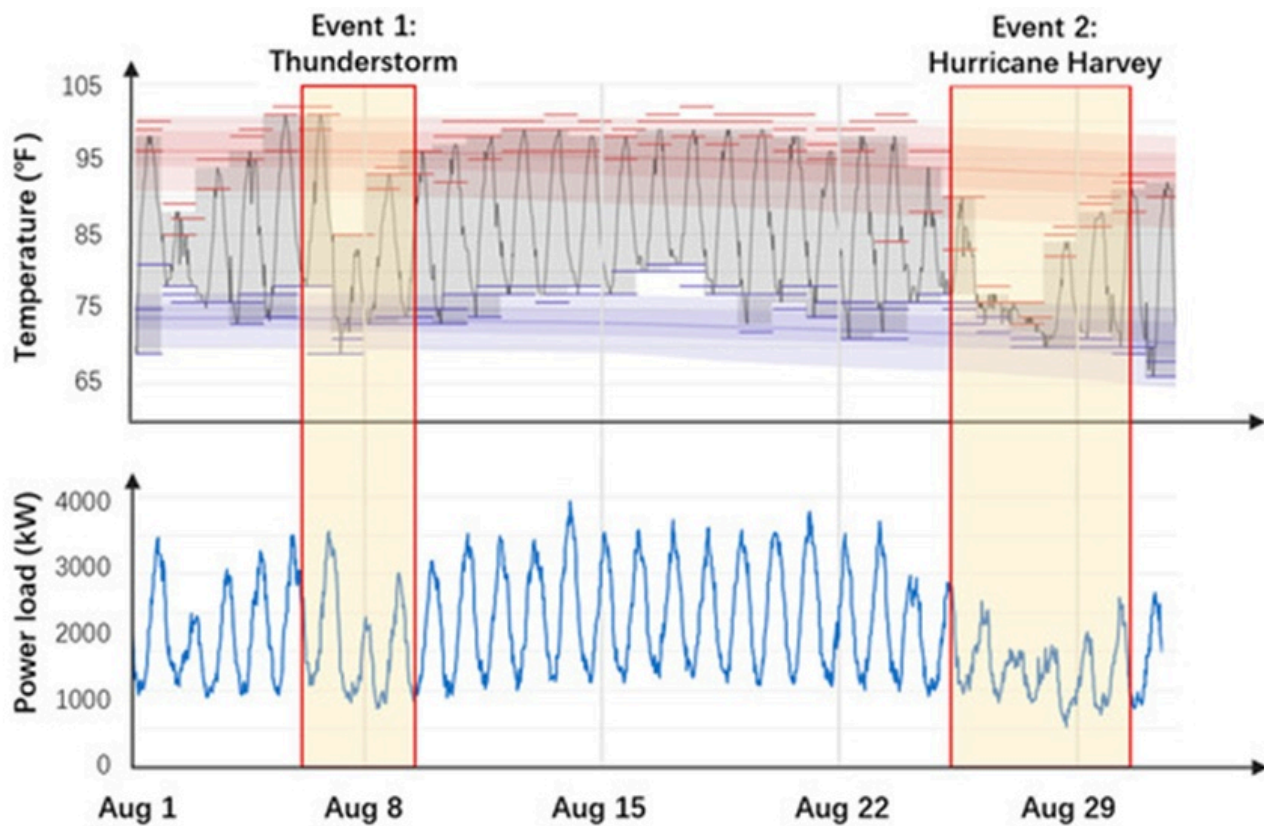


Figure 1. Temperature profile and the aggregated residential load profile in Austin, Texas, in August 2017. In the upper subfigure, red and blue ticks represent 24-hour temperature highs and lows. Faint red and blue lines represent the daily average high and low temperatures, with 25th-75th and 10th-90th percentile bands.

Literature review

Research pertinent to forecasting during periods of abnormal load fluctuations can be grouped into the following three main categories:

The first category emphasizes feature engineering techniques applied to the load forecasting task. In particular, meteorological variables like temperature and humidity have received considerable attention due to their strong association with electricity demand. Fay and Ringwood [3] examine how inaccuracies in weather predictions affect the overall performance of load forecasting models. Chu *et al.* [4] incorporate the Heat Index as a critical input variable for load prediction models. This index combines temperature and humidity to represent perceived temperature, revealing a strong link to air conditioning use and thereby improving model effectiveness. Yu *et al.* [5] develop a hierarchical Intelligent Feature Engineering (IFE) system to support short-term load prediction, where the initial IFE stage processes weather data to capture dependencies between load and meteorological conditions. Wang *et al.* [6] address the difficult task of forecasting individual household electricity usage by integrating weather-related inputs into a Long Short-Term Memory architecture. Departing from the typical focus on temperature-load relationships, Xie *et al.* [7] identify humidity's role in influencing power demand and identify relative humidity as a significant driver, particularly during warmer seasons.

They outline a structured procedure for deriving a set of humidity-derived indicators that improve forecasting precision across various time horizons. Hong *et al.* [8] address the challenge of optimally selecting weather stations and automatically assigning varying numbers to different prediction areas. Their method ranks and selects from candidate stations to determine both the ideal count and the most suitable ones concurrently. Mansouri *et al.* [9] present a forecasting approach that uses dynamic mode decomposition with control, treating apparent temperature—which accounts for the combined effects of temperature and humidity on people—as the primary exogenous factor. Xie and Hong [10] evaluate three approaches for creating temperature scenarios in probabilistic load forecasting, using quantile scoring as the criterion, and offer general recommendations for choosing appropriate scenario-generation strategies. Nevertheless, although the effects of weather elements on load prediction have been thoroughly explored, abnormal load fluctuations still lack a standardized, numerical definition, which remains a crucial element in feature engineering. Establishing such a quantitative framework would enable the detection of abnormal load events and provide valuable support to subsequent prediction models.

The second category focuses on refining load forecasting architectures to better represent load behavior and its interactions with weather factors. Dehalwar *et al.* [11] assess the predictive performance of an Artificial Neural Network (ANN) versus Bagged Regression Trees for municipal power demand estimation, incorporating weather predictions as inputs and confirming their significance for accuracy. Lu *et al.* [12] separate load series into baseline and weather-responsive portions, forecasting each independently; meteorological inputs train a Support Vector Regression model specifically for the weather-sensitive part. Adopting a comparable decompose-and-recombine strategy, Xu *et al.* [13] introduce a probabilistic model tailored for building-level demand prediction, including weather elements such as dry-bulb temperature and relative humidity as part of the standard load component. Kim *et al.* [14] compare multiple forecasting techniques for peak demand in institutional settings, demonstrating that weather variables serve as vital external predictors that improve performance. Accounting for distinct consumption patterns on weekdays versus weekends, Li *et al.* [15] develop a semi-parametric approach centered on weekend demand that incorporates meteorological influences and temperature accumulation dynamics. Pinheiro *et al.* [16] outline a comprehensive load prediction strategy spanning from grid-wide to low-voltage scales, employing numerical weather forecasts—particularly temperature—as key explanatory terms. Tests on 100,000 secondary substations in Portugal indicate that the method enhances accuracy while preserving strong usability, explainability, and replicability. Guo *et al.* [17] broaden traditional power load forecasting to encompass integrated energy systems that include cooling, heating, and electricity demand. They add an attention mechanism to identify inter-load relationships and propose an attentive temporal convolutional network for quantile regression with probabilistic outputs, including weather parameters such as temperature, humidity, and wind speed as essential inputs. Chen and Zhang [18] present a theory-informed deep learning model for load prediction that integrates domain expertise with machine learning; weather forecasts are among the inputs, and synthetic perturbations during training make the model resilient to inaccuracies in weather forecasts. Li *et al.* [19] describe a dual-stage method for electrical demand forecasting that leverages the rapidly advancing Internet of Things (IoT) infrastructure and performs sensitivity studies to rank factor importance, identifying temperature as a dominant contributor. It should be noted, however, that while substantial progress has been made in improving forecasting architectures, the challenge of limited sample sizes for abnormal load variation prediction remains largely unresolved. Since abnormal fluctuations are rare, the number of corresponding data points is far smaller than in typical conditions, posing difficulties for training sophisticated models. Developing augmentation strategies to expand the pool of abnormal-load variation instances artificially is therefore essential to enable effective training of advanced forecasting approaches.

The third category centers on the analysis and modeling of extreme weather phenomena, which, though infrequent, trigger pronounced and abrupt changes in electricity demand. Deng *et al.* [20] initially propose an identification framework for extreme weather conditions that incorporates load data, meteorological variables, and temporal factors to estimate the likely duration of impending peak-demand periods. They subsequently apply an Extreme Gradient Boosting (XGBoost) model enhanced with a Bagging ensemble to perform short-term load prediction. With a specific emphasis on wildfires, Yang *et al.* [21] introduce a load forecasting approach designed to maintain accuracy during wildfire-prone periods. The primary external input in their deep learning framework is the Fire Weather Index. Zhang *et al.* [22] investigate how hurricanes affect cascading outages within electric vehicle charging infrastructure. They construct a probabilistic graphical

model using a Bayesian Network to capture interdependencies among network nodes and implement an AC power-flow-based simulation to replicate the progression of cascading failures across the grid. Shield *et al.* [23] assess the effects of diverse severe weather types—including thunderstorms, winter storms, and tropical cyclones—on power grid infrastructure. They develop several quantitative metrics to assess disruption severity, providing policymakers and utility operators with useful insights. In light of growing renewable energy integration, Watson and Etemadi [24] evaluate grid resilience under hurricane conditions when solar and wind generation constitute a large share of capacity. Their simulations indicate markedly higher generation shortfalls and elevated restoration expenses in systems heavily reliant on renewables. Using Hurricane Maria as a case study, Kwasinski *et al.* [25] analyze the comprehensive effects of the storm on Puerto Rico's electricity infrastructure across generation, transmission, and distribution segments. Findings show that the island's grid was less resilient than mainland U.S. systems during hurricane strikes, underscoring the need for additional capital investment and technological advancements. Fatima *et al.* [26] conduct a comprehensive survey of techniques for predicting power outages during hurricanes, concluding that reliance on a single machine learning algorithm often introduces substantial bias, which can be mitigated through ensemble approaches. They further highlight that future efforts should prioritize improvements in data quality, algorithmic sophistication, and model explainability. Nonetheless, while these works predominantly address the consequences of extreme weather for grid robustness and financial impacts, their implications for the specific challenge of load forecasting remain underexplored.

Research gap

Drawing from the survey, three primary shortcomings emerge in the domain of abnormal load variation forecasting, outlined as follows:

First, within the realm of feature engineering, prior research [3-10] has concentrated predominantly on measuring and modeling relationships between load and weather variables. A standardized, numerically grounded definition of abnormal load variations is still missing, despite its potential to serve as a valuable additional feature that could strengthen forecasting performance.

Second, in the development of prediction models, earlier contributions [11-19] have largely emphasized the creation of sophisticated architectures capable of capturing intricate load dynamics. Yet because abnormal load variations are relatively rare, the corresponding training samples are severely limited—often insufficient to optimize increasingly elaborate models effectively. Consequently, there is a clear need for data augmentation techniques that can expand the modest set of abnormal variation instances to a larger volume, thereby improving the robustness of model training.

Third, regarding extreme weather events (such as hurricanes), previous investigations [20-26] have primarily examined their effects on overall system resilience and associated economic damages, with comparatively little attention devoted to their direct influence on load forecasting accuracy. Given that such events can dramatically alter typical load profiles, addressing the forecasting challenge under these conditions constitutes an important and understudied area.

Contribution

This paper introduces a hybrid prediction framework specifically designed to address abnormal load variations in urban environments. Initially, Seasonal and Trend decomposition using Loess (STL) [27] is applied to the raw load time series. As a well-established time-series decomposition technique, STL separates the series into trend, seasonal, and residual components, making it especially appropriate for power load data, which exhibits multiple periodic patterns. Following STL decomposition, abnormal load variation episodes are defined by examining the residual component, which captures irregular and non-repetitive load fluctuations.

Next, Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP) [28] are employed to expand the scarce set of abnormal load variation samples into a substantially larger collection, thereby overcoming the data scarcity challenge. Unlike standard GANs, WGAN-GP learns the underlying distribution from limited abnormal samples more reliably, generates diverse synthetic instances, improves training stability, and effectively mitigates mode collapse.

