symmetry

MDPI

Article

An Improved Adversarial Learning Method for Cross-Scene Reconstruction of Industrial Load Symmetry Power Data Based on Denoising Diffusion

Yuxiu Zang [1], Jia Cui [1,*], Jiaqi Shi [2], Yan Zhao [2] and Weichun Ge [2]

[1] School of Electrical Engineering, Shenyang University of Technology, Economic & Technological Development Zone, No.111, Shenliao West Road, Shenyang 110000, China; xiudiubiu@163.com
[2] Shenyang Institute of Engineering, No. 18 Puchang Road, Shenbei New District, Shenyang 110000, China; phdshijiaqi@foxmail.com (J.S.); zhaoyan@sie.edu.cn (Y.Z.)
* Correspondence: cuijiayouxiang@sut.edu.cn

Abstract

Symmetry power integrity is a core issue for power system data acquisition. However, industrial load data integrity is affected by missing values, abnormal disturbances, and low-reliability observations. A reliability-aware cross-scene industrial load symmetry power data reconstruction method is proposed based on adversarial learning. Firstly, an industrial electricity scene classification is proposed. Temporal and frequency-domain features are jointly encoded by a multilayer perceptron. The scene affiliation of the data is identified by cosine similarity to improve the cross-scene generalization capability. Secondly, a diffusion denoising generative adversarial reconstruction framework is proposed. For missing data, a conditional diffusion model is constructed with historical temporal distributions. Data structures are recovered by forward diffusion and reverse denoising processes. For low-reliability observations, original observations, first-order differences, and second-order differences are adopted to construct local shape constraints. In addition, the residual-correction guidance mechanism is introduced to estimate and correct observation deviations to improve the data reconstruction accuracy. Finally, simulations are conducted with industrial load datasets in Liaoning Province. The results validated the effectiveness of the proposed method. The average accuracies of data correction and data reconstruction reach 97.21% and 97.17%, respectively. Moreover, reconstruction accuracies exceeding 90% are maintained in cross-scene conditions involving different seasons and regions.

Keywords: industrial load; symmetry power data reconstruction; anomaly detection; adversarial learning; denoising diffusion; cross-scene adaptation

Check for updates

Academic Editor: Jie Yang

Received: 25 June 2026
Revised: 28 July 2026
Accepted: 7 August 2026
Published: 12 August 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

Power data acquisition is affected by missing and unreliable observations caused by equipment failures, communication interruptions, and environmental interference. The data quality issues reduce data integrity and may conceal critical abnormalities, such as electricity theft and equipment faults. Therefore, anomaly perception and data reconstruction are of great importance for ensuring reliable data analysis and stable power system operation.

Considerable research has been conducted on anomaly detection and data-quality diagnosis in power systems. In Ref. [1], a symmetry-aware anomaly alarm optimization

framework integrating crowdsourced active learning, hierarchical attention-based deep reinforcement learning, and interpretable analysis is developed to improve the identification and interpretation of abnormal power-system alarms. In Ref. [2], multidimensional sensing information is systematically investigated for next-generation battery fault diagnosis, demonstrating the importance of integrating heterogeneous observations for identifying subtle fault characteristics. In Ref. [3], adaptive instantaneous-frequency synchrosqueezing analysis is combined with an enhanced AdaBoost classifier to identify complex power-quality disturbances under noisy conditions. In Ref. [4], the innovation-reduction properties of the iterated extended Kalman filter are employed to detect and distinguish different power-system anomalies. In Ref. [5], bilateral data fusion is further introduced for anomaly detection in cyber-physical power systems. These studies demonstrate that anomaly detection has gradually evolved from model-based state estimation and manually designed features toward multidimensional feature fusion and intelligent learning. However, most of these studies primarily focus on the detection, classification, or localization of abnormal states, while the subsequent reconstruction of abnormal observations and the recovery of continuous temporal information receive comparatively limited attention. In Ref. [6], a Long Short-Term Memory (LSTM)-based reconstruction model is established to recover missing monitoring data caused by sensor failures in offshore wind turbines. In Ref. [7], the operating characteristics of energy-storage batteries are analyzed to splice and reconstruct fragmented operational data under different charging and discharging conditions.

Existing studies have extensively investigated anomaly detection, data cleaning, and data reconstruction from different perspectives. Refs. [1,3–5] mainly focus on anomaly alarm optimization, power-quality disturbance identification, and abnormal-state detection in power systems, thereby improving the capability to identify and locate abnormal observations. However, relatively limited attention is paid to the subsequent reconstruction of the detected abnormal data and the recovery of temporal continuity before and after anomalous events. Refs. [2,8] investigate multidimensional sensor-based fault diagnosis and power-data cleaning and interpolation, respectively, contributing to the improvement of data quality; nevertheless, these methods mainly concentrate on fault identification or the correction of local missing and corrupted observations. Furthermore, Refs. [6,7,9] investigate data reconstruction for virtual power plant loads, offshore wind turbine sensor measurements, and energy-storage battery operating data, respectively, promoting the development of data recovery from conventional interpolation toward deep-learning-based time-series reconstruction. Nevertheless, most of these methods are developed for specific equipment or individual operating scenarios, and their adaptability to distribution differences among different industrial load types and operating conditions remains limited. Meanwhile, Refs. [10,11] demonstrate that multi-source feature fusion and operating-characteristic modeling can provide richer prior information for power-data analysis. However, such prior information has not yet been sufficiently integrated into generative reconstruction models to support cross-scene knowledge transfer and high-fidelity recovery. Therefore, existing studies still lack a unified reconstruction framework that can simultaneously exploit cross-scene prior information, distinguish data with different reliability levels, and reconstruct missing and low-reliability observations while preserving both local temporal characteristics and overall data-distribution consistency.

Therefore, current research still lacks a universal data reconstruction method that could simultaneously handle data loss and low-reliability anomalies, while also considering local variation characteristics and the overall consistency of the reconstructed data distribution. Especially in situations where industrial load operating modes differ significantly and the degree of anomalies changes dynamically, how to achieve cross-scenario prior transfer and high-fidelity data recovery remains a hot research topic. Therefore, an integrated

approach capable of simultaneously performing anomaly detection and data reconstruction is still lacking. The following issues still exist in current anomaly perception and data reconstruction studies:

- Industrial load data exhibit multi-scale fluctuations, high-dimensional correlations, and significant scene-dependent characteristics. Existing methods still have room for improvement in modeling complex load behaviors and enhancing cross-scene transferability.
- In practical industrial environments, missing data, abnormal disturbances, measurement noise, and local distortions often coexist. The information differences between completely missing data and low-reliability observations are rarely distinguished, which may limit reconstruction performance under complex abnormal conditions.
- Low-reliability observations may still contain useful information. However, the effective utilization of preserved information in low-reliability observations remains insufficient, thereby restricting reconstruction accuracy and capability in the algorithm.

The main innovations are elaborated as follows:

- A scene feature classification model based on a multilayer perceptron is proposed. The time and frequency domain features of industrial load are extracted, and cosine similarity is used to identify similar source scenes to enhance cross-scene transferability and generalization ability.
- Construct missing and anomaly masks based on the reliability observation matrix. Missing data is generated based on the conditional distribution of historical data. Combining a conditional denoising diffusion model and a generative adversarial learning mechanism, the missing data from the conditional distribution of historical load data is reconstructed, improving the accuracy of the reconstruction results and enhancing data perception and reconstruction capabilities in complex anomaly scenes.
- For low-reliability data, a diffusion-adversarial reconstruction method based on weak consistency constraints is proposed. Local morphological constraints are constructed using the original observations, first-order differences, and second-order differences. A residual correction guidance mechanism is introduced to learn and correct the bias of low-reliability observations; while preserving the original time-series information, the accuracy and robustness of data reconstruction are improved.

This paper is structured as follows: Section 2 constructs a cross-scene similarity model with conditional encoding. An improved generative adversarial network diffusion algorithm is built based on data discrimination in Section 3. Section 4 details the cross-scene denoising diffusion adversarial reconstruction method with low-reliability data; Section 5 constructs a diffusion adversarial correction method based on missing data. Section 6 constructs the Deviation Residual-Guided Weak Consistency Constraint Low- Reliability Correction Mod-el The proposed method is verified by simulations. The superiority of the method is proved by the comparisons in the Section 7.

2. Industrial Condition Distribution Modeling Considering Scene Similarity

This paper proposes an industrial load power conditional distribution model based on scene similarity. By jointly encoding temporal-domain and frequency-domain features with a multilayer perceptron, cosine distance is calculated to measure scene similarity. It enhances the ability to transfer scene feature knowledge. The term "symmetry power data" used in this study is reflected in the consistency of temporal variation trends, frequency-domain characteristics, and statistical distributions between reconstructed data and reliable original data. At the model-mechanism level, symmetry is represented by the opposite evolutionary processes formed by forward diffusion and reverse denoising, while dis-

tributional consistency between real and generated data is further constrained through adversarial learning.

2.1. Temporal-Domain Metrics

The temporal-domain set S^{time} is proposed to describe the industrial load electricity consumption behavior based on the temporal-domain characteristics $S^{\text{time}} = [\max p, \min p, V, \sigma, CF, HVG]$.

1. Daily Load Extremes. The degree of daily load variation is described by the maximum and minimum values of daily load power.

$$\begin{aligned} \max P &= \max(P_1, \ldots, P_t, \ldots, P_T) \\ \min P &= \min(P_1, P_t, \ldots, P_T) \end{aligned} \tag{1}$$

where P_t represents load consumption data. T is the number of industrial electricity consumption data during the scheduling cycle.

2. Difference in Peak and Valley. It is adopted to calculate the difference between positive and negative peaks in a load waveform.

$$V = \max P_t - \min P_t \tag{2}$$

where P_t represents load consumption data. $\max P_t$ is the maximum power. $\min P_t$ is the minimum power.

3. Daily standard deviation σ. It is adopted to characterize the dispersion of load data.

$$\sigma = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(P_t - \left(\frac{\max P_t - \min P_t}{T}\right)\right)^2} \tag{3}$$

4. Margin CF. It is used to quantify the distance between the system current load level and stability threshold.

$$CF = \max P_t / \left(\frac{1}{T}\sum_{t=1}^{T}\sqrt{|P_t|}\right)^2 \tag{4}$$

5. Hirsch index HVG. It is used to measure the long-term memory and trend persistence of time series.

$$\begin{aligned} HVG &= \frac{\log(R/S)}{\log T} \\ R &= \max P_t - \min P_t \\ S &= \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(P_t - \overline{P}_t\right)^2} \end{aligned} \tag{5}$$

where R represents the extreme value difference. S denotes the standardized measure of short-term volatility for the sequence. R/S signifies the relabeled extreme difference. T is the number of industrial electricity consumption data points during the scheduling cycle.

2.2. Frequency Domain Metrics

The frequency-domain set is built, $S^{\text{fre}} = [FC, MSF, \gamma_k, H, N]$. The metrics are described as follows:

1. Center of mass frequency FC. It is used to describe the center of the frequency distribution.

$$FC = \frac{\sum_m f_m \cdot \left| X[m] \right|^2}{\sum_m \left| X[m] \right|^2} \tag{6}$$

where $|X[m]|$ is the signal amplitude at frequency f_k. $\left| X[m] \right|^2$ is the energy of the signal at frequency f_m. f_m is the frequency index in the discrete Fourier transform.

2. Frequency variance MSF. It is used to show the degree of the frequency signal component around its mean value.

$$MSF = \frac{\sum_m (f_m - FC)^2 \left| X[m] \right|^2}{\sum_m \left| X[m] \right|^2} \tag{7}$$

where m is the frequency index in the discrete Fourier transform.

3. Spectral skewness γ_m. The shift in the "center of gravity" of signal energy across frequencies.

$$\gamma_m = \frac{\sum_m (f_m - MSF)^3 \left| X[m] \right|^2}{\sigma_f^3 \sum_m \left| X[m] \right|^2} \tag{8}$$

where MSF is the frequency variance. f_m is the frequency index in the discrete Fourier transform. m is the frequency index in the discrete Fourier transform.

4. Spectral Entropy H. It is used to measure the disorder and randomness of the spectral energy distribution.

$$\begin{aligned} H &= \sum_m \widetilde{P}(f_m) \log_2 \left(\widetilde{P}(f_m) \right) \\ \widetilde{P}(f_m) &= \frac{\left| X[m] \right|^2}{\sum_m \left| X[m] \right|^2} \end{aligned} \tag{9}$$

where H is the spectral entropy. m is the discrete frequency point index. $\widetilde{P}(f_m)$ is the normalized spectral energy ratio of the frequency component m.

5. Entropy Power N. It is used to measure the disorder of frequency components.

$$N = \frac{1}{2\pi e} e^{2H(P)} \tag{10}$$

where $H(P)$ is the spectral entropy corresponding to the normalized spectral energy distribution. $P = \widetilde{P}(f_m)$ is the normalized spectral energy percentage of the frequency component m.

In summary, the scene classification model could be expressed based on temporal-domain and frequency-domain features as follows:

$$S = [S^{\text{time}} S^{\text{fre}}] \tag{11}$$

2.3. Data Condition Distribution Model Based on Conditional Encoder

High-dimensional condition variables could lead to unstable model training and an exponential increase in parameter size. A multilayer perceptron [12–15] is adopted with fewer parameters, stable training and fast convergence in the paper to build an encoder model. After the temporal condition vector is output as a low-dimensional condition by the multilayer perceptron network, feature fusion is performed according to the weight ratio.

$$\begin{aligned} h^{\text{time}} &= \phi^{\text{time}}\left(S^{\text{time}}\right) \\ h^{\text{fre}} &= \phi^{\text{fre}}\left(S^{\text{fre}}\right) \\ \phi^{\text{time}}\left(S^{\text{time}}\right) &= \omega_2^{\text{time}}\text{ReLU}\left(\omega_1^{\text{time}}S^{\text{time}} + b_1^{\text{time}}\right) + b_2^{\text{time}} \\ \phi^{\text{fre}}\left(S^{\text{fre}}\right) &= \omega_2^{\text{fre}}\text{ReLU}\left(\omega_1^{\text{fre}}S^{\text{fre}} + b_1^{\text{fre}}\right) + b_2^{\text{fre}} \end{aligned} \tag{12}$$

where ϕ^{time} is the temporal-domain conditional encoder. ϕ^{fre} is the frequency-domain conditional encoder; α is the feature fusion weight ratio. ω_1^{time} and ω_1^{fre} are the weight matrices for the first layer. b_1^{time} and b_1^{fre} are the bias vectors for the first layer. ω_1^{time} and ω_1^{fre} is the weight matrix for the second layer. b_1^{time} and b_1^{fre} are the bias vectors for the second layer.

$$h^{\Sigma} = \alpha h^{\text{time}} + (1-\alpha)h^{\text{fre}} \tag{13}$$

After feature dimension reduction and fusion via the multi-layer perceptron algorithm, the output is h^{Σ}. Cosine similarity is employed to calculate distances between features for load scene. By adjusting the similarity parameters of scene classification, cross-scene anomaly detection and data reconstruction could be achieved.

$$SI^{(i,j)} = \frac{h^{\Sigma(i)}h^{\Sigma(j)}}{\left\|h^{\Sigma(i)}\right\|_2\left\|h^{\Sigma(j)}\right\|_2}$$
$$SI^{(i,j)}\begin{cases} 0 \le SI^{(i,j)} \le 0.5, \text{not similar} \\ 0.5 < SI^{(i,j)} \le 0.7, \text{ similar} \\ 0.7 < SI^{(i,j)} \le 1, \text{ verysimilar} \end{cases} \tag{14}$$

where scene similarity $SI^{(i,j)}$ is the similarity parameter concerning scene i and scene j. $\left\|h^{\Sigma(i)}\right\|_2$ and $\left\|h^{\Sigma(j)}\right\|_2$ represent the L2 norm of the two feature vectors, respectively. $\Sigma(i)$ and $\Sigma(j)$ represent the i and j load scenarios, respectively. i and j are the sample indices.

The similarity between different scenes is calculated to identify historical data that are most similar to the target scene. For each clustered scene, the fused feature $h^{\Sigma i}$ is employed to characterize the corresponding conditional distribution, which is further utilized as a prior condition for data reconstruction $f_\theta\left(p|h^{\Sigma i}\right)$. In this way, data reconstruction is guided by cross-scene conditional distributions. The architecture diagram of the model based on the conditional encoder is shown in Figure 1.

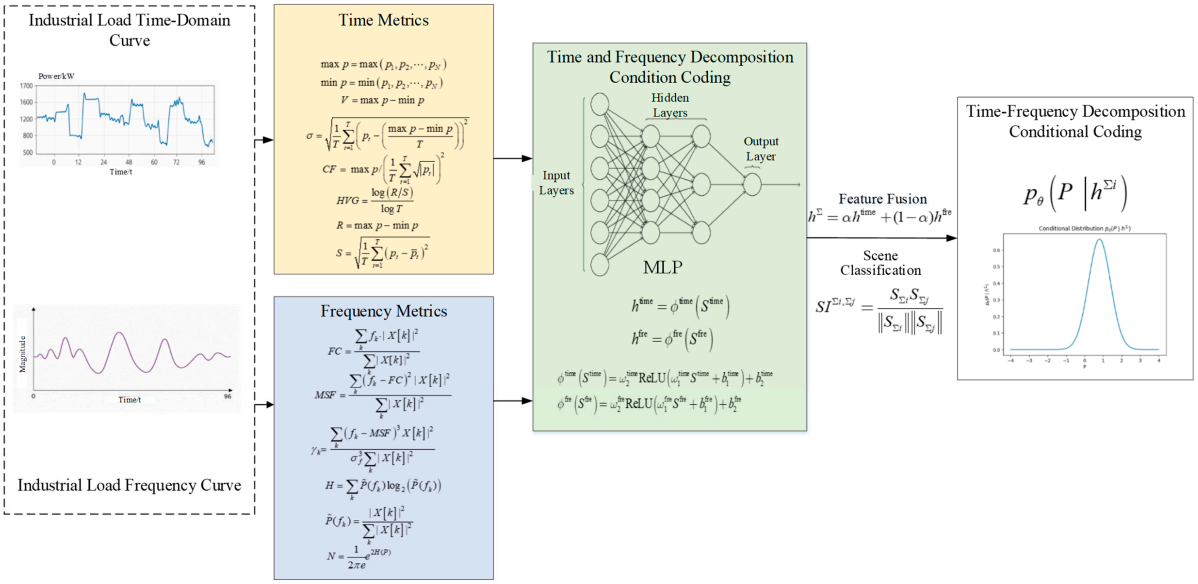


Figure 1. Scene data conditional distribution model based on a conditional encoder: temporal and frequency-domain features are jointly encoded by a multilayer perceptron. The scene affiliation of the data is identified by cosine similarity to improve the cross-scene generalization capability.

3. An Improved Generative Adversarial Network Denoising and Diffusion Algorithm Based on Reliability Mask

Based on generative adversarial networks (GANs), the method continuously reduces the distribution gap between the generated data and the real data during the process of missing data recovery. This method demonstrates strong capabilities in complex nonlinear data modeling and high-dimensional data generation. However, existing GAN-based data reconstruction methods still have certain limitations in the aforementioned scenarios, mainly reflected in three aspects: the ability to recover continuously missing data, the consideration of the information contained in low-reliability data, and the ability to transfer load behavior distribution and cross-scenario features in complex industrial scenarios.

An Improved Adversarial Learning Generative Algorithm Based on Denoising Diffusion

Traditional adversarial generative models [16] still have certain limitations in characterizing complex data distributions. This paper proposes a reliability-aware data reconstruction method for generative adversarial networks with an improved denoising diffusion model. The observed data is classified and modeled by data reliability analysis. For abnormal data, a noise-adding diffusion process is introduced to reduce the original observation noise and anomalous interference. A denoising generator is then used to invert and reconstruct the diffused data. During this process, an adversarial learning mechanism between the generator and the discriminator is employed to achieve high-quality reconstruction. The generation efficiency and distribution consistency are improved.

Firstly, a data reliability observation method is proposed that classifies raw data $P_t \in R^{n\times t}$ into missing data, low-reliability data, and high-reliability data. The reliability observation method consists of a missing mask A and an anomaly mask R.

The missing mask A is used to characterize data integrity.

$$\begin{aligned} A &= \begin{cases} 0\,, \text{Data is not collected} \\ 1\,, \text{Data is collected} \end{cases} \\ A &\in \{0,1\}^{n\times T} \end{aligned} \tag{15}$$

where A represents the missing mask, n denotes the sampling time span, and t indicates the scheduling cycle, $T = 96$. If data collection fails $A = 0$, otherwise, $A = 1$.

Anomaly mask R is used to indicate whether data contains outliers:

$$\begin{aligned} R &= R_1 \cdot R_2 \\ R_1 &= \exp\left(-\frac{|P_t - P_{t-1}|}{\sigma_\Delta}\right) \\ R_2 &= \exp\left(\frac{\max(0, |P_t - P_{0.5}| - \kappa)}{\sigma_D}\right) \\ R &\in R^{n\times T} \end{aligned} \tag{16}$$

$$\begin{aligned} \sigma_\Delta &= \frac{1}{N-2}\sum_{t=2}^{T}\left(\Delta P_t - \Delta\overline{P}_t\right)^2 \\ \Delta\widetilde{P}_t &= \frac{1}{T-1}\sum_{t=2}^{T}\Delta P_t \end{aligned} \tag{17}$$

where R represents the anomaly mask, $R \in [0,1]^{n\times T}$. R_1 represents the mutation anomaly judgment mask. R_2 represents the persistent anomaly judgment mask. P_t and P_{t-1} denote the power values at time points t and $t-1$, respectively. σ_Δ denotes the historical load variation standard deviation. $\Delta\overline{P}_t$ denotes the average historical variation; $P_{0.5}$ represents the median of the load sequence. κ is the tolerance value. σ_D is the permissible normal fluctuation threshold and serves as the scale parameter for distribution offset. Reliability Mask W is used to indicate the level of trustworthiness of the data.

$$W = A \cdot R \tag{18}$$

when an anomaly occurs but the value exists $A = 1,\ R \approx 0$. At this point, $W \approx 0$, the data is considered unreliable. $A = 1,\ R \approx 1$. At this point, $W \approx 1$, the data is reliable. If $W = 0$, data is currently missing.

4. The Reconstruction Training Model with Low-Reliability Data

For low-reliability data, this paper introduces different noises and employs data weakening to preserve references to original observations. For missing data, reconstruction is achieved through diffusion and denoising starting from unbiased noise.

The historical time-series structure condition $h^{\Sigma i}$ is constructed as C; generator $\mathcal{G}_\theta$ generates a distribution based on historical conditional generation.

$$\hat{P}_{0,x}^{\text{miss}} \sim f_\theta^{\Omega_0}\left(P_0^{\text{miss}} \middle| C\right) \tag{19}$$

where P_0^{miss} is the actual load data distribution corresponding to the region under study. $\hat{P}_{0,t}^{\text{miss}}$ represents the corrected or reconfigured load value output by the generator in the study area. $f_\theta^{\Omega_0}$ is the conditional diffusion generation distribution with parameter θ.

The intermediate noise power P^x is obtained at each step x of the diffusion process with abnormal power consumption data. The improved model guides the generator to produce missing data and reconstruct unreliable data via a reliability mask. The generator outputs samples generated in the original data space through reverse denoising.

$$\hat{P}_0 = \mathcal{G}_\theta(P_x, x) \tag{20}$$

where the generator outputs a generated sample $\hat{P}_0$ located in the original data space at diffusion step x.

The generator loss function based on conditional denoising diffusion is shown in the following equation within the "missing or low-confidence region", thus learning how to map the noisy intermediate state P_x back to the original data space. The denoising loss is $L_{\text{re}}(\mathcal{G}_\theta)$, and the generator adversarial loss is $L_{\mathcal{G}_\theta}$.

The final loss function of the generator is shown in the following equation:

$$L_{\Sigma\mathcal{G}} = L_{\text{re}}(\mathcal{G}_\theta) + \lambda_{\mathcal{G}} L_{\mathcal{G}} \tag{21}$$

where $\lambda_{\mathcal{G}}$ is the generator weight.

The discriminator determines the authenticity of data given known low-reliability and missing regions. The input is:

$$D\left(\hat{P}_0^{\Omega_0}, A, R\right) \tag{22}$$

The loss function for the improved method is L_D. The adversarial learning could be expressed based on denoising diffusion as:

$$\begin{aligned} &\min_{\mathcal{G}_\theta} \max_{D} L\ (\mathcal{G}_\theta, D) = \\ &L_{\Sigma\mathcal{G}} + L_D = L_{\text{re}}(\mathcal{G}_\theta) + \lambda_{\mathcal{G}_\theta} L_{\mathcal{G}_\theta} + \lambda_{\text{GAN}} L_{\text{D}} \end{aligned} \tag{23}$$

where L_{re} represents the denoising loss. λ_{GAN} denotes the adversarial loss weight. L_{GAN} signifies the adversarial loss.

The discriminator evaluates the authenticity of data generated by the generator. The loss functions are defined as follows concerning the generator and discriminator:

$$L_{\mathcal{G}} = -\mathbb{E}\left[A \cdot \log D\left(\hat{P}_0^{\text{miss}}, A, R\right)\right] \tag{24}$$

$$\begin{aligned} L_D = &-\mathbb{E}\left[A \cdot \log D\left(\hat{P}_0^{\text{miss}}, A, R_t\right)\right] \\ &-\mathbb{E}\left[(1-A) \cdot \log\left(1 - D\left(\hat{P}_0^{\text{miss}}, A, R\right)\right)\right] \end{aligned} \tag{25}$$

Building upon Section 1, this section categorizes data reconstruction scenes into data missing scenes and data low-reliability scenes. By default, data regions with higher reliability are retained as true values [16–18].