Finally, to effectively model the intricate and nonlinear dynamics observed during abnormal load variation periods, the cutting-edge TimesNet architecture [29] is adopted as the core forecasting engine. TimesNet processes the load series in the frequency domain to extract dominant periodic components, then reorganizes the data to simultaneously capture patterns within individual cycles and relationships across different cycles. This approach aligns closely with the multi-periodic nature of power load behavior, enabling TimesNet to learn sophisticated patterns during periods of anomalous variation robustly.

The primary contributions and novelties of this work are summarized below:

(1) A systematic quantitative approach is presented for identifying and describing abnormal load variation events. Using the residual component from STL decomposition, abnormal instances are identified as those falling in the tails of the residual probability distribution, indicating rare occurrences. A numerical abnormality score is then computed based on each sample's occurrence probability, which can serve as an informative feature to improve forecasting model performance. This method for quantitatively defining and characterizing abnormal load variations represents an original contribution of the present study.

(2) A tailored hybrid forecasting framework is developed to predict abnormal load variation. In contrast to conventional load forecasting techniques, the proposed framework integrates three key elements: a procedure for defining and quantifying abnormal load events, a data augmentation step to alleviate sample scarcity, and an advanced neural network as the predictor. This combination is customized to the unique challenges of abnormal load forecasting. To the authors' best knowledge, such an integrated and specialized framework targeting abnormal load conditions has seldom appeared in the prior literature.

Materials and Methods

The overall structure of the proposed forecasting framework for abnormal load variations is outlined in **Figure 2**. The process begins by applying STL decomposition to the historical load time series, separating it into trend, seasonal, and residual components. Next, abnormal load variation events are detected by analyzing the residual component; the corresponding samples are then augmented to create a substantially larger dataset, thereby strengthening the training of the forecasting model. Subsequently, TimesNet is used to independently predict each of the three decomposed components (trend, seasonal, and residual), and the individual forecasts are combined to yield the final predicted load values. Benefiting from an enriched set of abnormal-variation samples and the robust pattern-learning ability of TimesNet, the presented framework more effectively captures the complex behaviors observed during abnormal load variation periods, ultimately leading to enhanced forecasting accuracy.

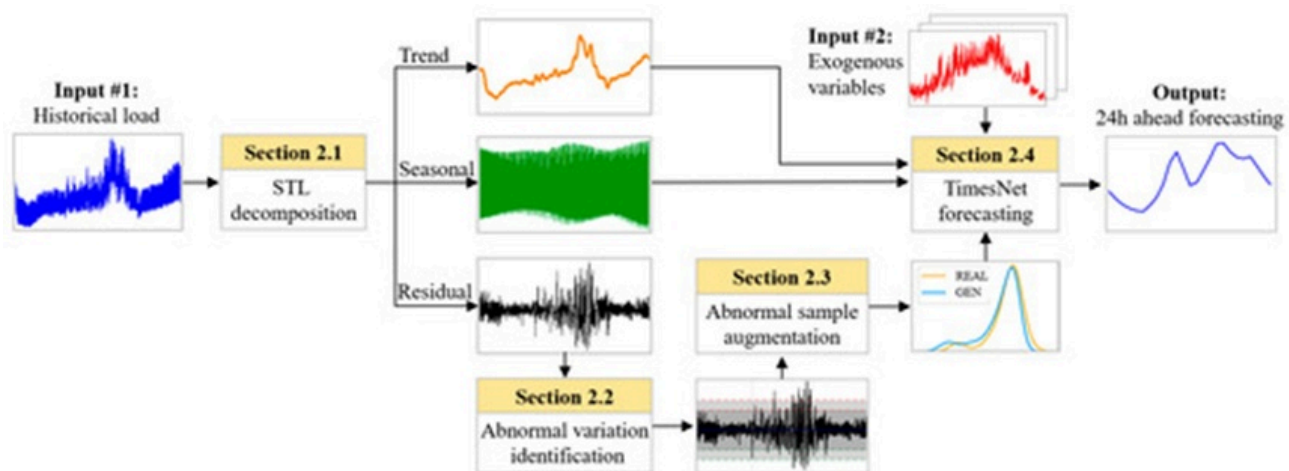


Figure 2. Flowchart of the proposed abnormal load variation forecasting framework.

STL algorithm for load profile decomposition

The Seasonal-Trend decomposition using Loess (STL) is a widely recognized time series decomposition technique, especially well-suited for sequences that exhibit distinct trend and seasonal patterns [27]. For a power load time series denoted as Y , the STL procedure can be expressed as

$$Y = T + S + R \quad (1)$$

Where T represents the low-frequency trend component, S denotes the high-frequency seasonal component, and R stands for the residual component that captures irregular and non-systematic load fluctuations.

The core mechanism of STL relies on Locally Weighted Regression (Loess), denoted by the smoothing function $g(\cdot)$. Let x be the sequence of independent variables corresponding to Y . For any point $x_0 \in x$, consider a local neighborhood containing q neighboring samples, denoted x_i ($i = 1, 2, \dots, q$). The tricube weight function determines the weight assigned to the i th sample

$$w_i = \begin{cases} (1 - u^3)^3, & 0 \leq u < 1 \\ 0, & u \geq 1 \end{cases} \quad (2)$$

$$u = \frac{|x_i - x_0|}{\lambda_q(x_0)} \quad (3)$$

Where $\lambda_q(x_0) = \max(|x_0 - x_i|)$, $i = 1, 2, \dots, q$ is the maximum distance from x_0 to its q nearest neighbors. The local estimate $\hat{y}_0$ at x_0 is then obtained through a weighted polynomial regression fitted to the q neighboring points (x_i, w_i) :

$$\hat{y}_0 = g(x_0) = \beta_0 + \beta_1 x_0 + \beta_2 x_0^2 + \dots + \beta_D x_0^D \quad (4)$$

In which the coefficients $\beta_0, \beta_1, \dots, \beta_D$ are determined via weighted least squares regression using the q local samples, and D is the chosen polynomial degree.

The application of STL to decompose power load series involves the following main stages:

(1) Initialization. Apply Loess smoothing to the original load series Y to extract an initial trend component $T = g(x)$. The seasonal component S is preliminarily set by averaging every L_s sample across Y , and the residual is initialized as $R = Y - S - T$. The robustness weight vector p is initialized as a vector of ones.

(2) Detrending. Remove the current trend T from Y to obtain a detrended series containing the seasonal and residual components.

(3) Smoothing the detrended series. Perform Loess smoothing on successive segments of the detrended series to generate a smoothed version. To ensure continuity at segment boundaries, Loess is reapplied to the entire smoothed, detrended series, yielding the final smoothed, detrended series C .

(4) Seasonal component estimation. Since C may still include low-frequency variations, a low-pass filter (implemented as a moving average with window length L_s) is applied to C to estimate and subtract this low-frequency part L , resulting in the refined seasonal component $S = C - L$.

(5) Residual estimation. Compute the residual component as $R = Y - T - S$.

(6) Trend component re-estimation. Recompute the trend component by applying Loess smoothing with window length L_t to the deseasonalized series $Y - S$.

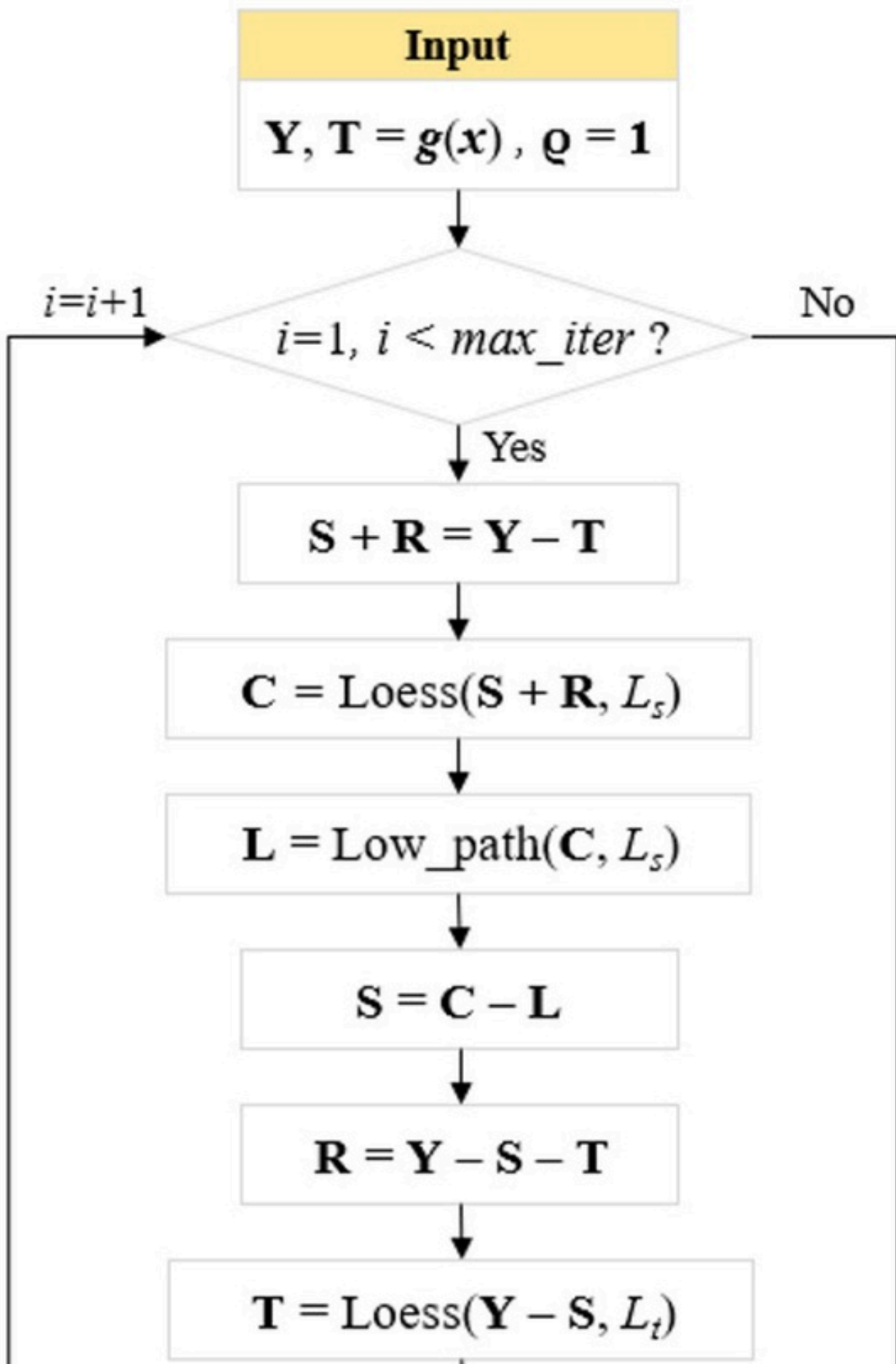
(7) Robustness weighting update. To mitigate the influence of large residuals, compute a robustness weight vector ρ using [Eqs. 5–7](#), where h is six times the median of the absolute residuals, and $B(\cdot)$ is the bisquare weighting function. The updated weights are obtained by multiplying the original weights w by ρ . This adjustment assigns higher weights to points with smaller residuals and lower weights to outliers, improving robustness in subsequent iterations.

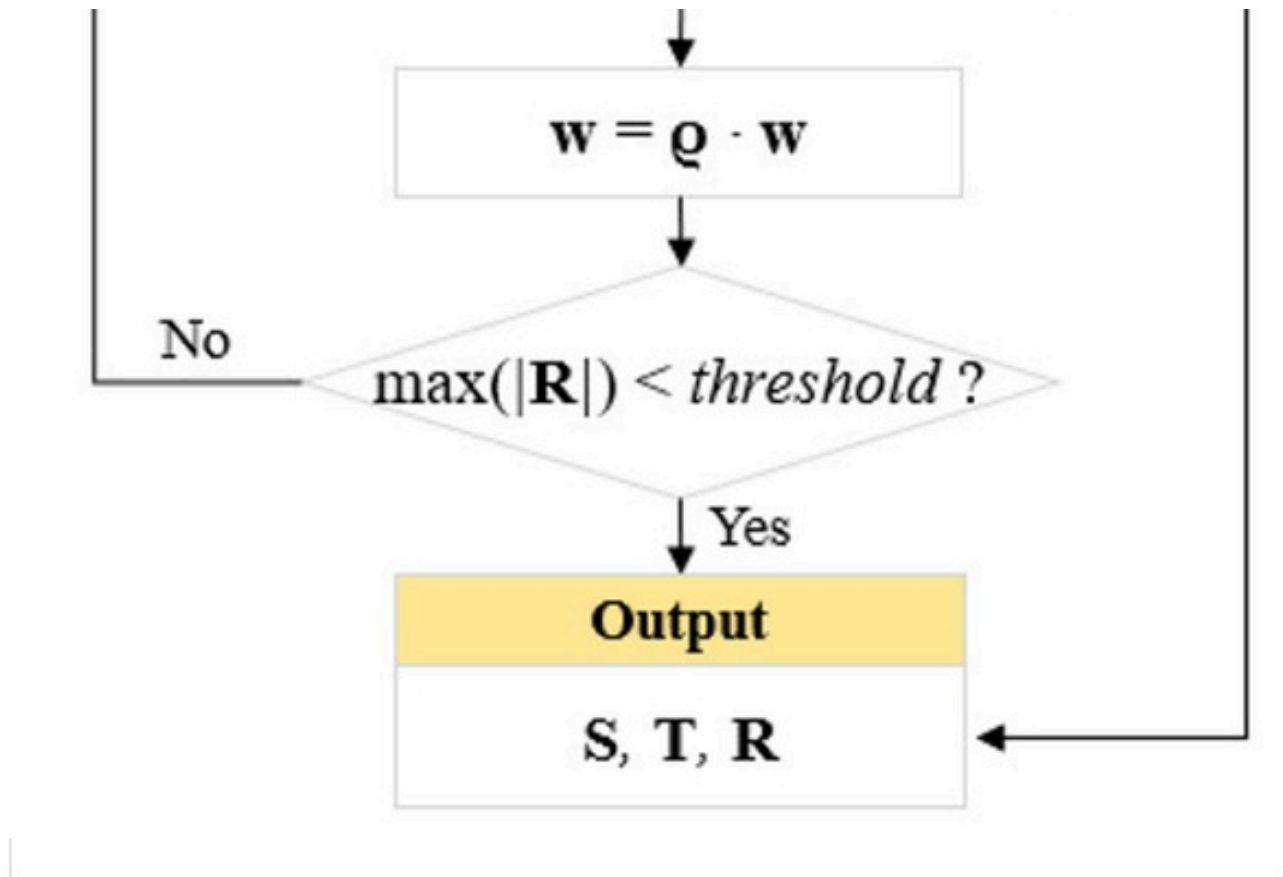
$$\rho = B\left(\frac{|\mathbf{R}|}{h}\right) \quad (5)$$

$$B(u) = \begin{cases} (1 - u^2)^2 & 0 \leq u < 1 \\ 0 & u \geq 1 \end{cases} \quad (6)$$

$$h = 6 \times \text{median}(|\mathbf{R}|) \quad (7)$$

The iterative cycle comprising steps (2) through (7) continues until the residual component R falls below a predefined convergence threshold. The trend T obtained in one iteration serves as the starting point for the next. A visual representation of the STL decomposition process, along with its pseudocode, is provided in [Figure 3](#) and [Algorithm 1](#).





Algorithm 1. STL decomposition algorithm

Abnormal load variation identification and characterization

Based on the outcomes of STL decomposition, the trend component is understood to reflect long-term load evolution, influenced by factors such as economic expansion, population growth, and climate shifts. The seasonal component captures recurring load behavior tied to periodic influences, such as seasonal cycles. Both trend and seasonal elements exhibit predictable and structured patterns. In contrast, the residual component accounts for unpredictable load deviations arising from irregular influences, including sudden weather shifts, social activities, or random fluctuations in user demand. Accordingly, this study identifies and quantifies abnormal load variations primarily through analysis of the residual component.

Given that the residual component lacks a clear systematic structure and resembles random noise, it is modeled as a Gaussian distribution, as expressed in Eqs. 8–10. Here, N denotes the length of the decomposed residual series R ; $r \in R$ represents an individual residual value; and μ_r and σ_r are the mean and standard deviation, respectively, of the fitted Gaussian distribution $G(r)$.

$$\mu_r = \frac{1}{N} \sum_{i=1}^N r_i \quad (8)$$

$$\sigma_r = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - \mu_r)^2} \quad (9)$$

$$G(r) = \frac{1}{\sigma_r \sqrt{2\pi}} e^{-\frac{(r - \mu_r)^2}{2\sigma_r^2}} \quad (10)$$

Drawing on $G(r)$, a discrete abnormality indicator $\mathbb{Z}$ is introduced to classify varying degrees of abnormal load variations as follows:

- Severe abnormal load variation: when $r \in (-\infty, \mu_r - 3\sigma_r] \cup [\mu_r + 3\sigma_r, +\infty)$, the deviation is considered extremely rare and substantial, lying well outside the typical range of load fluctuations. These pronounced anomalies are typically triggered by extreme meteorological conditions or major societal occurrences and pose significant forecasting challenges. In such cases, $\mathbb{Z}$ is assigned a value of 1.
- Mild abnormal load variation: when $r \in (\mu_r - 3\sigma_r, \mu_r - 2\sigma_r] \cup [\mu_r + 2\sigma_r, \mu_r + 3\sigma_r)$, the variation is deemed abnormal yet less extreme and less infrequent than the severe category. Here, $\mathbb{Z}$ is set to 0.5.
- Normal load variation: when $r \in (\mu_r - 2\sigma_r, \mu_r + 2\sigma_r)$, the fluctuations remain within a conventional range and are regarded as ordinary load variations. These are generally attributable to routine stochastic factors—such as daily behavioral variability, production schedule adjustments, and equipment outages and maintenance—and do not substantially impair forecasting accuracy. In this case, $\mathbb{Z}$ is assigned a value of 0.

The proposed feature $\mathbb{Z}$ can serve as a supplementary input variable for subsequent load forecasting models, enabling them to place greater emphasis on periods of abnormal load behavior and thereby enhancing predictive performance during such challenging conditions.

Abnormal load variation sample augmentation

To mitigate the challenge of limited samples for abnormal load variation forecasting, Generative Adversarial Networks (GANs) [30] are utilized to synthetically expand the small set of abnormal load variation instances into a much larger collection before training the forecasting model.

As depicted in **Figure 4**, a GAN architecture consists of two main networks: a generator (G) and a discriminator (D). The generator receives a random latent vector z —typically drawn from a Gaussian noise distribution—and produces synthetic samples $\hat{x} = G(z)$ that aim to mimic the distribution of genuine samples x . The discriminator evaluates inputs from both the real distribution x and the generated distribution $\hat{x}$, assigning labels of “real” or “fake.” Training proceeds as an adversarial game in which G strives to generate samples indistinguishable from real data, while D aims to distinguish them from real data correctly. This competition is mathematically expressed as a minimax optimization problem:

$$\min_G \max_D [\mathbb{E}_{x \in P_r} [\log D(x)] + \mathbb{E}_{\hat{x} \in P_g} [\log(1 - D(\hat{x}))]] \quad (11)$$

Where E denotes the expectation, and P_r and P_g represent the probability distributions of the real and generated data, respectively.

To improve training stability and prevent mode collapse—a common issue in standard GANs—this work adopts the Wasserstein GAN (WGAN) with Gradient Penalty (GP). The corresponding loss function is modified to:

$$\min_G \max_D [E_{\hat{x} \in P_g} [D(\hat{x})] - E_{x \in P_r} [D(x)]] + \lambda E_{\tilde{x} \in P_{\tilde{x}}} [(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2] \quad (12)$$

Where $\tilde{x} = \varepsilon x + (1-\varepsilon)\hat{x}$ with ε uniformly sampled from $[0,1]$, $P_{\tilde{x}}$ is the distribution of interpolated samples $\tilde{x}$, ∇ is the gradient operator, and λ is a weighting hyperparameter that enforces the Lipschitz constraint.

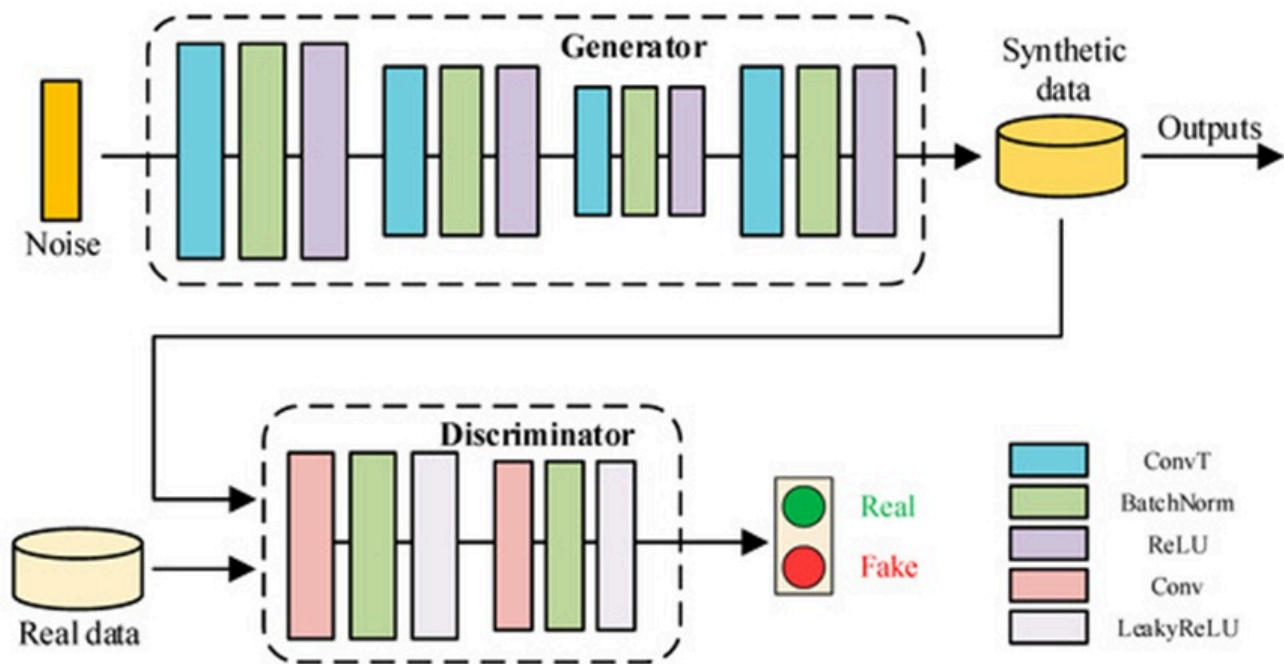


Figure 4. GAN framework in this paper.

It is worth noting that the discriminator loss in WGAN-GP reflects the Wasserstein distance between the real and generated distributions. Convergence is therefore indicated when the discriminator loss stabilizes near zero, signifying that the generated samples closely match the real data distribution. Additionally, an independent performance metric—the Fréchet Inception Distance (FID) score—is employed to assess the quality of generated samples and confirm model convergence:

$$FID = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}) \quad (13)$$

Where μ_r and μ_g are the mean vectors of the real and generated samples, Σ_r and Σ_g are their covariance matrices, and $\text{Tr}(\cdot)$ denotes the trace of a matrix. The WGAN-GP is considered converged when the FID score levels off near zero.

Following the augmentation procedure, the synthetic samples are combined with the original real samples to form an enriched training dataset. This expanded set enables the forecasting model to allocate greater focus to abnormal load

variation events, thereby improving predictive accuracy under such conditions.