5. The Reconstruction Training Model with Missing Data

When the reliability observation matrix is $W = 0$, the data is considered missing. This paper proposes a prior distribution based on similar load patterns. Missing data is defined as conditionally generated with diffusion.

The historical time-series structure condition of $h^{\Sigma i}$ is constructed as follows:

$$C^{\Omega_0} = \left[P_{t-1}, P_{t-D}, P_{t-7D}, P_{t-Y}, h^{\Sigma i}\right] \tag{26}$$

where P_{t-1} represents the power at the previous time window period. P_{t-D} denotes the load electric consumption at this time during the whole preceding scheduling cycle. P_{t-7D} indicates the load power consumption at this time step one week prior. P_{t-Y} signifies the load power consumption at this time step one year prior. $h^{\Sigma i}$ represents the conditional distribution of the scene classified scene i.

For the set of missing positions $\Omega_0 = \{P|W = 0\}$, generator $\mathcal{G}_\theta$ generates a distribution based on historical conditional generation.

$$\hat{P}_{0,x}^{\text{miss},\Omega_0} \sim f_\theta^{\Omega_0}\left(P_0^{\text{miss},\Omega_0} \middle| C^{\Omega_0}\right) \tag{27}$$

where $P_0^{\text{miss},\Omega_0}$ is the actual load data distribution corresponding to the missing regions. $\hat{P}_{0,t}^{\text{miss},\Omega_0}$ represents the reconstructed load value output by the generator in the missing region Ω_0. $f_\theta^{\Omega_0}$ is the conditional diffusion generation distribution with parameter θ.

When data is missing, historical load patterns serve as the primary conditional information. Multi-cycle data already incorporates most of the information, and more robust constraints are provided in cases of complete data absence.

1. Forward Additive Noise

To learn the conditional generative distribution, missing power values are simulated in the training phase for the complete historical sequence, with only noise injected at the missing positions. During the training phase, the model is trained with actual values P_t. When $W = 0$, the intermediate noise power obtained in step x is rewritten as:

$$\begin{aligned} &P_x^{\text{miss},\Omega_0} = (1-A)P_0 + A \odot \left(\sqrt{\textstyle\prod_{x=1}^{X} \alpha_x} P_0 + \sqrt{1 - \textstyle\prod_{x=1}^{X} \alpha_x}\varepsilon\right) \\ &\varepsilon \sim N(0, I) \end{aligned} \tag{28}$$

where $P_x^{\text{miss},\Omega_0}$ represents the missing region state during the x diffusion step. P_0 is the original complete load data. $\odot$ is the Hadamard element-wise product. α_x is the noise weighting coefficient.

2. Conditional Denoising

The generator models the conditional denoising process with a noise prediction network.

$$\hat{P}_0^{\text{miss},\Omega_0} = \mathcal{G}_\theta\left(P_x^{\text{miss},\Omega_0}, x, C\right) \tag{29}$$

where $\mathcal{G}_\theta$ is the conditional diffusion generator. $\hat{P}_0^{\text{miss},\Omega_0}$ is the missing region reconstruction value predicted by the generator. $P_x^{\text{miss},\Omega_0}$ is the diffusion state at step x. x is the current diffusion step number.

3. Generator Training Loss Function

$$L_{\text{re}}^{\Omega_0}(\mathcal{G}_\theta) = E_{P_0,\varepsilon,x}\left[\left\|A \cdot (P_x^{\text{miss},\Omega_0} - \mathcal{G}_\theta\left(P_x^{\text{miss},\Omega_0}, x, C\right)\right\|_2^2\right] \tag{30}$$

where $L_{\text{re}}^{\Omega_0}(\mathcal{G}_\theta)$ represents the reconstruction loss of the missing region. $P_x^{\text{miss},\Omega_0}$ represents the true value of the missing region.

4. Refined Training Based on Adversarial Learning

For missing positions in the training samples, the trained diffusion model is employed to perform full reverse diffusion. For completely missing position sets Ω_0, data reconstruction is achieved by reverse denoising with the constraint of historical conditions C.

$$\begin{aligned} P_X^{\text{miss},\Omega_0} &= (1-A)P_0 + A \cdot \varepsilon, \varepsilon \sim N(0, I) \\ P_{x-1}^{\text{miss},\Omega_0} &= \mathcal{G}_\theta\left(P_x^{\text{miss},\Omega_0}, x, C\right), x = X, X-1, \ldots, 1 \end{aligned} \tag{31}$$

where X is the total diffusion steps. $P_X^{\text{miss},\Omega_0}$ is the initial state of the back diffusion. $P_{x-1}^{\text{miss},\Omega_0}$ is the result of the back diffusion at step $x-1$.

5. The restructured output values $\hat{P}^{\Omega_0}$ are as follows:

$$\hat{P}_x^{\Omega_0} = A \cdot P_0^{\text{miss},\Omega_0} + (1-A)P_x \tag{32}$$

where $\hat{P}_x^{\Omega_0}$ is the final reconstructed load sequence. $P_0^{\text{miss},\Omega_0}$ is the missing region reconstruction value output by the generator. P_x is the original observed load data.

6. Discriminator $D(\cdot)$.

$D(\cdot)$ is built to determine whether the input originates from the true data distribution.

7. The loss functions

The discriminator evaluates the authenticity of data generated by the generator. The loss functions are defined as follows concerning the generator and discriminator:

$$L_{\mathcal{G}} = -\mathbb{E}\left[A \cdot \log D\left(\hat{P}_0^{\text{miss},\Omega_0}, A, R\right)\right] \tag{33}$$

$$\begin{aligned} L_D = &-\mathbb{E}\left[A \cdot \log D\left(\hat{P}_0^{\text{miss},\Omega_0}, A, R_t\right)\right] \\ &-\mathbb{E}\left[(1-A) \cdot \log\left(1 - D\left(\hat{P}_0^{\text{miss},\Omega_0}, A, R\right)\right)\right] \end{aligned} \tag{34}$$

8. Adversarial learning could be formulated based on denoising diffusion:

$$\begin{aligned} \min_{\mathcal{G}_\theta} \max_D L(\mathcal{G}, D) &= L_{\Sigma\mathcal{G}} + L_D \\ &= L_{\text{re}}(\mathcal{G}_\theta) + \lambda_{\mathcal{G}} L_{\mathcal{G}} + \lambda_{\text{GAN}} L_{\text{D}} \end{aligned} \tag{35}$$

where L_{re} represents the denoising loss. λ_{GAN} denotes the adversarial loss weight. L_{GAN} signifies the adversarial loss.

In summary, missing data reconstruction is achieved by a two-stage algorithm. The diffusion algorithm is employed to build a conditional generative reconstruction model, while adversarial learning is proposed to improve the quality of reconstructed data. Pseudocode is presented in Table 1.

Table 1. Data imputation reconstruction training pseudocode.

Pseudocode
• Identify missing sampling values $\Omega_0 = \{ P \mid W = 0 \}$
• Construct the historical conditions C
• Apply forward noise only to missing positions $P_x^{\text{miss},\Omega_0}$
• Perform noise prediction $\hat{P}_0^{\text{miss},\Omega_0} \leftarrow \mathcal{G}_\theta\left(P_x^{\text{miss},\Omega_0}, x, C\right)$
• Calculate diffusion loss only for missing positions $L_{\text{re}}^{\Omega_0}(\mathcal{G}_\theta)$
• Update θ
• Output reconstructed values $\hat{P}_x$

6. Deviation Residual-Guided Weak Consistency Constraint Low-Reliability Correction Model

For the reconstruction of data in low-reliability areas, directly ignoring the original data may disrupt the consistency of the original temporal sequence. In this section, the original data is used as a weak constraint, and a low-reliability data set is first built, $W \approx 0$.

$$\Omega_u = \{ P \mid 0 < W < \delta \} \tag{36}$$

Firstly, the historical time series structure condition C injects the original observation P^{Ω_u} and its difference characteristics as weak conditional features into the condition space. To avoid local temporal morphological distortion of low-reliability data during the conditional diffusion generation process, first-order and second-order difference equations are introduced to construct local morphological constraints based on the original observations and historical time-series conditions. The temporal consistency is improved by the low-reliability data reconstruction results. The difference is used to characterize the direction and speed of change of the first order, expressed as follows:

$$\Delta P_t^{\text{obs}} = P_t^{\text{obs}} - P_{t-1}^{\text{obs}} \tag{37}$$

where P_t^{obs} is the power at the moment t. P_{t-1}^{obs} is the power at the previous moment.

The second-order difference is used to ensure the curvature and peak-valley periods of the data reconstruction.

$$\nabla^2 P_t^{\text{obs}} = P_t^{\text{obs}} - 2P_{t-1}^{\text{obs}} + P_{t-2}^{\text{obs}} \tag{38}$$

where P_{t-2}^{obs} is the power at the previous moment of $t-1$.

Therefore, a conditional distribution C^{Ω_u} is proposed based on weak consistency, which follows the historical condition:

$$C^{\Omega_u} = \left[P_{t-1}, P_{t-D}, P_{t-7D}, P_{t-Y}, P^{\Omega_u}, h^{\Sigma i}, \Delta P_t^{\text{obs}}, \nabla^2 P_t^{\text{obs}} \right] \tag{39}$$

where ΔP_t^{obs} and $\nabla^2 P_t^{\text{obs}}$ are first-order difference and second-order difference. P_{t-1} represents the power at the previous time window period. P_{t-D} denotes the load electric consumption at this time during the whole preceding scheduling cycle. P_{t-7D} indicates the load power consumption at this time step one week prior. P_{t-Y} signifies the load power consumption at this time step one year prior. $h^{\Sigma i}$ represents the conditional distribution of the scene classified scene i.

Furthermore, low-reliability data contains valuable information that could be used to guide model reconstruction. This paper proposes a low-reliability data diffusion reconstruction model guided by residual correction.

$$e = P_t^{\text{fact}} - P_t^{\text{obs}} \tag{40}$$

where e represents the data error. P_t^{fact} represents the actual data value. P_t^{obs} represents the data collection value.

$\hat{P}_0^{\text{obs},\Omega_u}$ is not reconstructed directly from the generator but first learns the bias residuals from the low-reliability data. The residuals could be expressed as follows:

$$\hat{e} = E_\phi\left(P^{\text{obs}}, A, W, \Delta P_t^{\text{obs}}, \nabla^2 P_t^{\text{obs}}\right) \tag{41}$$

where E_ϕ is the residual estimation network. A is the reliability mask. ΔP_t^{obs} is the power difference, and $\nabla^2 P_t^{\text{obs}}$ is the second-order differences of the low-reliability data.

For the set $\Omega_u = \{P|0 < W < \delta\}$ that needs to be corrected, the generator $\mathcal{G}_\theta$ generates a distribution that follows the historical condition.

$$\hat{P}_t^{0,\Omega_u} \sim f_\theta^{\Omega_u}\left(P_t^{\Omega_u}\middle|C^{\Omega_u}\right) \tag{42}$$

Unlike data loss, low-reliability data contains valid information that could guide model reconstruction by effective utilization of information. This paper proposes a low-reliability data diffusion reconstruction model guided by residual correction.

Unlike missing data, it first learns the bias residuals in low-reliability data, which could be expressed by the following equation:

$$\hat{e} = E_\phi\left(P, A, W, \Delta P_t^{\text{obs}}, \nabla^2 P_t^{\text{obs}}\right) \tag{43}$$

where E_ϕ is the residual estimation network. A is the reliability mask.

Then, the low-reliability observations are corrected.

$$\hat{P}_t^{\text{obs}} = P_t^{\text{obs}} - \hat{e} \tag{44}$$

The loss function and refined training based on adversarial learning are different in the low-reliability correction model. The Equations are shown as follows:

1. The Loss Function:

$$L_{\text{re}}^{\Omega_u}(\mathcal{G}_\theta) = E_{P_0,\varepsilon,x}\left[\left\|A\cdot\left(P_x^{\text{obs},\Omega_u} - \mathcal{G}_\theta\left(P_x^{\text{obs},\Omega_u}, x, C\right)\right\|_2^2\right] + \|(1-A)(\hat{e}-e)\|_2^2 \tag{45}$$

2. Refined Training Based on Adversarial Learning.

For the low-reliability locations in the training samples, a complete reverse diffusion is performed with the trained diffusion model. For the set of Ω_u, reconstruction is performed with a conditional generation distribution and reverse denoising with the constraint of historical condition C:

$$\begin{aligned} P_X^{\text{obs},\Omega_u} &= (1-A)P_0 + A\cdot\varepsilon, \varepsilon \sim N(0,I) \\ P_{x-1}^{\text{obs},\Omega_u} &= \mathcal{G}_\theta\left(P_x^{\text{obs},\Omega_u}, x, C\right), x = X, X-1, \ldots, 1 \end{aligned} \tag{46}$$

The data reconstruction flowchart is shown in Figure 2.