TimesNet

This section presents TimesNet [29], an advanced neural network architecture selected as the primary forecasting engine to model intricate patterns during abnormal load variation periods effectively. As illustrated in **Figure 5**, TimesNet is built upon a stack of TimesBlock modules. The internal processing within each TimesBlock consists of the following key stages:

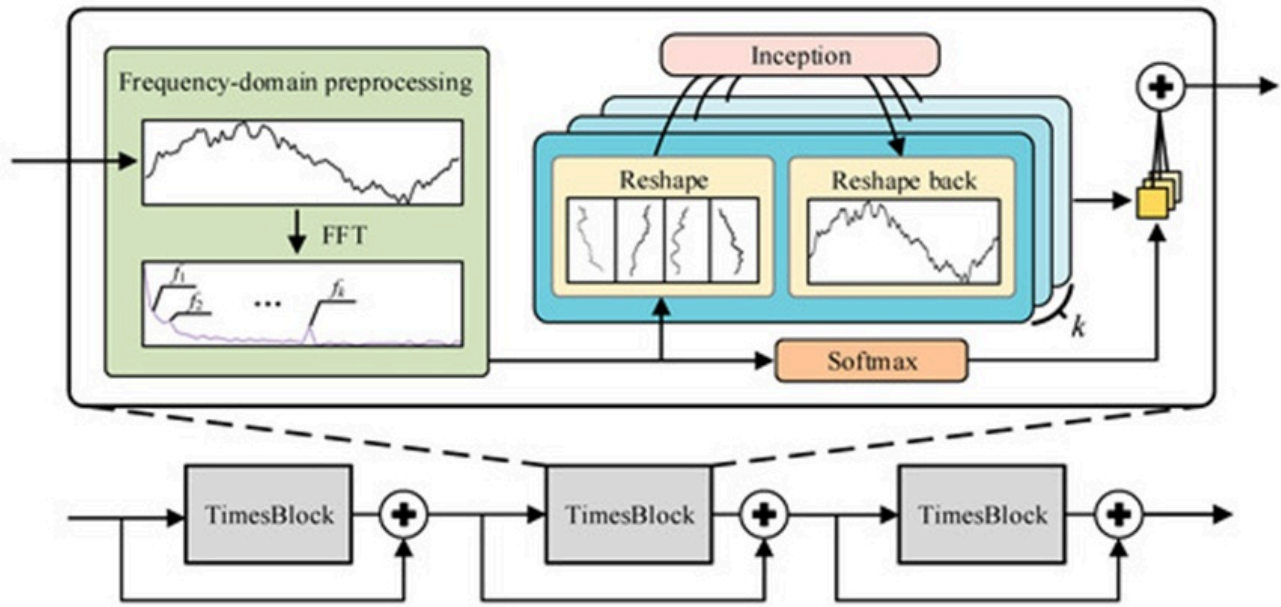


Figure 5. TimesNet framework.

(1) Frequency-domain preprocessing. Real-world time series, such as power load data, often comprise multiple superimposed components with diverse periodicities, which complicates direct learning. To overcome this, TimesBlock first converts the input time series into the frequency domain using the Fast Fourier Transform (FFT). For a one-dimensional multivariate time series $\mathbf{X}_{1D} \in \mathbb{R}^{T \times C}$ with temporal length T and C channels (e.g., load together with explanatory variables), the dominant periodic components are identified by:

$$\mathbf{A} = \text{avg}(\text{amp}(\text{FFT}(\mathbf{X}_{1D}))) \quad (14)$$

$$f_1, \dots, f_k = \underset{f \in \{1, \dots, T/2\}}{\text{argTop-}k}(\mathbf{A}) \quad (15)$$

$$p_1, \dots, p_k = \frac{T}{f_1}, \dots, \frac{T}{f_k} \quad (16)$$

Where $\text{FFT}(\cdot)$ computes the Fourier transform, $\text{amp}(\cdot)$ extracts amplitude values for each frequency, and $\text{avg}(\cdot)$ averages across channels. The top- k frequencies with the highest amplitudes are selected and denoted $\{f_1, \dots, f_k\}$. The

corresponding periods are then computed as $p_i = T / f_i$ for $i = 1, \dots, k$, yielding the set $\{p_1, \dots, p_k\}$.

(2) Reshaping time series into 2D tensors. Using the identified periods $\{p_1, \dots, p_k\}$, the original 1D series X_{1D} is reorganized into multiple 2D representations:

$$X_{2D}^i = \text{Reshape}_{[p_i, f_i, C]}(\text{Padding}(X_{1D})), \quad i \in \{1, \dots, k\} \quad (17)$$

Where $\text{Padding}(\cdot)$ appends zeros along the time axis to enable reshaping into tensors of shape $[p_i, f_i, C]$. The resulting collection of 2D tensors $\{X_{2D}^1, X_{2D}^2, \dots, X_{2D}^k\}$ encodes intra-period dynamics along rows and inter-period relationships along columns.

(3) Feature extraction by the Inception network. An Inception-style module [31], built from 2D convolutional layers, is applied to each 2D tensor to extract multi-periodic features:

$$\hat{X}_{2D}^i = \text{Inception}(X_{2D}^i) \quad (18)$$

Where $\hat{X}_{2D}^i$ is the resulting feature map, preserving the same spatial dimensions as the input 2D tensor.

(4) Reshaping the feature map back to a 1D series. The learned feature maps $\hat{X}_{2D}^i$ are converted back into one-dimensional sequences for subsequent processing and aggregation:

$$\hat{X}_{1D}^i = \text{Trunc}(\text{Reshape}_{[1, p_i \times f_i]}(\hat{X}_{2D}^i)), \quad i \in \{1, \dots, k\} \quad (19)$$

Where $\text{Trunc}(\cdot)$ removes the zero-padding previously added, restoring the original temporal length.

(5) Dynamic fusion. The resulting one-dimensional feature sequences, where $\hat{X}_{1D}^i$, $i \in \{1, 2, \dots, k\}$, are merged into a composite sequence by applying weights derived from the amplitude vector A , which encodes the strength of each sequence in the frequency domain.

$$\hat{X}_{1D} = \sum_{i=1}^k \frac{A[i]}{\|A\|} \times \hat{X}_{1D}^i \quad (20)$$

In implementation, multiple TimesBlocks are chained together to form the full TimesNet model. The 1D feature output $\hat{X}$ generated by the $(l-1)^{\text{th}}$ TimesBlock is fed directly into the l^{th} block, allowing progressive and thorough capture of time-dependent patterns present in the raw power load data. Furthermore, skip (residual) connections are added between neighboring TimesBlocks to stabilize training and boost overall effectiveness, especially as network depth increases.

$$\mathbf{X}_{1D}^l = TimesBlock(\mathbf{X}_{1D}^{l-1}) + \mathbf{X}_{1D}^{l-1} \tag{21}$$

Results and Discussion

Experimental setup

This study’s evaluation was conducted using real-world historical electricity load and meteorological records collected in Chongqing, China. The time span runs from January 1, 2020, through August 23, 2023, with data recorded every hour; the series is visualized in [Figure 6](#). Distinct outliers and sharp fluctuations are visually evident, most notably during the summer and winter months when cooling and heating demand dominate consumption. The prediction task involves taking the most recent 24 hours of load and weather observations to generate a forecast for the coming 24 hours, thereby supporting day-ahead operational planning [\[32\]](#). The full dataset was partitioned into 70% for training and 30% for testing to assess model accuracy.

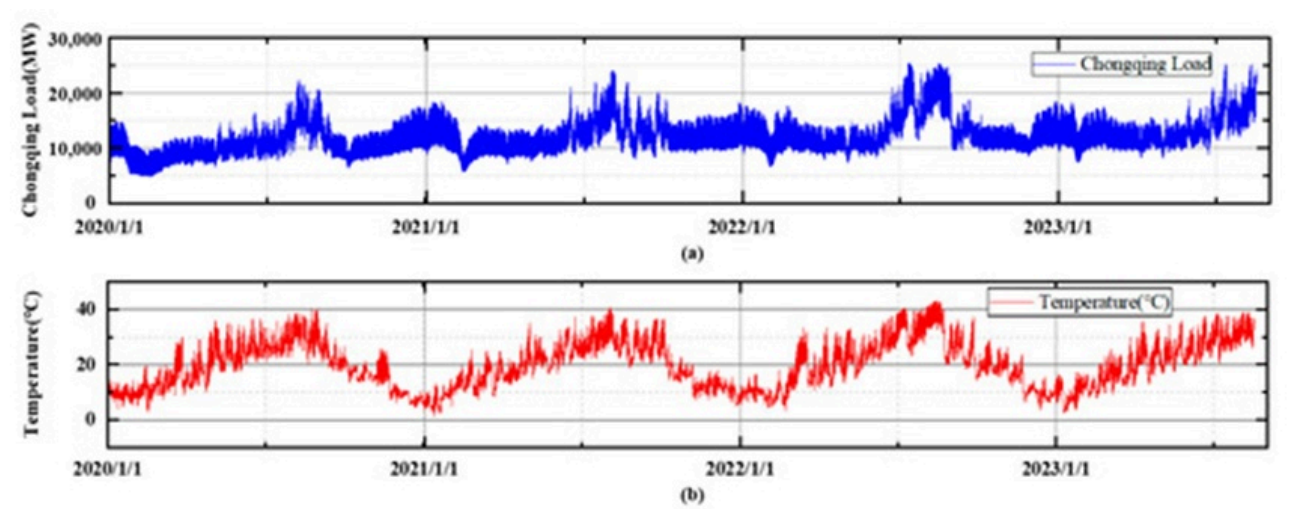


Figure 6. Historical hourly data recorded in Chongqing: (a) electricity load; (b) ambient temperature.

For performance comparison, the proposed TimesNet architecture was benchmarked against three other deep learning approaches: Long Short-Term Memory networks (LSTM) [\[33\]](#), Temporal Convolutional Networks (TCN), and a hybrid model combining TCN with Gated Recurrent Units (TCN-GRU). Optimal hyperparameter values for each model—determined through extensive grid-search and manual tuning—are listed in [Table 1](#). As a non-learning baseline, the similar-day method was also included, which replicates the load curve from the previous 24 hours as the forecast for the next 24 hours.

Table 1. Hyperparameter settings for the compared models.

Model	Hyperparameter configuration
LSTM	3 hidden layers, 48 neurons, 0.1 dropout rate, Adam optimizer, MSE loss function
TCN	2 layers, 3 kernel sizes, 64 convolution kernels, dilation rates [1, 2], 0.3 dropout rate, Adam optimizer, MSE loss function
TCN-GRU	2 TCN layers, 3 kernel sizes, 64 convolution kernels, dilation rates [1, 2], 2 GRU layers, 96 neurons, 0.2 dropout rate, 0.0001 learning rate, Adam optimizer, MSE loss function

TimesNet	6 encoder layers, 1 decoder layer, 256 neurons, 0.1 dropout rate, 0.0001 learning rate, Adam optimizer, MSE loss function
----------	---

Model quality was quantified with five well-established error measures: Mean Squared Error (MSE), its normalized variant (nMSE), Mean Absolute Error (MAE), normalized MAE (nMAE), and Mean Absolute Percentage Error (MAPE) [34]. The normalized versions nMSE and nMAE were included to reduce sensitivity to very small load values that can distort MAPE. To keep the presentation concise, only the predictions corresponding to the first hour of each 24-hour forecast horizon are displayed and analyzed below.

Forecasting without preprocessing (direct approach)

Here, the models are applied directly to the unprocessed historical time series, omitting STL seasonal-trend decomposition, outlier/anomaly handling, and data augmentation. These plain results establish a performance floor for each model and serve as a reference point for evaluating the value added by the full proposed pipeline in later sections. Test-set predictions appear in Figure 7 and are quantitatively summarized in Table 2.