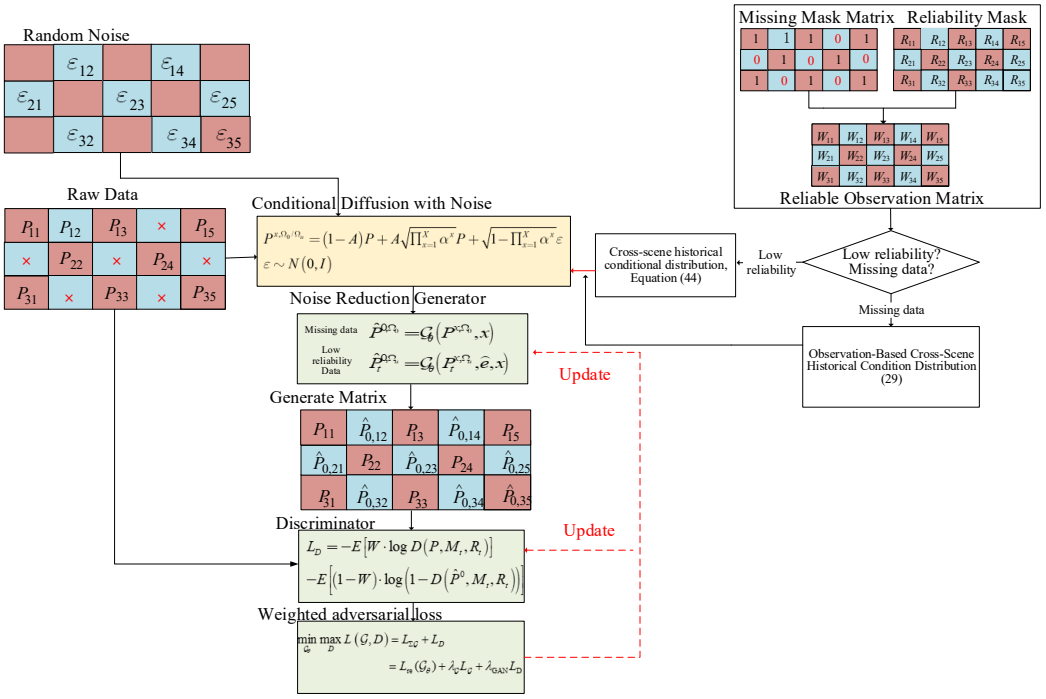


Figure 2. Data reconstruction algorithm flowchart: historical conditional distributions are incorporated to collaboratively regulate the diffusion process, denoising inversion, and optimization objectives. Anomaly perception and reconstruction performance in complex industrial scenes are improved.

A reliability-aware reconstruction framework is proposed to correct missing data and low-reliability data. The framework of the paper is shown in Figure 3.

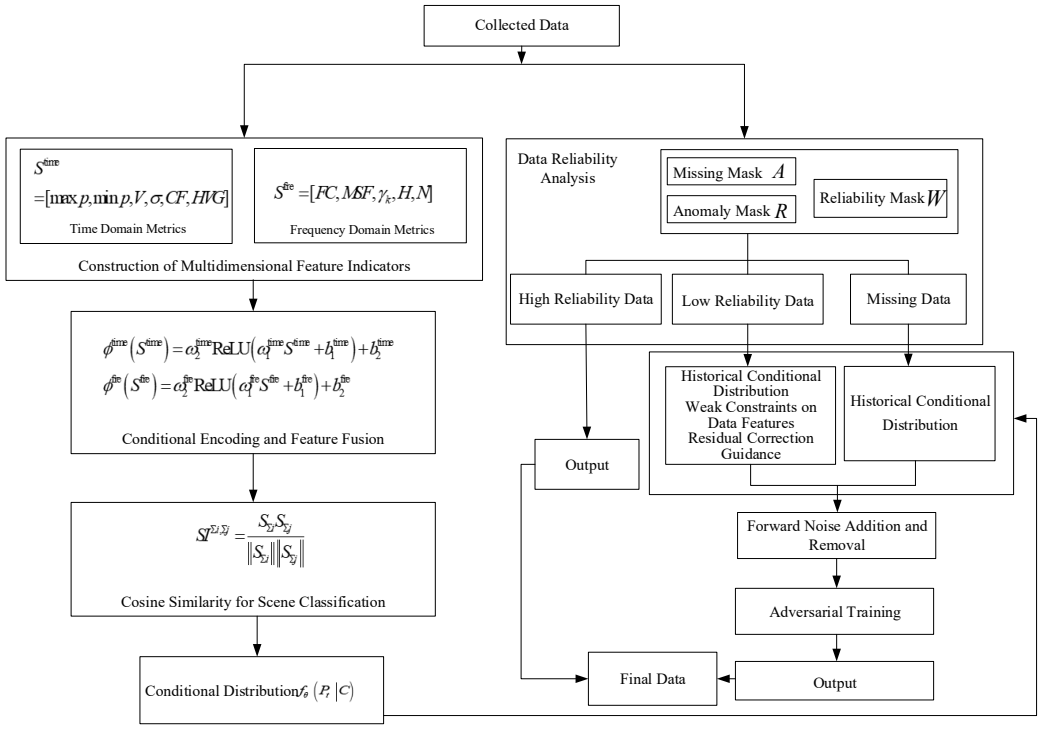


Figure 3. Cross-scene industrial load electricity anomaly detection method: logic diagram of algorithm for correcting and reconstructing low-reliability and missing data.

7. Discussion

The data comes from actual industrial users in Liaoning Province, including load 1 (furniture manufacturing plant), load 2 (feed processing plant), load 3 (rubber manufacturing plant), and load 4 (cotton textile processing plant), with rated capacities of 900 kVA, 1000 kVA, 2200 kVA, and 5500 kVA, respectively. The selected loads cover different manufacturing industries and capacity levels, with rated capacities ranging from 900 to 5500 kVA, reflecting the differences in electricity consumption among industrial users of different sizes. Furthermore, different industries exhibit significant differences in production processes, equipment operation modes, and production shifts, resulting in load curves displaying different characteristics in terms of base load level, peak–valley features, fluctuation degree, and periodicity.

The simulation platform used in this paper is configured with an Intel Core i7 processor with a clock speed of 3.60 GHz, 32 GB of memory, and Windows 11 operating system. The model building and data processing were completed in the MATLAB R2023b environment.

Therefore, the selected data can provide a representative test scenario for verifying the model's ability to reconstruct industrial electricity consumption data under different capacity levels and load patterns.

When data gaps occur in these industries, the accuracy of load-side electricity consumption data is improved through reconstructive simulation analysis. Simulation comparisons are conducted against seven other methods.

Figure 4 shows the typical daily load curves for loads 1 to 4.

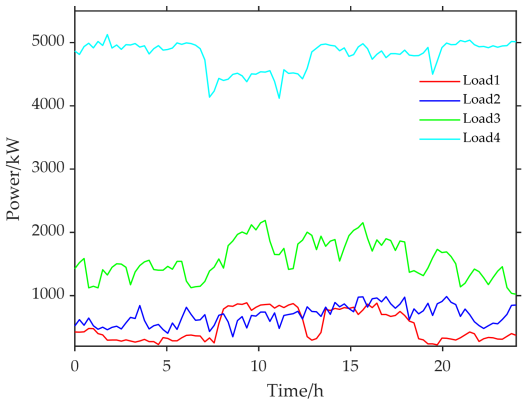


Figure 4. Typical daily load curves of load 1-4. Load 1 (furniture manufacturing plant), load 2 (feed processing plant), load 3 (rubber manufacturing plant), and load 4 (cotton textile processing plant), with rated capacities of 900 kVA, 1000 kVA, 2200 kVA, and 5500 kVA, respectively.

The load dataset is collected from an industrial park in Liaoning Province, covering the period from 2023 to 2024 with a sampling interval of 15 min. In total, 70% of the data are used for training, 10% are used for validation, and the remaining 20% are tested. The model is trained for 200 epochs, with the batch size set to 64. The learning rate and momentum coefficient are set to 0.0002 and 0.5, respectively. Detailed parameter settings of the proposed method are provided in Supplementary Materials Table S1.

7.1. Accuracy Analysis of Missing Mask and Reliability Mask Judgment

To evaluate the capability of the proposed missing-data mask in accurately identifying missing locations, random missingness, consecutive missingness, and block missingness are artificially introduced into the datasets of loads 1–4, with missing ratios of 10%, 20%, 30%, 40%, and 50%, respectively. Algorithm metrics under different data missing rates from 10% to 50% are shown in Supplementary Materials Table S2.

Precision measures the proportion of correctly identified unreliable observations out of all observations classified as abnormal, while *Recall* characterizes the model's ability to detect actual abnormal observations. The F1 score provides a balanced evaluation of precision and recall, particularly useful when abnormal samples are unevenly distributed.

Precision reflects how much of the anomaly identification results given by the model are actually abnormal data.

$$Precision = \frac{TP}{TP + FP} \tag{47}$$

where TP represents the number of correctly identified anomalous or unreliable data, and FP represents the number of data incorrectly identified as anomalous.

$$Recall = \frac{TP}{TP + FN} \tag{48}$$

where FN represents the number of actual anomalies that were not identified.

F1 considers both *Precision* and *Recall*, avoiding the impact of a high proportion of normal samples when using only accuracy.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{49}$$

To simulate low-reliability data caused by sensor drift, communication interference, measurement noise, and equipment operating abnormalities during industrial load acquisition, a certain proportion of data points are randomly selected from the original complete load sequence and designated as abnormal regions. For the selected locations, random perturbations are injected into the observations while preserving the original load trends and temporal structures, thereby generating abnormal observations as follows:

$$\begin{aligned} &P_t = P_t^1 + P_t^2 + P_t^3 + P_t^4 + P_t^5 + P_t^6 \\ &\left(P_t^1, P_t^2, P_t^3, P_t^4, P_t^5, P_t^6\right) \in \Omega_u \end{aligned} \tag{50}$$

1. Gaussian Noise

$$\begin{aligned} P_t^1 &= P_t^{\text{fact}} + \varepsilon_t^1 \\ \varepsilon_t^1 &\sim N\left(0, \sigma^2\right) \end{aligned} \tag{51}$$

where P_t^{fact} is the true load value. $P_t^{\Omega_u}$ is the low-reliability observation value. ε_t is a random disturbance that follows a Gaussian distribution.

2. Progressive Drift

$$\begin{aligned} P_t^2 &= P_t^{\text{Fact}} + \varepsilon_t^2 \\ \varepsilon_t^2 &= \alpha^2 (t - t_0) \end{aligned} \tag{52}$$

where α^2 is the drift rate.

3. Sensor Stuck

The power maintains a fixed value for a continuous period of time.

$$\begin{aligned} &P_t^3 = P_{t0}^{\text{fact}} \\ &t_0 \le t \le t_0 + L \end{aligned} \tag{53}$$

4. Duplicate Values

It is used to simulate data duplication caused by communication caching or repeated uploads.

$$P_{t:t+L}^4 = P_{t-L:t}^4 \tag{54}$$

5. Time Misalignment

$$P_t^5 = P_{t-\tau}^5 \tag{55}$$

where τ is the time offset.

6. Pulse Contamination

$$P_t^6 = P_t^{\text{Fact}} + \delta_t A \tag{56}$$

where $\delta_t \in (0, 1)$; only at random moments do large-amplitude instantaneous disturbances occur.

Unlike missing data, low-reliability observations still preserve partial authentic load information. Although deviations exist, the information is not completely lost, thereby reflecting practical industrial scenes in which the data remain available but exhibit low reliability.

As shown in Table 2, the identification accuracy of missing regions remains above 97.17%, while that of low-reliability regions remains above 97.21% under 50% anomaly ratios. Although a slight decrease in accuracy is observed as the anomaly ratio increases, the overall performance is consistently maintained at a high level. These results demonstrate that the constructed missing mask and reliability mask are capable of accurately locating missing and low-reliability regions, thereby providing reliable prior location information for subsequent differentiated data reconstruction.

Table 2. Reliability mask and missing mask accuracy table.

Noise Type	Accuracy %	Abnormal Ratio of 10%	Abnormal Ratio of 20%	Abnormal Ratio of 30%	Abnormal Ratio of 40%	Abnormal Ratio of 50%
Gaussian noise	Missing Accuracy	99.98	98.98	98.97	98.94	97.17
	Low-Reliability Accuracy	99.99	98.99	98.98	98.95	97.21
Progressive drift	Missing Accuracy	99.94	99.91	99.9	99.96	99.97
	Low-Reliability Accuracy	99.98	99.9	99.99	99.92	99.94
Sensor stuck	Missing Accuracy	99.99	99.99	99.91	99.97	99.88
	Low-Reliability Accuracy	99.99	99.91	99.91	99.91	99.91
Duplicate values	Missing Accuracy	99.95	99.98	99.94	99.93	99.92
	Low-Reliability Accuracy	99.92	99.9	99.97	99.94	99.88
Time misalignment	Missing Accuracy	99.91	99.95	99.9	99.96	99.89
	Low-Reliability Accuracy	99.99	99.98	99.95	99.97	99.93
Pulse contamination	Missing Accuracy	99.93	99.93	99.93	99.96	99.96
	Low-Reliability Accuracy	99.99	99.9	99.94	99.98	99.97

A total of 550 missing observations and 500 low-reliability observations were included in the task-specific evaluation sets. Among them, 532 missing observations and 491 low-reliability observations were correctly identified, respectively, as shown in Table 3.