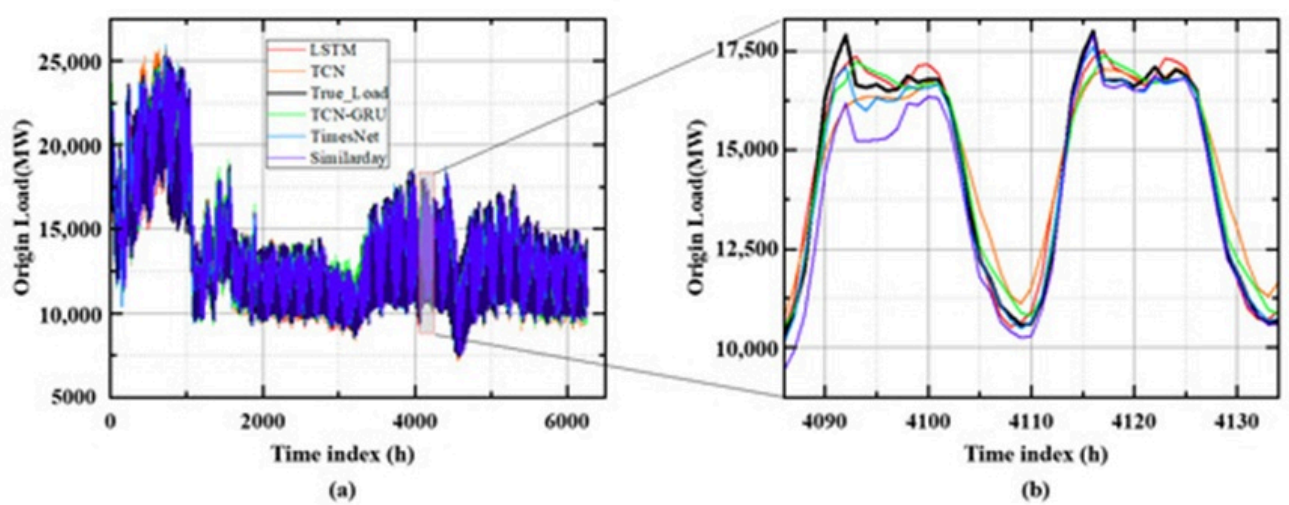


Figure 7. Direct forecasting outcomes on the test period: (a) complete test-set overview; (b) magnified view of selected regions.

Table 2. Quantitative accuracy summary for direct load forecasting.

Metric	LSTM	TCN	TCN-GRU	TimesNet	Similar Day
MSE (MW ²)	1,002,757	978,383	896,066	649,335	846,459
Normalized MSE	0.084	0.082	0.075	0.055	0.068
MAE (MW)	709	696	680	490	556
Normalized MAE	0.051	0.050	0.049	0.035	0.040
MAPE (%)	4.88	4.87	4.72	3.33	3.86

As shown in Table 2, TimesNet outperforms all competitors across all metrics, achieving gains of 1.55% over LSTM, 1.53% over TCN, and 1.39% over TCN-GRU. Interestingly, the elementary similar-day strategy also delivers competitive

accuracy in this unprocessed setting, which aligns with the strong calendar-day regularity typically observed in urban-scale electricity demand patterns.

STL-based decomposition and component-wise prediction

This subsection begins by splitting the input load time series into three parts using STL: a slowly varying trend component, a repeating seasonal component, and a residual (irregular) component. Predictions are generated for each part and then combined by simple addition to yield the complete forecasted load profile. This breakdown strategy enables analysis of model behavior across distinct signal types and quantification of the benefit of applying STL decomposition.

The decomposition relies on two main control parameters: the seasonal smoothing window L_s and the cycle length L_p . L_s governs how wide a neighborhood the Loess smoother uses when estimating the seasonal signal—choosing a bigger L_s creates a gentler, less wiggly seasonal curve. In contrast, a smaller L_s allows more rapid changes. L_p sets the assumed repeat interval of the underlying periodic behavior; increasing L_p tends to produce a flatter, more stable trend estimate, while decreasing it allows the trend to follow shorter variations more closely. After systematic testing, the combination $L_s = 25$ and $L_p = 168$ was found to yield the cleanest, most balanced separation of components. See the resulting decomposition in [Figure 8](#).

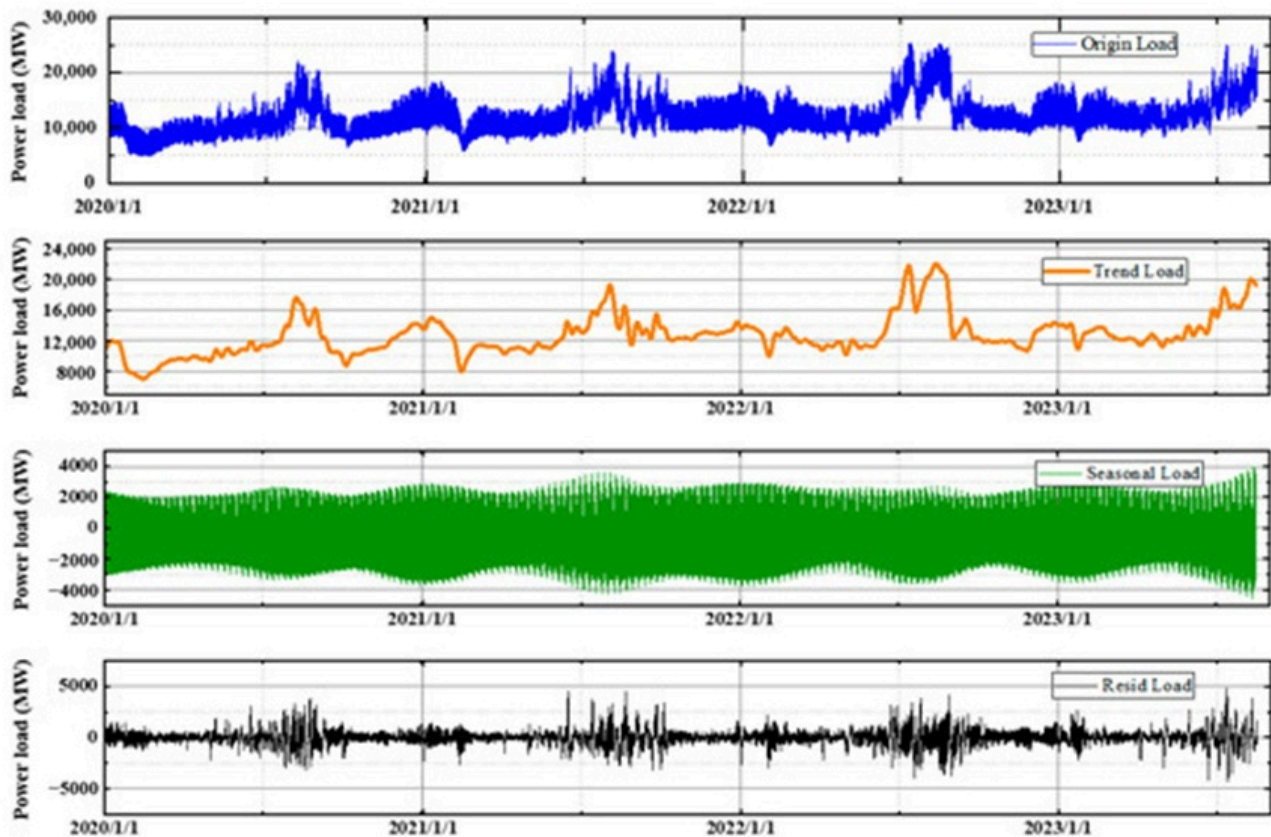


Figure 8. Original load series together with its STL-extracted components.

[Figure 8](#) clearly illustrates that the extracted trend is very smooth and faithfully tracks the gradual, multi-month evolution visible in the raw load data. The seasonal signal repeats reliably with an average near zero. Both trend and seasonal signals display structured, low-noise behavior. By contrast, the residual signal contains mostly erratic, seemingly random fluctuations. Because these three parts exhibit fundamentally different dynamics, it makes sense to model and forecast each one separately before summing the predictions.

(1) Forecasting the trend component

Five candidate models were trained and tested specifically on the trend signal. Models were trained on trend samples from the training period and evaluated on the held-out test period; results appear in [Figure 9](#) and [Table 3](#). The newly proposed TimesNet clearly leads on this task, recording an MAPE of only 0.08%. Across the four reference models, it reduces MAPE by 0.32%-1.36%, highlighting its particular strength in modeling persistent, long-horizon load movements.

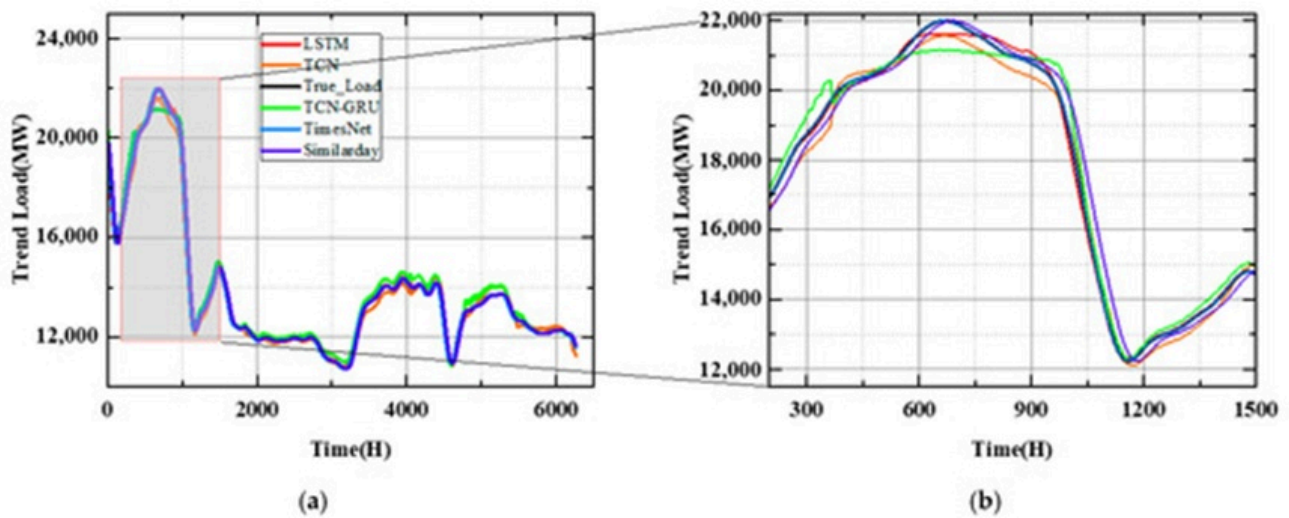


Figure 9. Trend-component predictions on the test data: (a) full test-period view; (b) close-up of selected intervals.

Table 3. Accuracy metrics for trend-component forecasting.

Metric	LSTM	TCN	TCN-GRU	TimesNet	Similar Day
MSE (MW ²)	9,911	64,877	51,362	435	75,564
Normalized MSE (nMSE)	0.0013	0.0086	0.0068	0.000053	0.0092
MAE (MW)	62	207	141	11	158
Normalized MAE (nMAE)	0.0044	0.015	0.010	0.00082	0.011
MAPE (%)	0.40	1.44	0.94	0.080	1.070

(2) Forecasting the seasonal component

The same modeling workflow was repeated for the seasonal part. Performance is documented in [Figure 10](#) and [Table 4](#). Since the seasonal signal fluctuates around a mean very close to zero, MAPE becomes unreliable (risk of division by near-zero) and is therefore omitted from the comparison. The results indicate that TCN-GRU and TimesNet achieve substantially lower errors than the other three approaches.

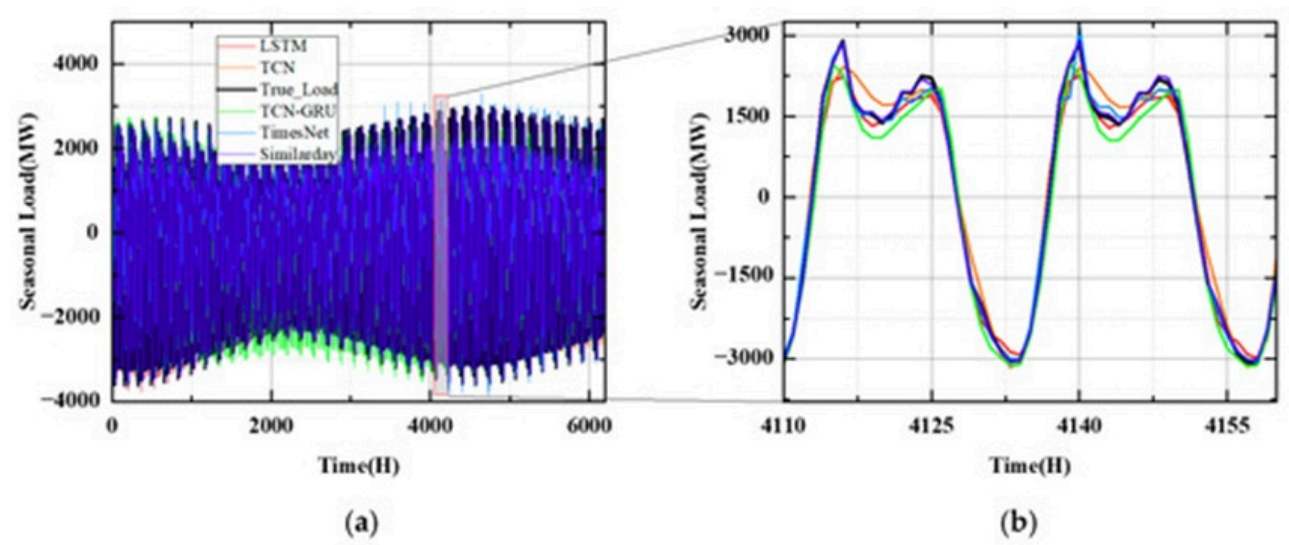


Figure 10. Seasonal-component predictions on the test data: (a) full test-period view; (b) close-up of selected intervals.

Table 4. Accuracy metrics for seasonal-component forecasting.

Metric	LSTM	TCN	TCN-GRU	TimesNet	Similar Day
MSE (MW ²)	157,186	182,484	95,018	98,102	141,784
Normalized MSE (nMSE)	0.050	0.058	0.030	0.032	0.046
MAE (MW)	289	331	244	197	258
Normalized MAE (nMAE)	0.18	0.21	0.16	0.13	0.16

(3) Forecasting the residual component

The residual captures unpredictable short-term load swings, where outdoor temperature is usually the single most important driver. For this reason, temperature readings are fed in as an auxiliary (exogenous) feature during residual modeling to boost accuracy. Key results are compiled in [Table 5](#) and [Figure 11](#). Even on this noisier component, TimesNet maintains a noticeable performance advantage.

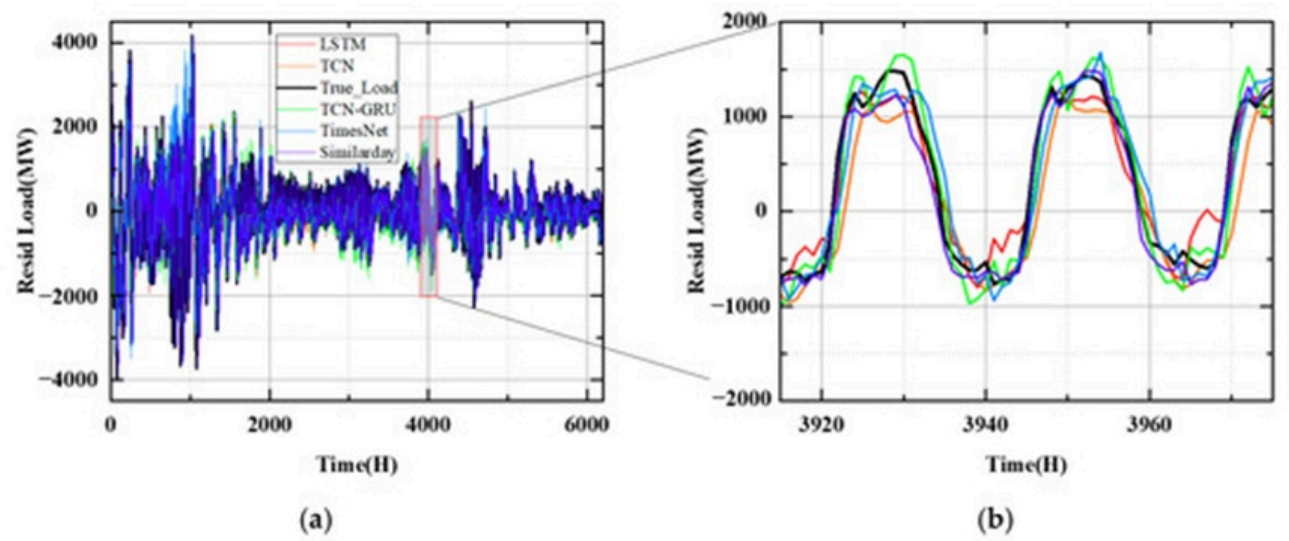


Figure 11. Residual-component predictions on the test data: (a) full test-period view; (b) close-up of selected intervals.

Table 5. Accuracy metrics for residual-component forecasting.

Metric	LSTM	TCN	TCN-GRU	TimesNet	Similar Day
MSE (MW ²)	467,428	517,234	462,912	454,757	577,822
Normalized MSE (nMSE)	0.58	0.64	0.57	0.56	0.86
MAE (MW)	445	473	440	413	434
Normalized MAE (nMAE)	0.72	0.76	0.71	0.67	0.76

Aggregated (final) load forecasting

Once the separate trend, seasonal, and residual forecasts are combined, the complete reconstructed load predictions are obtained, as shown in [Figure 12](#) and quantified in [Table 6](#).

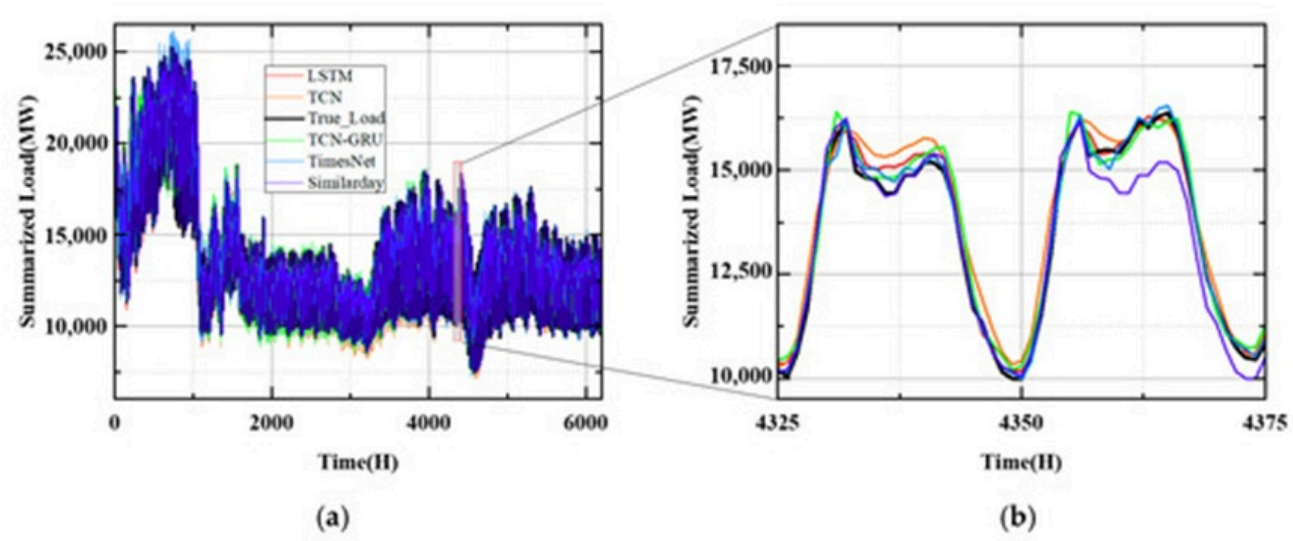


Figure 12. Final aggregated load predictions on the test data: (a) full test-period view; (b) close-up of selected intervals.

Table 6. Accuracy metrics for the complete (aggregated) load forecast.

Metric	LSTM	TCN	TCN-GRU	TimesNet	Similar Day
MSE (MW ²)	627,577	881,442	663,031	542,877	846,459
Normalized MSE (nMSE)	0.052	0.074	0.056	0.044	0.068
MAE (MW)	557	721	562	467	556
Normalized MAE (nMAE)	0.040	0.051	0.040	0.034	0.040
MAPE (%)	3.86	4.34	3.91	3.24	3.86

Across the board, TimesNet emerges as the top performer. Taken together, the experiments show that the presented TimesNet framework consistently delivers strong, reliable predictions regardless of whether the input is a smooth trend, a

periodic seasonal pattern, irregular residuals, or the full composite signal. Therefore, in the sections that follow, TimesNet is adopted as the primary model and further enhanced to improve forecast quality.

Prediction enhancement via abnormal variation labeling and data expansion

To direct the forecasting model's emphasis toward periods of unusual load behavior, abnormal events were first detected and labeled according to the rules. As depicted in [Figure 13](#), a probability density was fitted to the residual component, yielding estimates of its mean and standard deviation. The resulting labeling outcomes appear in [Figure 14](#). Data points falling within the region between upper bound 1 and lower bound 1 (shaded dark grey) represent typical fluctuations and receive a label $Z = 0$. Points lying between upper bound 1 and upper bound 2, or between lower bound 1 and lower bound 2 (shaded light grey), indicate moderate anomalies and are labeled $Z = 0.5$. All remaining points (unshaded/white regions) signify severe anomalies and are assigned $Z = 1$. This Z label is then added as an extra input feature to the forecasting model, enabling it to allocate greater focus to instances of irregular load changes.

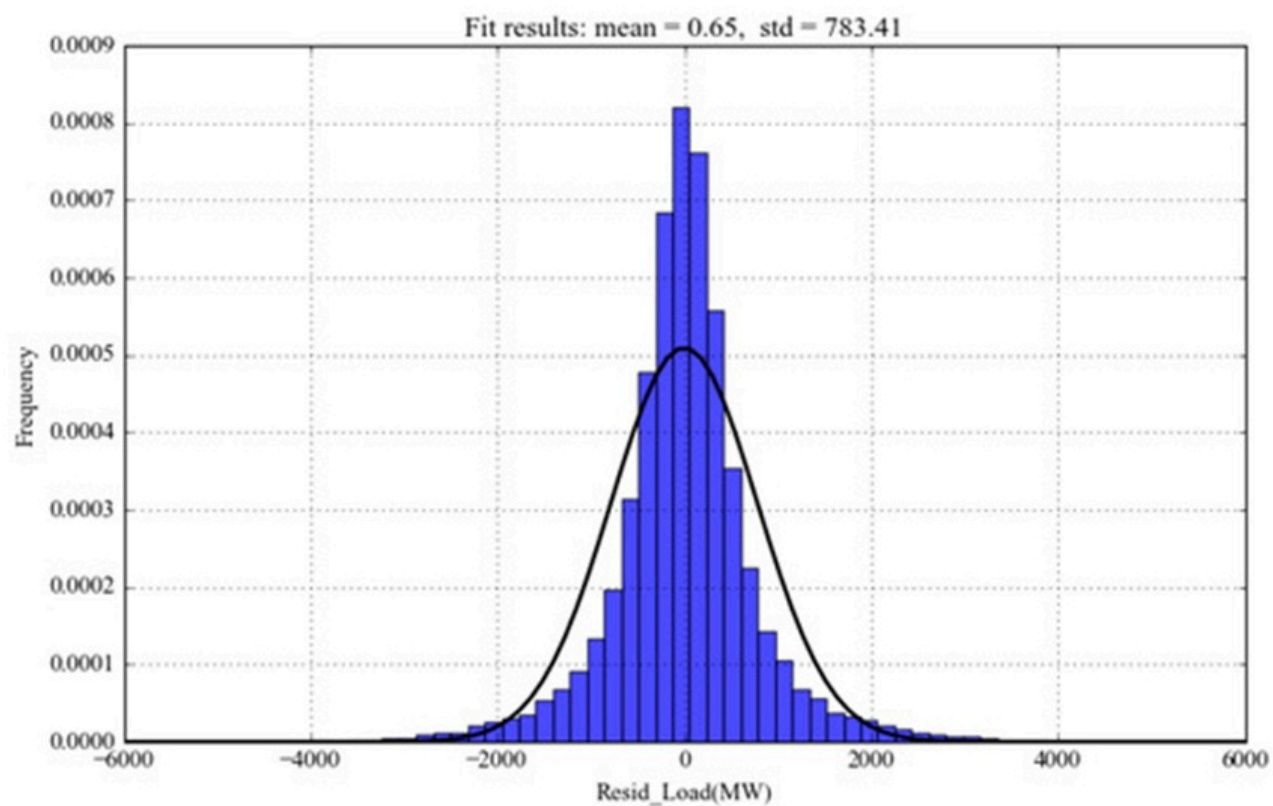


Figure 13. Fitted probability distribution for the residual component.