Table 3. Evaluation indicators of low-reliability and missing data.

Type	TP	FP	FN	TN	Precision	Recall	F1	Accuracy
Reliability mask	491	15	9	1815	0.970	0.982	0.976	98.97%
Missing mask	532	7	18	1823	0.987	0.967	0.977	98.95%

7.2. Analysis of Low-Reliability Threshold Setting for Load Data

This paper determines the optimal decision threshold by setting different values of δ. The values of δ are set to 0.1, 0.3, 0.5, 0.7, and 0.9, as shown in Figure 5.

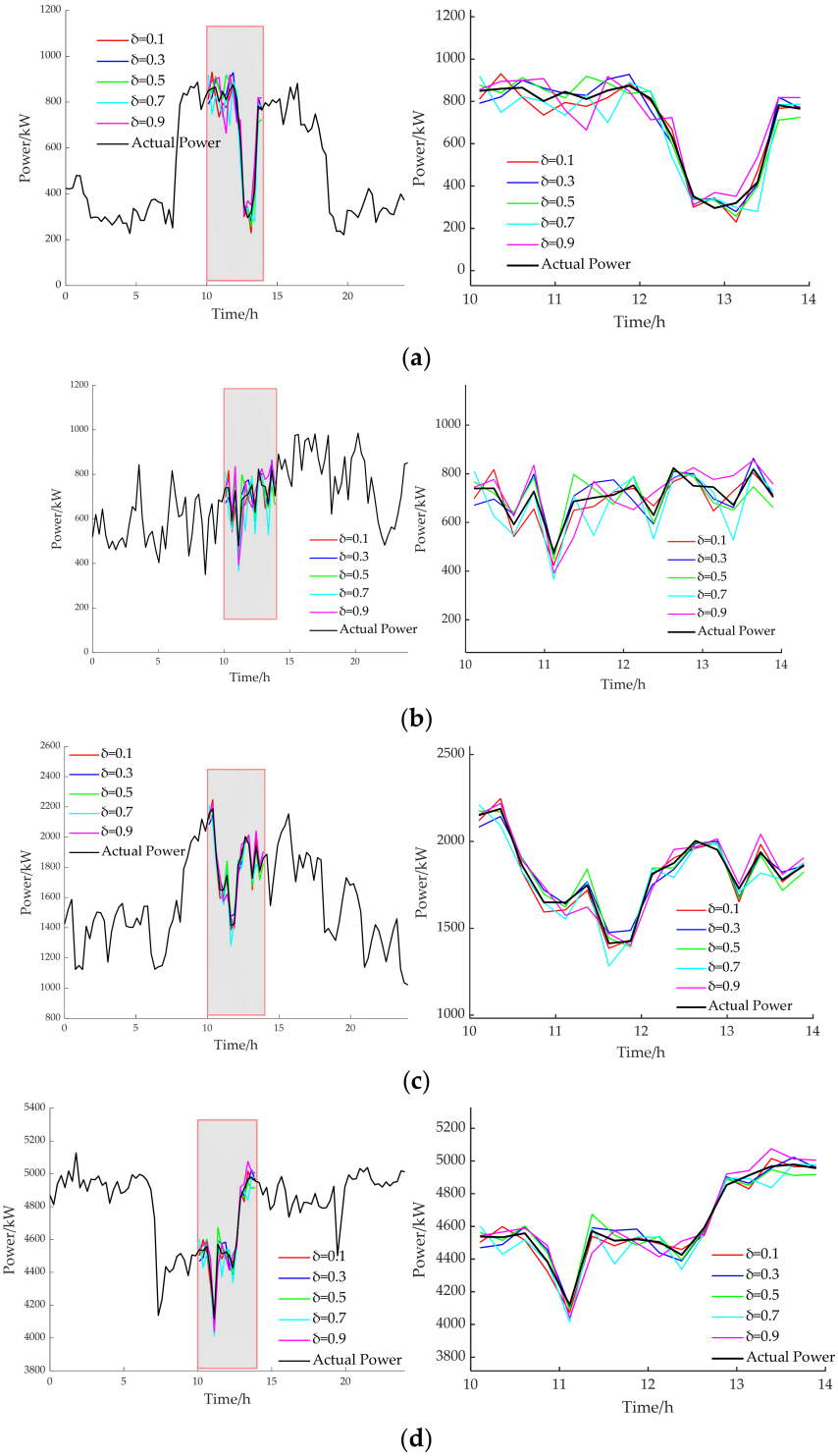


Figure 5. Reconfiguration results for different thresholds across four loads: (**a**) reconstruction results for different thresholds in furniture manufacturing plants; (**b**) reconstruction results for different thresholds in feed processing plants; (**c**) reconstruction results for different thresholds in the rubber manufacturing plant; (**d**) reconstruction results for different thresholds in cotton textile processing plants. As the threshold rises, the quality of data reconstruction first improves and then deteriorates. Increasing the low-reliability threshold, data reconstruction is shifted from conservative anomaly correction to aggressive generative smoothing.

As shown in Figure 5, as the reliability threshold is progressively raised, the load curve is progressively smoother. When $\delta = 0.1$, only excessively outlying anomalous data points are identified and reconstructed; not all outliers are detected. As the threshold increases, sudden changes in actual electricity consumption are classified as anomalies. Therefore, as the threshold rises, the quality of data reconstruction first improves and then deteriorates. Increasing the low-reliability threshold δ, data reconstruction is shifted from conservative anomaly correction to aggressive generative smoothing. Experiments demonstrate that, when $\delta = 0.3$, an optimal balance between fidelity and robustness is achieved.

As shown in Table 4, the reconstruction accuracy analysis result is shown with varying low-reliability threshold settings. Both Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics describe the trend when the reliability threshold rises. It initially decreases and then increases.

Table 4. Accuracy analysis of data reconstruction at low-reliability thresholds.

Reliability Threshold	Indicator	Load 1	Load 2	Load 3	Load 4
0.1	MAE	0.055	0.059	0.042	0.067
	RMSE	0.189	0.268	0.148	0.185
	R	0.975	0.968	0.987	0.979
	KS	0.179	0.186	0.171	0.174
0.3	MAE	0.046	0.054	0.026	0.033
	RMSE	0.16	0.199	0.136	0.142
	R	0.9875	0.9836	0.992	0.987
	KS	0.082	0.088	0.073	0.077
0.5	MAE	0.049	0.062	0.028	0.039
	RMSE	0.227	0.33	0.166	0.173
	R	0.973	0.968	0.986	0.981
	KS	0.133	0.139	0.119	0.122
0.7	MAE	0.081	0.091	0.059	0.061
	RMSE	0.284	0.47	0.187	0.235
	R	0.961	0.959	0.976	0.972
	KS	0.133	0.136	0.1272	0.132
0.9	MAE	0.119	0.137	0.105	0.107
	RMSE	0.407	0.42	0.331	0.355
	R	0.933	0.921	0.945	0.941
	KS	0.136	0.138	0.131	0.135

When δ is small, extreme residual anomalies dominate the squared error, resulting in a large RMSE. As δ increases, anomalies are effectively corrected, leading to a significant reduction in RMSE. However, an excessively large δ would lead to excessive smoothing of the reconstructed values, distorting the true peaks. Systematic errors are introduced, causing the RMSE to rise again. Compared to the MAE, the RMSE is more sensitive to residual anomalies and excessive smoothing effects. As the low-reliability threshold increases, the correlation coefficient R first increases and then decreases. This occurs because a larger δ causes indiscriminate smoothing of data fluctuations. The authenticity of load samples is diminished. Similarly, the load data distribution is compressed, thereby increasing the Kolmogorov–Smirnov distance (KS distance). A moderate value of δ could effectively correct anomalies while preserving normal measurement data. Overall, it achieves the optimal balance between accuracy, temporal consistency, and distribution fidelity when $\delta = 0.3$.

7.3. Accuracy Analysis of Load Data Reconstruction with Missing Values

Take the load curve reconstructed with 50% data missing on a certain day as an example. The reconstructed curve is shown in Figure 6. With the condition of 50% data missing, the reconstruction accuracy rates of the four types of loads are 96.65%, 95.54%, 98.63%, and 97.87%, respectively. The average reconstruction accuracy is 97.17%. Within the time period from 10:00 to 22:00, the data is missing, accounting for 50% of the total scheduling cycle data. Based on the stability of electricity consumption, the reconstruction accuracy of load 3 and 4 is generally higher than that of load 1 and 2.

The accuracy of load reconstruction is evaluated with four metrics: MAE, RMSE, Pearson correlation coefficient R, and KS distance.

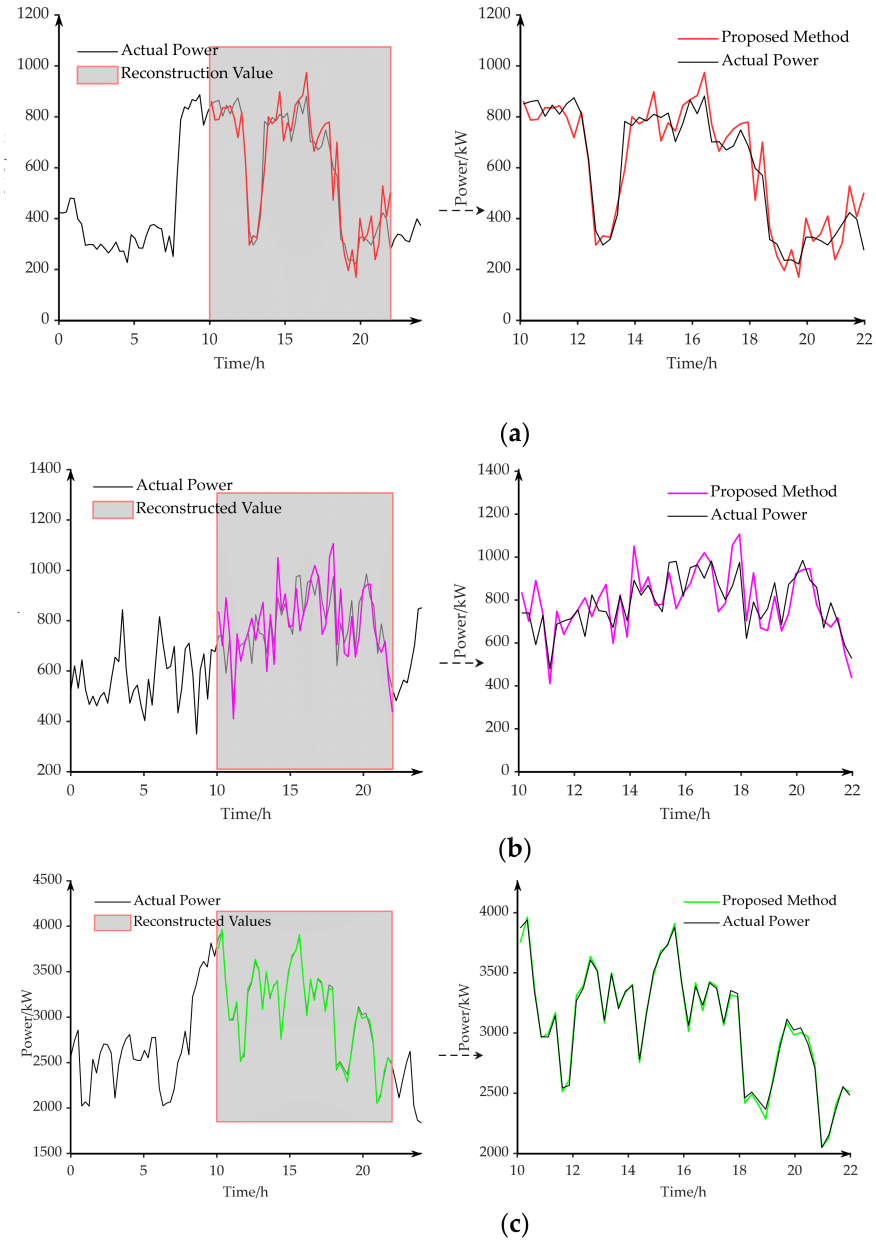
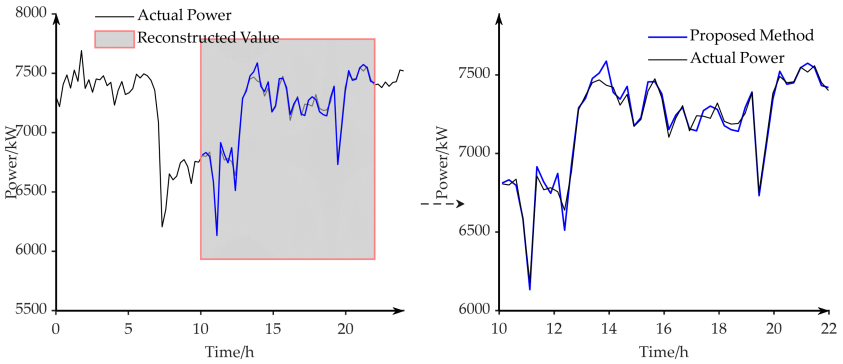


Figure 6. *Cont.*



(**d**)

Figure 6. Example of reconstructing missing data for 50% of four load categories on a given day: (**a**) reconstructed results for 50% missing data in furniture manufacturing plants; (**b**) reconstructed results for 50% missing data with feed processing plants; (**c**) reconstructed results for 50% missing data with rubber manufacturing plants; (**d**) reconstructed results for 50% missing data with cotton textile processing plants. With the condition of 50% data missing, the reconstruction accuracy rates of the four types of loads are 96.65%, 95.54%, 98.63%, and 97.87%, respectively.