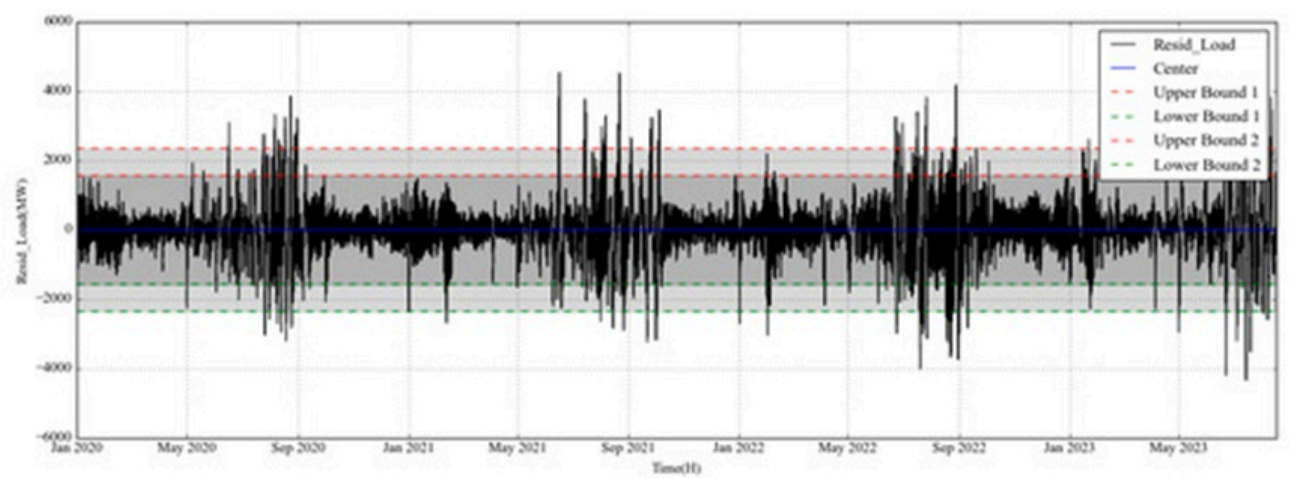


Figure 14. Labeling outcomes for abnormal load variation incidents.

Moreover, since anomalous samples form only a small fraction of the overall dataset, the method was applied to artificially increase their count, thereby strengthening the model's learning during training. **Figure 15** reveals that 625 severe abnormal samples ($Z = 1$) and 1408 moderate abnormal samples ($Z = 0.5$) were detected across the historical load record. A clear pattern emerges: the majority of these irregularities cluster in the summer months, driven primarily by extreme temperature swings. Each anomalous instance was reformatted into a prediction-ready structure, pairing the 24-hour target window containing the anomaly with the corresponding 24-hour historical input sequence, and supplementing it with temperature as an external covariate.

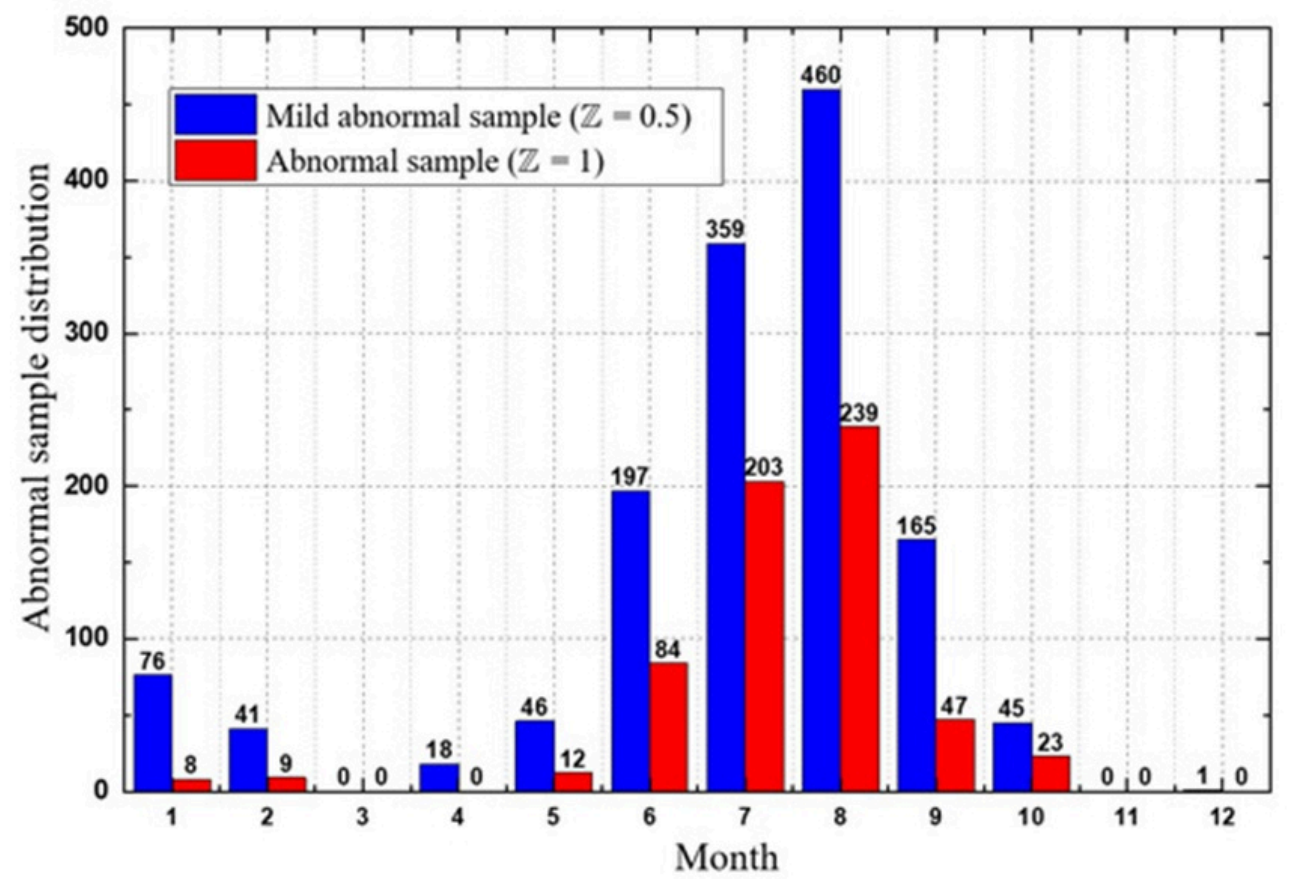


Figure 15. Yearly spread of abnormal load variation occurrences.

Next, a WGAN-GP architecture was trained on the collected abnormal samples to capture their underlying distribution. Training progress was monitored by tracking the generator loss (Gloss), the discriminator loss (Dloss), and the Fréchet Inception Distance (FID) score over 10,000 epochs, as shown in **Figure 16**. After approximately 1000 epochs, both the discriminator loss and FID stabilize near zero, signaling successful convergence and close alignment between the distributions of real and synthesized samples. At the 1000th epoch, the FID reaches 0.0028.

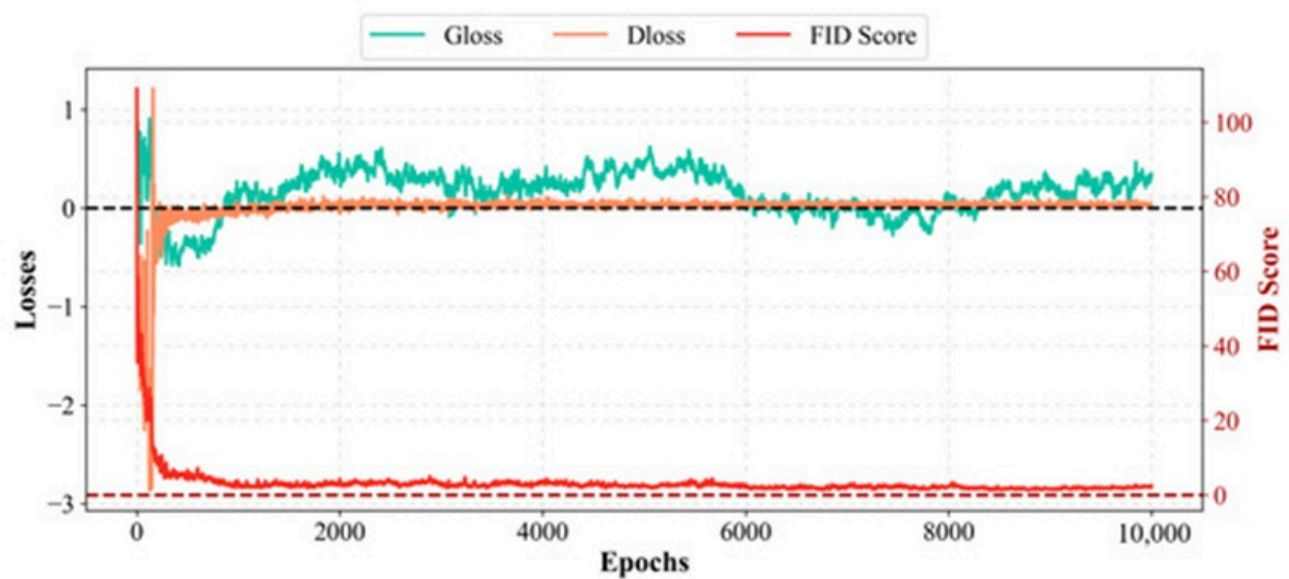


Figure 16. Convergence trajectory of the WGAN-GP during training.

With the WGAN-GP fully trained and stabilized, an additional 2000 artificial abnormal samples were produced. **Figure 17** compares the statistical profiles of the original abnormal samples against those of the newly created ones. The close matching of means and peak shapes across both sets confirms the high realism of the synthetic data.

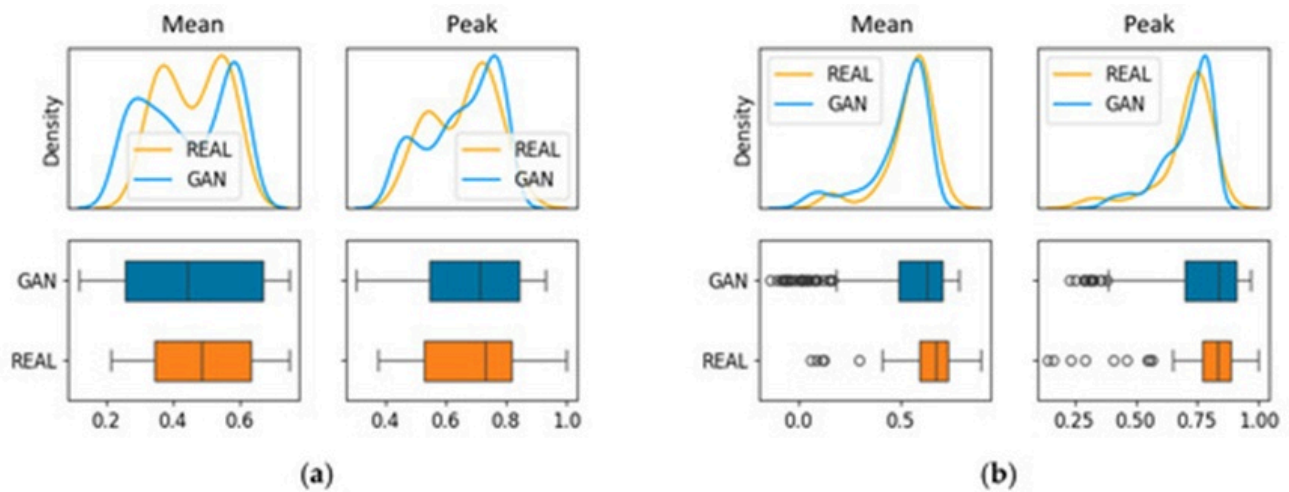


Figure 17. Comparison of distributions between real and synthetic abnormal samples: (a) load values, and (b) temperature values.

These synthetic abnormal instances were then added to the forecasting model's training data to improve its handling of outlier load behavior. The updated residual component prediction accuracies—incorporating the $\mathbb{Z}$ feature and the expanded samples—are reported in [Table 7](#), while performance specifically on the abnormal subset is detailed in [Table 8](#). Adding the $\mathbb{Z}$ indicator alone boosts TimesNet's accuracy on the residual component. However, including the augmented samples yields a modest decline in residual component performance but a substantial gain on the abnormal samples themselves. Relative to the no-augmentation baseline, MSE for the residual component rises by 2.65%, whereas MSE restricted to abnormal samples drops sharply by 38.86%.

Table 7. Prediction accuracy metrics for the residual component.

Metric	Sample augmentation	$\mathbb{Z}$ Included	Baseline
MSE (MW ²)	466,827	429,996	454,757
Normalized MSE (nMSE)	0.69	0.53	0.56
MAE (MW)	450	401	413
Normalized MAE (nMAE)	0.81	0.64	0.67

Table 8. Prediction accuracy metrics focused on abnormal load variation samples.

Metric	Sample Augmentation	$\mathbb{Z}$ Included	Baseline
MSE (MW ²)	2,086,058	2,739,755	3,411,852
Normalized MSE (nMSE)	0.43	0.51	0.63
MAE (MW)	1,172	1,263	1,344
Normalized MAE (nMAE)	0.54	0.56	0.60

The experiments were structured as a series of controlled tests to systematically verify the value added by each element of the suggested hybrid prediction pipeline. In greater detail:

- The superiority of the proposed TimesNet predictor is confirmed by comparing its accuracy with several established baselines under varied prediction conditions. While TimesNet generally delivers stronger results, the degree of its edge varies markedly across scenarios. For instance, when predicting the trend component, TimesNet attains remarkably low error with an MAPE of just 0.08% ([Table 3](#)), far surpassing the comparison models. In contrast, the seasonal component forecast lags slightly behind the TCN-GRU approach ([Table 4](#)). This pattern suggests that TimesNet excels particularly at handling smooth, stable sequences rather than highly erratic or volatile ones.
- A large body of prior research has established that breaking down the raw load time series into distinct sub-components, modeling each separately, and then recombining the outputs constitutes a robust and practical approach to load prediction. Evidence supporting this strategy emerges from a direct comparison between the outcomes in [Table 2](#) (direct prediction without decomposition) and [Table 6](#) (decomposition-based approach). Every machine learning model examined shows substantial gains in precision when using the decompose–predict–reassemble workflow rather than predicting the aggregate load directly. The reason lies in the fact that individual components reveal clearer, more isolated patterns that are far simpler for models to capture than the composite signal, which blends multiple overlaid dynamics.

Findings in **Tables 7 and 8** show that enriching the dataset with synthetic abnormal samples markedly improves prediction quality, particularly during outlier load episodes. Yet, it comes at the cost of a slight overall performance dip across all periods. This outcome is intuitive: by increasing the relative share of anomalous examples in the training pool, the model naturally shifts its learning emphasis toward the distinct, non-typical patterns that diverge sharply from routine behavior. The key takeaway is that careful calibration is essential when blending synthetic outliers into the full dataset—to maximize gains on abnormal cases without unduly sacrificing reliability on normal ones.

Additionally, the empirical findings from the case study highlight several remaining shortcomings in the proposed prediction pipeline that warrant further refinement:

The underlying drivers behind abnormal load deviations remain untraceable in the current setup. Anomaly detection relies solely on morphological traits of the load curve, without probing causal mechanisms. Consequently, although 625 abnormal load events were flagged, the specific triggers (e.g., weather extremes, events, or equipment issues) remain unidentified, limiting opportunities for deeper, cause-informed modeling. Future efforts should prioritize gathering richer exogenous data sources and performing causal inference to uncover the roots of these irregularities, potentially leading to more targeted and accurate forecasts.

The abnormality indicator $\mathbb{Z}$ is discretized into three levels (0, 0.5, 1) to gauge deviation severity. While straightforward, this coarse-grained scale fails to capture nuanced differences in anomaly intensity. For example, two instances both labeled $\mathbb{Z} = 1$ could exhibit markedly dissimilar behaviors. Adopting a continuous-valued abnormality metric in subsequent work could offer finer discrimination and better representational power.

Several aspects of the methodology involve subjective choices that can affect end-to-end performance. STL decomposition outcomes depend heavily on user-selected parameters, such as the seasonal window L_s and period L_p , whose optimal values are determined heuristically. Similarly, both the neural forecaster architecture and the anomaly labeling procedure incorporate empirically tuned elements that draw heavily on expert judgment. Achieving reliable results may therefore demand extensive iterative experimentation to fine-tune the integrated system.

Conclusion

This study introduces a specialized hybrid prediction pipeline designed to enhance the precision of urban load forecasting, particularly during periods of irregular load fluctuations. Core elements include a procedure for detecting and quantifying abnormal load events, a data-expansion technique to mitigate the rarity of outlier samples, and integration of the cutting-edge TimesNet neural architecture as the core predictor—rendering the framework well-adapted to the challenges of abnormal load prediction. Validation draws on three years of real hourly load and weather records from Chongqing, China, yielding the following principal insights: (1) Applying the proposed anomaly definition and labeling scheme successfully pinpointed 625 severe abnormal load events and 1408 moderate ones within the three-year dataset. Encoding these anomalies as a numeric input feature ($\mathbb{Z}$) and feeding it into the model lifts residual-component prediction accuracy by 5.44% across the entire test set and by 19.70% when restricted to abnormal instances. These gains affirm the practical utility of the anomaly characterization step for boosting forecast reliability under irregular conditions. (2) Leveraging the 625 real abnormal samples, an additional 2000 synthetic abnormal samples were created to enrich training. This augmentation yields a dramatic 38.86% reduction in error on abnormal cases, accompanied by only a minor 2.65% decline in accuracy across the full test set. Together, these controlled comparisons substantiate the overall effectiveness of the presented hybrid pipeline.

Recommended directions for future research include: First, investigating the influence of major societal disruptions—such as large public gatherings or significant policy changes—and developing quantitative measures of their effects on city-wide electricity demand to refine predictive models, and second, exploring a few-shot or meta-learning strategies to better address the inherent data scarcity in abnormal load scenarios. Third, advancing granular modeling of individual consumer

behaviors so that aggregate load forecasts can be derived more precisely by summing simulated end-user consumption patterns.

Acknowledgements

None

Conflict of interest

None

Financial support

This work was supported in part by the National Natural Science Foundation of China under Grant #52307121, and in part by the Shanghai Sailing Program under Grant #23YF1419000.

Ethics statement

None

Received: 26 Sep 2023

Revised: 03 Nov 2023

Accepted: 12 Dec 2023

Published online: 20 March 2024

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

Pecan Street. <https://www.pecanstreet.org> (accessed on 11 December 2024).

Time and Date. <https://www.timeanddate.com/weather/usa/austin> (accessed on 5 November 2024).

Fay D, Ringwood JV. On the influence of weather forecast errors in short-term load forecasting models. *IEEE Trans Power Syst.* 2010; 25:1751-8.

Chu WC, Chen YP, Xu ZW, Lee WJ. Multiregion short-term load forecasting in consideration of HI and load/weather diversity. *IEEE Trans Ind Appl.* 2010;47:232-7.

Yu B, Li J, Liu C, Sun B. A novel short-term electrical load forecasting framework with intelligent feature engineering. *Appl Energy.* 2022;327:120089.

Wang Y, Zhang N, Chen X. A short-term residential load forecasting model based on LSTM recurrent neural network considering weather features. *Energies*. 2021;14:2737.

Xie J, Chen Y, Hong T, Laing TD. Relative humidity for load forecasting models. *IEEE Trans Smart Grid*. 2016;9:191-8.

Hong T, Wang P, White L. Weather station selection for electric load forecasting. *Int J Forecast*. 2015;31:286-95.

Mansouri A, Abolmasoumi AH, Ghadimi AA. Weather sensitive short term load forecasting using dynamic mode decomposition with control. *Electr Power Syst Res*. 2023;221:109387.

Xie J, Hong T. Temperature scenario generation for probabilistic load forecasting. *IEEE Trans Smart Grid*. 2016;9:1680-7.

Dehalwar V, Kalam A, Kolhe ML, Zayegh A. Electricity load forecasting for urban area using weather forecast information. In: *Proceedings of the 2016 IEEE International Conference on Power and Renewable Energy (ICPRE)*; Shanghai, China; 2016. p. 355-9.

Lu Q, Cai Q, Liu S, Yang Y, Yan B, Wang Y. Short-term load forecasting based on load decomposition and numerical weather forecast. In *Proceedings of the IEEE: 2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*, Beijing, China, 2017; IEEE: Piscataway, NJ, USA, 2017. p. 1–5.

Xu L, Wang S, Tang R. Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load. *Appl Energy*. 2019;237:180-95.

Kim Y, Son H, Kim S. Short term electricity load forecasting for institutional buildings. *Energy Rep*. 2019;5:1270-80.

Li B, Lu M, Zhang Y, Huang J. A weekend load forecasting model based on semi-parametric regression analysis considering weather and load interaction. *Energies*. 2019;12:3820.

Pinheiro MG, Madeira SC, Francisco AP. Short-term electricity load forecasting—a systematic approach from system level to secondary substations. *Appl Energy*. 2023;332:120493.

Guo H, Huang B, Wang J. Probabilistic load forecasting for integrated energy systems using attentive quantile regression temporal convolutional network. *Adv Appl Energy*. 2024;14:100165.

Chen Y, Zhang D. Theory-guided deep-learning for electrical load forecasting (TgDLF) via ensemble long short-term memory. *Adv Appl Energy*. 2021;1:100004.

Li L, Ota K, Dong M. When weather matters: IoT-based electrical load forecasting for smart grid. *IEEE Commun Mag*. 2017;55:46-51.

Deng X, Ye A, Zhong J, Xu D, Yang W, Song Z, et al. Bagging-XGBoost algorithm based extreme weather identification and short-term load forecasting model. *Energy Rep*. 2022;8:8661-74.

Yang W, Sparrow SN, Wallom DCH. A comparative climate-resilient energy design: wildfire resilient load forecasting model using multi-factor deep learning methods. *Appl Energy*. 2024;368:123365.

Zhang T, Tang D, Fan P, Wang Q, Wang P. A probabilistic graphical model for predicting cascade failures of electric vehicle charging networks caused by hurricanes. *IEEE Trans Smart Grid*. 2024;16:627-39.

Shield SA, Quiring SM, Pino JV, Buckstaff K. Major impacts of weather events on the electrical power delivery system in the United States. *Energy*. 2021;218:119434.

Watson EB, Etemadi AH. Modeling electrical grid resilience under hurricane wind conditions with increased solar and wind power generation. *IEEE Trans Power Syst*. 2019;35:929-37.

Kwasinski A, Andrade F, Castro-Sitiriche MJ, O'Neill-Carrillo E. Hurricane Maria effects on Puerto Rico electric power infrastructure. *IEEE Power Energy Technol Syst J*. 2019;6:85-94.

Fatima K, Shareef H, Costa FB, Bajwa AA, Wong LA. Machine learning for power outage prediction during hurricanes: an extensive review. *Eng Appl Artif Intell*. 2024;133:108056.

Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: A seasonal-trend decomposition. *J Off Stat*. 1990;6:3-73.

Jin Q, Lin R, Yang F. E-WACGAN: Enhanced generative model of signaling data based on WGAN-GP and ACGAN. *IEEE Syst J*. 2019;14:3289-300.

Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M. TimesNet: Temporal 2D-variation modeling for general time series analysis. *arXiv*. 2022:arXiv:2210.02186.

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. *Commun ACM*. 2020;63:139-44.

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Las Vegas, NV, USA; 2016. p. 2818-26.

Gao Y, Ai Q. A novel optimal dispatch method for multiple energy sources in regional integrated energy systems considering wind curtailment. *CSEE J Power Energy Syst*. 2022;10:2166-73.

Shang Y, Li D, Li Y, Li S. Explainable spatiotemporal multi-task learning for electric vehicle charging demand prediction. *Appl Energy*. 2025;384:125460.

Feng C, Shao L, Wang J, Zhang Y, Wen F. Short-term load forecasting of distribution transformer supply zones based on federated model-agnostic meta learning. *IEEE Trans Power Syst*. 2024;40:31-45.