Reconstruction Accuracy: The accuracy formula could be defined as:

$$Acc^{\Omega_0} = \left(1 - \frac{\sum_{n=1}^{N}\left|P_t^{\text{fact}} - \hat{P}_x^{\Omega_0}\right|}{\sum_{n=1}^{N}\left|P_t^{\text{fact}}\right|}\right) \times 100\% \tag{57}$$

$$Acc^{\Omega_u} = \left(1 - \frac{\sum_{n=1}^{N}\left|P_t^{\text{fact}} - \hat{P}_x^{\Omega_u}\right|}{\sum_{n=1}^{N}\left|P_t^{\text{fact}}\right|}\right) \times 100\% \tag{58}$$

where the true value is P_t^{fact} and corrected value is $\hat{P}_x$.

The MAE, RMSE, correlation coefficient, and KS are as follows:

MAE directly reflects the mean absolute error between the reconstructed results and the actual data; it could directly measure how much the model deviates from the true value on average for each sampling point.

The MAE is expressed as follows:

$$MAE = \frac{1}{N}\sum_{n=1}^{N}\left|P_t^{\text{fact}} - \hat{P}_x^{\Omega_0}\right| \tag{59}$$

RMSE could be used to check whether a model has severe local reconstruction distortion.

$$RMSE = \frac{1}{N}\sum_{n=1}^{N}\sqrt{\left(P_t^{\text{fact}} - \hat{P}_x^{\Omega_0}\right)^2} \tag{60}$$

MAE and RMSE primarily evaluate the magnitude of numerical errors, but they could not fully determine whether the reconstructed data retain the actual load variation patterns. The correlation coefficient is mainly used to evaluate the consistency between the reconstructed and actual sequences in terms of time-series trends and fluctuation characteristics. It could determine whether the model accurately restores the peak–valley variation patterns of industrial load. It is expressed as follows:

$$
\begin{aligned}
R &= \frac{\sum_{n=1}^{N}\left(P_t^{\text{fact}}-\overline{P}_t^{\text{fact}}\right)\left(\hat{P}_x-\overline{\hat{P}}_x\right)}{\sqrt{\sum_{n=1}^{N}\left(P_t^{\text{fact}}-\overline{P}_t^{\text{fact}}\right)^2}\sqrt{\sum_{n=1}^{N}\left(\hat{P}_x-\overline{\hat{P}}_x\right)^2}} \\
\overline{P}_t^{\text{fact}} &= \frac{1}{N}\sum_{n=1}^{N}\left(P_t^{\text{fact}}-\overline{P}_t^{\text{fact}}\right) \\
\overline{\hat{P}}_x &= \frac{1}{N}\sum_{n=1}^{N}\left(\hat{P}_x-\overline{\hat{P}}_x\right)
\end{aligned}
\tag{61}
$$

The KS distance is used to measure the degree of restoration of a statistical distribution. For real data and reconstructed data, empirical cumulative distribution functions and are constructed, respectively. The KS distance is defined as follows:

$$
KS = \sup\left|F^{\text{FACT}}\left(P_t^{\text{fact}}\right)-F\left(\hat{P}_x\right)\right| \tag{62}
$$

where sup represents the maximum absolute distance between two cumulative distribution functions.

Load 3 and 4 exhibit higher overall reconstruction accuracy than load 1 and 2 due to their stable electricity consumption patterns. Load 3 shows the best reconstruction performance, while load 2 shows relatively poorer results. As the missing data ratio increases, reconstruction accuracy decreases. However, it still performs well at a 50% missing data ratio, demonstrating engineering feasibility, as shown in Table 5.

Table 5. Analysis of reconstruction accuracy with 50% data missing ratios.

Missing Ratio	Indictor	Load 1	Load 2	Load 3	Load 4
50%	MAE	0.0435	0.0498	0.0417	0.0434
	RMSE	0.0507	0.0538	0.0431	0.0447
	R	0.932	0.925	0.951	0.942
	KS	0.163	0.176	0.129	0.173

In Figure 7, it shows the reconstruction results for four loads within the low data reliability interval. It could be seen that the reliability separately sets persistent data anomalies and sudden anomalies in the proposed method. The time intervals 12:30–14:30 are used for validation. The proposed method exhibits a certain degree of reconfiguration capability under low-reliability conditions, with reconfiguration accuracies of 96.68%, 95.81%, 98.59%, and 97.77%, respectively, and an average reconfiguration accuracy of 97.21%.

The reconstruction accuracy analysis metrics are presented in Table 6, indicating that the method possesses engineering implementation capability. Furthermore, the data reconstruction is achieved with low-reliability conditions during computations under different load characteristics.

Table 6. Analysis of reconstruction accuracy with 50% data low-reliability ratios.

Indicator	Load 1	Load 2	Load 3	Load 4
MAE	0.0378	0.043	0.036	0.037
RMSE	0.157	0.169	0.107	0.155
R	0.978	0.969	0.991	0.987
KS	0.1268	0.1267	0.1275	0.1272

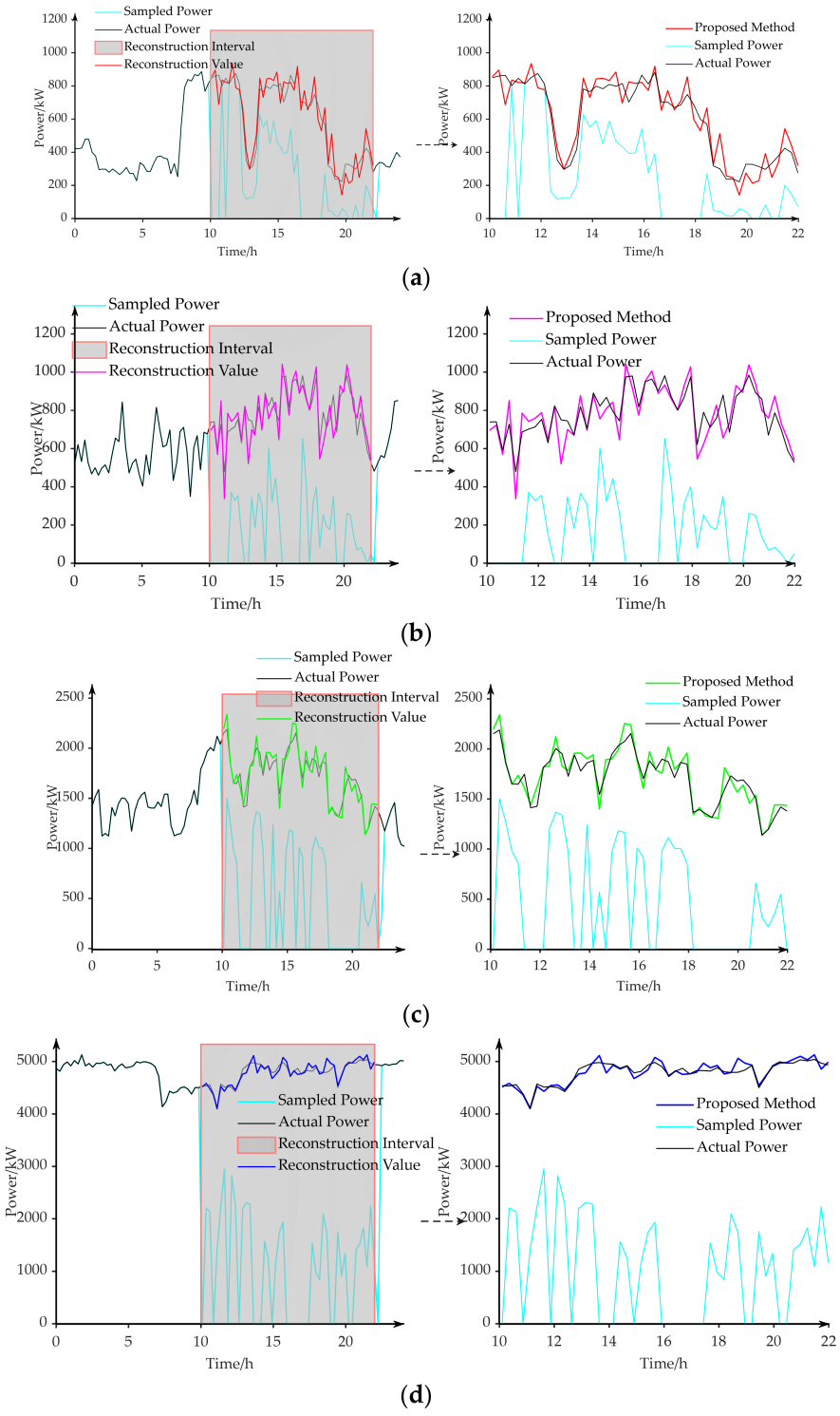


Figure 7. Example of reconstructing low-reliability data of four load categories on a given day: (**a**) low-reliability reconstruction results with furniture manufacturing plant data; (**b**) low-reliability reconstruction results with feed processing plant data; (**c**) low-reliability reconstruction results with rubber manufacturing plant data; (**d**) low-reliability reconstruction results with cotton textile processing plant data. The time intervals 12:30–14:30 are used for validation. The proposed method exhibits a certain degree of reconfiguration capability under low-reliability conditions, with reconfiguration accuracies of 96.68%, 95.81%, 98.59%, and 97.77%.

7.4. Cross-Scene Generalization Ability Analysis

Cross-Scene Analysis Considering Region and Season

The cross-scene reconstruction capability is evaluated by transferring knowledge from existing industrial load scenarios (loads 1–4) to different target industrial scenarios (loads 5–8) under different seasonal conditions. Moreover, to avoid information leakage, the historical conditions used in the proposed method (previous time step, daily, weekly, and yearly patterns) are only constructed from observations available before the target reconstruction point. Future observations and target reconstruction values are not used for model training or scene similarity calculation. The generalization ability and effectiveness of method proposed are illustrated by the comparison of four different types of enterprises (load 5, load 6, load 7, and load 8) in the north and south of China, as well as different seasons (spring, summer, autumn, and winter) of the year. The example in the paper shows the spring load 5–8 in the north to illustrate the superiority. The remaining seasons in the north and other seasons in the south could be found in the Supplementary Materials Figures S1–S4.

Without historical data, the existing collected data needs to be reconstructed and corrected for loads 5 (paper processing factory), 6 (food processing factory), 7 (plastic products factory), and 8 (printing and dyeing factory) by load 1–4.

Firstly, the difference features are calculated, and then the scene similarity with the existing industry data is calculated according to Equation (14). The average similarity is calculated to obtain the similarity of loads 5–8, as shown in Table 7.

Table 7. Similarity results of the 5–8 curve of load in north of China.

Load/Similarity	Load 5	Load 6	Load 7	Load 8
Load 1	0.68	0.25	0.51	0.47
Load 2	0.22	0.71	0.45	0.59
Load 3	0.37	0.49	0.73	0.28
Load 4	0.31	0.29	0.6	0.65

By calculating the similarity, it could be seen that load 1 has the highest degree of similarity with load 5, load 6 has the highest degree of similarity with load 2, load 7 has the highest degree of similarity with load 3, and load 8 has the highest degree of similarity with load 4. Therefore, the data of loads 5, 6, 7, and 8 is corrected by load 1, 2, 3, and 4, respectively. The cross-scene data correction diagram is shown in Figure 8.

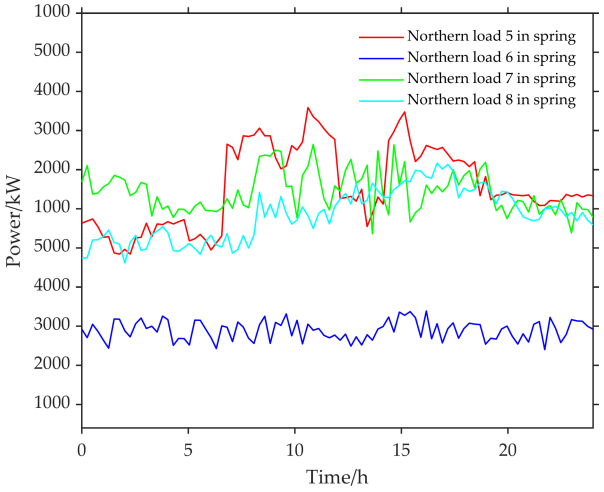


Figure 8. Typical daily load curves of loads 5–8 in northern regions during spring.

To demonstrate the effectiveness of proposed method, this paper assumes that load 5 is with abnormal data collection quality from 8:30 to 10:45 and failed data collection from 11:00 to 13:45. Load 6 collected anomalies between 4:45 and 9:30, with no data collected between 4:45 and 6:30. Load 7 experienced collection anomalies from 16:15 to 18:15 and collection gaps from 18:30 to 20:15. Load 8 experienced a situation where there was a gap in data collection and an abnormal crossover of data collection between 12:30 and 16:15.

After correcting the four types of loads using the proposed method, they are shown in Figure 9a–d. The accuracy of data vacancy reconstruction for loads 5, 6, 7, and 8 is 92.96%, 94.26%, 93.07%, and 91.02%, respectively. The accuracy of data anomaly reconstruction is 91.25%, 93.81%, 91.98%, and 90.17%. It can be seen that the accuracy of data reconstruction and data correction in this article are both greater than 90%.

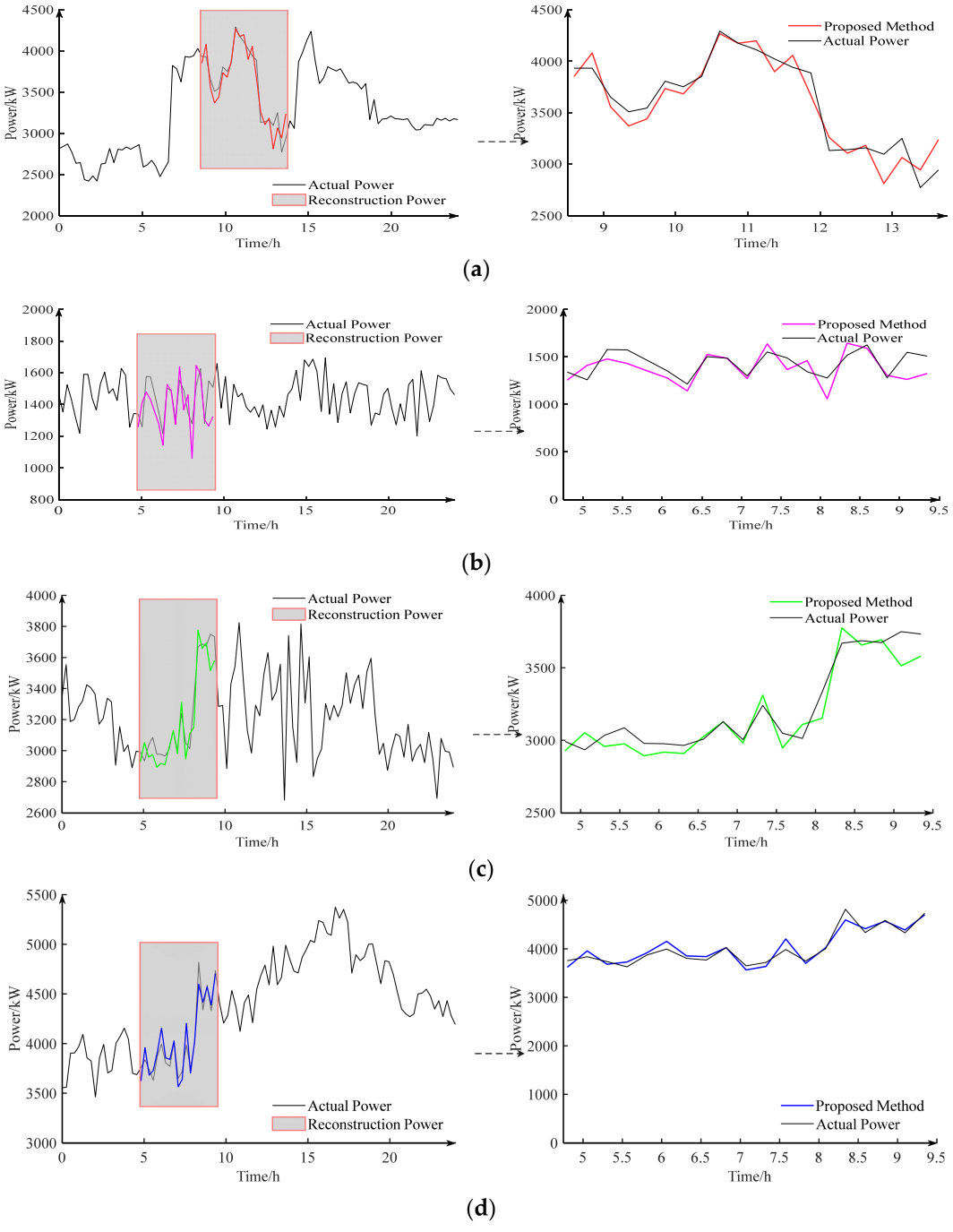


Figure 9. Cross-scene load reconstruction and correction: (**a**) northern load 5 spring curve reconstruction based on similar loads; (**b**) northern load 6 spring curve reconstruction based on similar loads; (**c**) northern

load 7 spring curve reconstruction based on similar loads; (**d**) northern load 8 spring curve reconstruction based on similar loads The accuracy of data vacancy reconstruction for loads 5, 6, 7, and 8 is 92.96%, 94.26%, 93.07%, and 91.02%, respectively. The accuracy of data anomaly reconstruction is 91.25%, 93.81%, 91.98%, and 90.17%. It can be seen that the accuracy of data reconstruction and data correction in this article are both greater than 90%.

7.5. Algorithm Accuracy

The necessity of the classification method in two scenes is illustrated to show the necessity of the proposed classification method.

Scene (1) is directly reconstructed without scene classification.

Scene (2) is classified with the proposed method and then reconstructed.

The table shows a comparison of the accuracy of data reconstruction results in two scenes. It could be seen that the method exhibits significant advantages in multiple dimensions such as error, trend consistency, and distribution consistency compared to scene (1), indicating that the proposed method is with good generalization ability. The model could effectively maintain the overall trend and dynamic structure of electricity consumption behavior. In addition, the significant decrease in KS distance further indicates that the reconstruction results are consistent with the real power data at the statistical distribution level, verifying the stable generalization ability with the conditions, as shown in Table 8.

Table 8. Cross-scene error analysis and comparison.

Method	MAE	RMSE	R	KS
Scene (1)	0.1512	0.183	0.77	0.267
Scene (2)	0.0262	0.035	0.967	0.055

The classification accuracy and algorithm accuracy of temporal-domain and frequency-domain data scene classification are proved by data reconstruction of four types of loads with data loss and low-reliability modes.

In Figure 10, the method visualizes and analyzes the extraction of electricity consumption data for loads 1 to 4 in this paper. It could be seen that, as shown in Figure 10a, the categories of different electricity consumption behaviors of industrial loads are highly mixed, and the boundaries between categories are not clear. After classification by the proposed method, different behavioral patterns form clear, compact, and separated cluster structures in the embedding space. It could effectively suppress redundant fluctuations and noise interference in the original load sequence in Figure 10b. The essential differences in users' electricity consumption behavior are highlighted, providing stronger feature representations for discriminative perception and modeling of anomalies.

The figures show the electricity consumption data of four types of loads for the whole year of 2023–2024, with one data point representing the daily electricity consumption characteristics of the load.

7.6. Method Comparison

Eight methods are compared to evaluate the performance of the proposed method.

1. Adapted Cross-Modal Generative Adversarial Network (Adapted CM-GAN) [19];
2. Adapted GAT-based reconstruction model derived from the feature-reconstruction module in [20];
3. Adapted TSFNN from the traffic-flow interpolation model in [21];
4. Convolutional Transformer Imputation Model (CTIM) [22];
5. Conditional Score-based Diffusion Imputation (CSDI) [23];
6. Bidirectional Recurrent Imputation for Time Series (BRITS) [24];

7. Self-Attention-based Imputation for Time Series (SAITS) [25];
8. The proposed method in this paper.

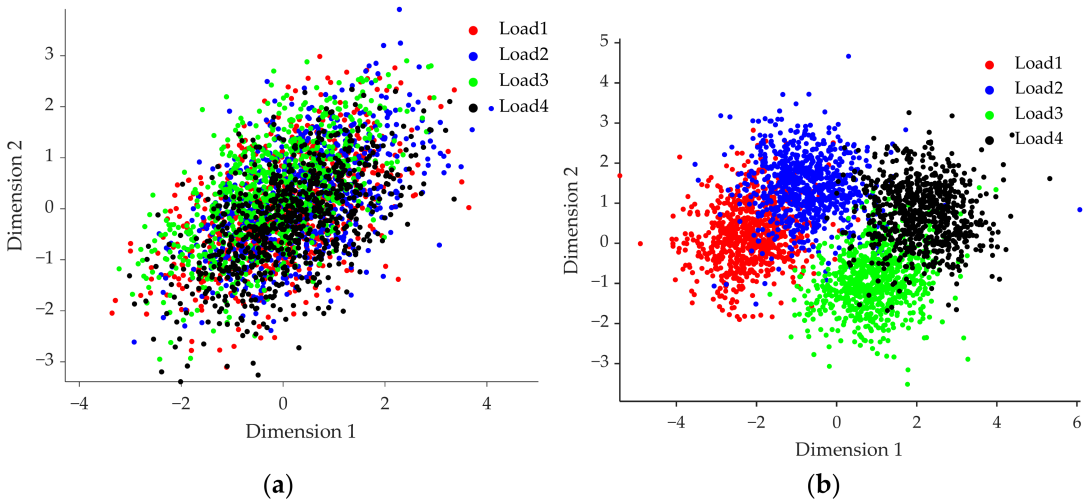


Figure 10. Distribution of load data before and after classification: (**a**) t-SNE visualization of the distribution prior to load classification; (**b**) distribution of load 1–4 with t-SNE classification. In the first figure, the categories of different electricity consumption behaviors of industrial loads are highly mixed, and the boundaries between categories are not clear. It could effectively suppress redundant fluctuations and noise interference in the original load sequence in the second figure.

Furthermore, the mechanism descriptions of all comparison methods are added in Table S10. The key hyperparameters are shown in Table S11. All comparison methods were trained and evaluated using information derived from the same industrial-load dataset. No external data or manually supplied privileged information was introduced. Model-specific masks, positional tags, noise variables, and conditional embeddings were retained only when they were required by the corresponding original architectures.

The performance of the eight methods is compared. All methods are based on the same dataset and compared using four types of industrial load data from 2024. The proposed method shows stronger robustness and adaptability at a high missing rate of 50%. However, the computational complexity of this method is also relatively high.

To evaluate the performance of different methods, seven evaluation metrics, including F1, Area Under the Curve (AUC), Mean Average Precision (MAP), MAE, RMSE, correlation coefficient *R*, and KS, are employed. The results are presented in Tables 9 and 10.

Table 9. Data missing data correction indicators for different methods.

Method	F1	AUC	MAP	MAE	RMSE	*R*	KS
Adapted CM-GAN	0.62	0.726	0.631	0.127	0.181	0.677	0.315
Adapted GAT	0.64	0.733	0.681	0.108	0.162	0.683	0.298
Adapted TSFNN	0.652	0.766	0.706	0.094	0.143	0.724	0.264
CTIM	0.675	0.778	0.755	0.081	0.126	0.752	0.231
CSDI	0.695	0.782	0.764	0.068	0.105	0.772	0.207
BRITS	0.724	0.801	0.776	0.059	0.087	0.789	0.185
SAITS	0.762	0.826	0.793	0.052	0.073	0.808	0.172
Proposed	0.804	0.849	0.806	0.045	0.048	0.938	0.16

The performance of the eight methods is compared using the same dataset and four types of industrial load data collected in 2024. The proposed method exhibits stronger robustness and adaptability under a high missing rate of 50%, although its computational complexity is relatively higher than that of the comparison methods.

Table 10. Low-reliability data correction indicators for different methods.

Method	F1	AUC	MAP	MAE	RMSE	*R*	KS
Adapted CM-GAN	0.591	0.634	0.679	0.119	0.291	0.681	0.341
Adapted GAT	0.69	0.644	0.708	0.103	0.254	0.724	0.311
Adapted TSFNN	0.74	0.671	0.734	0.087	0.221	0.759	0.278
CTIM	0.76	0.713	0.762	0.073	0.198	0.771	0.246
CSDI	0.781	0.788	0.811	0.061	0.176	0.795	0.213
BRITS	0.792	0.812	0.817	0.052	0.162	0.806	0.185
SAITS	0.808	0.837	0.824	0.045	0.153	0.819	0.156
Proposed	0.821	0.855	0.831	0.038	0.147	0.981	0.127

To evaluate the effectiveness of different reconstruction approaches, seven metrics, including F1, AUC, MAP, MAE, RMSE, *R*, and KS, are employed for missing-data reconstruction and low-reliability data correction. The comparison results are presented in Tables 9 and 10. The proposed reliability-aware cross-scene reconstruction method achieves the best overall performance in both tasks.

For missing-data reconstruction, the proposed method achieves an F1 score of 0.804, an AUC of 0.849, an MAP of 0.806, and an R value of 0.938. Compared with Adapted CM-GAN, these four indicators are improved by 29.68%, 16.94%, 27.73%, and 38.55%, respectively. In addition, the proposed method obtains the lowest MAE, RMSE, and KS values of 0.045, 0.048, and 0.160, respectively. These results indicate that the proposed method can accurately recover missing load values while preserving the temporal variation patterns and statistical distribution of the original data.

For low-reliability data correction, the proposed method achieves F1, AUC, MAP, and R values of 0.821, 0.855, 0.831, and 0.981, respectively. Compared with Adapted CM-GAN, the corresponding improvements are 38.92%, 34.86%, 22.39%, and 44.05%, respectively. The proposed method also achieves the lowest MAE, RMSE, and KS values of 0.038, 0.147, and 0.127, respectively. These results demonstrate that the reliability-aware mechanism can effectively utilize the partially preserved information in low-reliability observations, reduce the influence of abnormal disturbances, and preserve the temporal variation patterns of the original load data.

Compared with Adapted TSFNN, the proposed method improves F1, AUC, MAP, and R by 23.31%, 10.84%, 14.16%, and 29.56%, respectively, for missing-data reconstruction. For low-reliability data correction, the corresponding improvements are 10.95%, 27.42%, 13.22%, and 29.25%, respectively. Compared with CTIM, improvements of 19.11%, 9.13%, 6.75%, and 24.73% are obtained for missing-data reconstruction, while improvements of 8.03%, 19.92%, 9.06%, and 27.24% are obtained for low-reliability data correction.

BRITS and SAITS also achieve competitive performance because they model bidirectional temporal dependencies and self-attention relationships, respectively. Nevertheless, compared with BRITS, the proposed method improves F1, AUC, MAP, and R by 11.05%, 5.99%, 3.87%, and 18.88%, respectively, for missing-data reconstruction and by 3.66%, 5.30%, 1.71%, and 21.71%, respectively, for low-reliability data correction. Compared with SAITS, the corresponding improvements are 5.51%, 2.78%, 1.64%, and 16.09% for missing-data reconstruction and 1.61%, 2.15%, 0.85%, and 19.78% for low-reliability data correction.

Overall, the proposed method consistently outperforms the comparison methods in terms of numerical accuracy, temporal-pattern preservation, and distributional consistency. The performance improvements are mainly attributed to the integration of cross-scene prior information, diffusion-based conditional generation, reliability-aware constraints, residual correction, and adversarial refinement.

Based on the comprehensive reconstruction and correction results, the proposed method maintains the leading performance across the first four evaluation metrics and exhibits the most extensive overall coverage with a relatively balanced metric distribution in the radar chart shown in Figure 11. This indicates that the proposed method is capable of effectively identifying the missing areas and the low-reliability areas, thereby enabling accurate anomaly detection and high-quality data recovery under different data quality conditions.

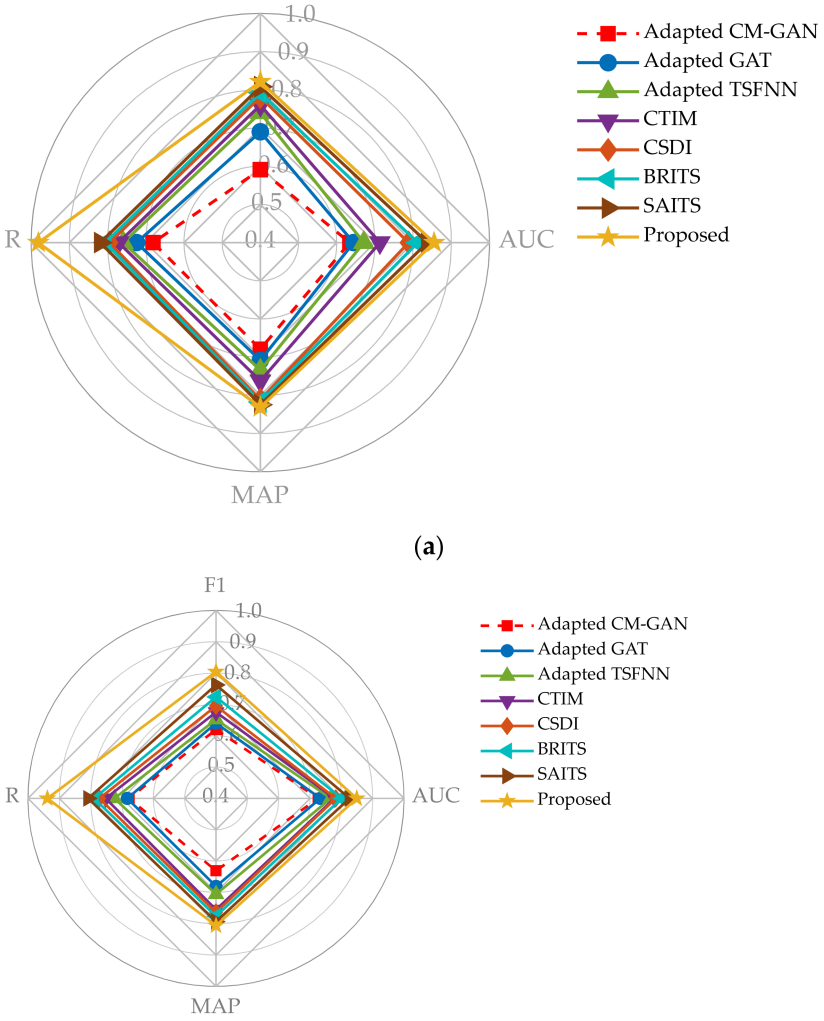


Figure 11. Radar-chart comparison of the eight methods for: (**a**) low-reliability data correction and (**b**) missing-data reconstruction at a missing rate of 50%.

The computational characteristics vary across different model architectures and do not change monotonically across all indicators. In particular, Adapted CM-GAN employs relatively large fully connected generator and discriminator networks, resulting in a comparatively large parameter count. As shown in Tables 11 and 12, under the 50% missing-data condition, Adapted CM-GAN contains 3.63 million parameters and requires 0.160 GFLOPs per sample, whereas the proposed method contains 1.20 million parameters and requires 0.306 GFLOPs per sample. Therefore, compared with Adapted CM-GAN, the proposed method contains 66.94% fewer parameters but requires 91.25% more FLOPs per sample.

Table 11. Algorithm performance improvement compared to Adapted CM-GAN with 50% data missing.

Method	Relative Change in Inference Time (%)	Relative Change in Memory Usage (%)	Relative Change in Parameter Count (%)	Average Improvement in Accuracy Metrics (%)
Adapted GAT	−0.8	19.17	−75.21	3.25
Adapted TSFNN	2.4	22.33	−75.12	7.37
CTIM	5.6	30.83	−72.42	11.69
CSDI	11.6	33.5	−69.7	13.73
BRITS	17.6	36.89	−69.15	16.66
SAITS	28.4	43.45	−68.04	20.43
Proposed	36	48.79	−66.94	28.23

Table 12. Algorithm performance with 50% data missing.

Method	Parameters (M)	FLOPs (G/Sample)	Inference Time (S/Sample)	Training Time (min)	Memory Usage (GB)
Adapted CM-GAN	3.63	0.16	0.025	12.12	4.12
Adapted GAT	0.9	0.223	0.0248	14.88	4.91
Adapted TSFNN	0.903	0.23	0.0256	15.35	5.04
CTIM	1.001	0.238	0.0264	15.84	5.39
CSDI	1.1	0.251	0.0279	16.74	5.5
BRITS	1.12	0.264	0.0294	17.65	5.64
SAITS	1.16	0.288	0.0321	19.18	5.91
Proposed	1.2	0.306	0.034	20.4	6.13

The inference time of the proposed method increases from 0.025 to 0.034 s per sample, corresponding to an increase of 36.00%, while its training time increases from 12.12 to 20.40 min, corresponding to an increase of 68.32%. Its memory usage increases from 4.12 to 6.13 GB, representing an increase of 48.79%. Despite this additional computational overhead, the proposed method achieves the highest average relative improvement of 28.23% across the four positively oriented accuracy metrics, namely F1, AUC, MAP, and R, compared with Adapted CM-GAN.

These results indicate that the proposed framework requires additional computational resources because of the combined use of cross-scene feature encoding, iterative diffusion-based reconstruction, and adversarial refinement. Nevertheless, the substantial improvements in reconstruction accuracy demonstrate a reasonable trade-off between computational cost and reconstruction performance.

Accordingly, the revised manuscript distinguishes the module-level computational characteristics of the MLP-based feature-fusion mechanism from the overall computational complexity of the complete framework. The MLP-based fusion mechanism efficiently integrates time- and frequency-domain features and controls the additional computational burden at the feature-encoding stage. This statement does not imply that the complete framework has a lower computational cost than all comparison methods. The diffusion-based reconstruction and adversarial refinement modules inevitably increase the FLOPs, training time, inference time, and memory consumption, although the total parameter count remains lower than that of Adapted CM-GAN.

In practical industrial monitoring systems, the proposed framework can be integrated into existing data-acquisition and monitoring platforms. Load measurements collected

by smart meters and sensors are first transmitted to the reconstruction framework. When missing or low-reliability observations are detected, the available measurements are used to extract time- and frequency-domain features and retrieve similar historical operating information as conditional inputs. The diffusion-based reconstruction module then recovers the missing or abnormal observations, while the adversarial refinement module further improves the local continuity and waveform consistency of the reconstructed sequences. The reconstructed data are subsequently returned to the industrial monitoring system for load analysis, forecasting, and operational decision-making. Depending on the available computing resources and application requirements, the framework can be deployed on edge computing devices near the data-acquisition layer or on centralized industrial monitoring servers. For large-scale industrial datasets, multiple load sequences can be processed in batches or in parallel to improve computational efficiency.

All models were evaluated on the same hardware and software platform, consisting of an Intel Core i7 3.60 GHz processor, 32 GB RAM, Windows 11, and MATLAB R2023b. The inference time was measured with a batch size of 1 after warm-up and averaged over 10 independent runs. The training time represents the wall-clock time required to reach the predefined stopping criterion.

Furthermore, the revised manuscript discusses the practical deployment of the proposed framework in industrial monitoring systems. Depending on the available computing resources and application requirements, the framework can be deployed on edge-computing devices close to the data-acquisition layer or on centralized industrial monitoring servers. For large-scale industrial datasets, multiple load sequences can be processed in batches or in parallel to improve computational efficiency.

8. Conclusions

To address the concurrent issues of missing values and anomalies in industrial plant electricity consumption monitoring data, a cross-scene industrial load data reconstruction method is proposed based on adversarial learning with improved denoising diffusion. It aims to address the limitations in handling complex data distributions and observational uncertainties. By combining a noisy diffusion process and a denoising generator, noise and anomalies in observational data could be effectively mitigated. Simultaneously, the adversarial learning mechanism between the generator and discriminator ensure efficient generation and distributional consistency during data reconstruction. The conclusions are as follows:

1. The MLP-based time–frequency feature-fusion module efficiently integrates scene features and controls the additional computational burden at the feature-encoding stage. The complete framework achieves average reconstruction and correction accuracies of 97.17% and 97.21% under the 50% missing-data and 50% low-reliability conditions, respectively.
2. By hierarchical model of data reliability, observational data are categorized into missing data, low-reliability data, and high-reliability data. Differentiated processing of varying data quality is achieved. The experiments demonstrate that $\delta = 0.3$ achieves the optimal balance among accuracy, temporal consistency, and distributional fidelity.
3. The proposed method can simultaneously handle missing data and low-reliability data. Particularly during the reconstruction of low-reliability data, the weak consistency constraint based on original observations effectively preserves potentially useful information, thereby promoting the robustness and accuracy of the reconstructed results. Compared with other algorithms, the superior MAE, RMSE, R, and KS distance indicate that the distribution differences and trend variations in this study exhibit smaller errors relative to actual data.

In addition, although the proposed method performs well on the datasets considered in this study, its generalization capability still needs to be further evaluated using larger-scale datasets covering more industrial sectors, longer observation periods, and more complex combinations of missing and anomalous data.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/sym18081358/s1, The supplementary materials has been uploaded along with this document. Figure S1. Southern load 5–8 spring cross-scene reconstruction and correction. Figure S2. North-south load 5–8 summer cross-scene reconstruction and correction. Figure S3. North-south load 5–8 autumn cross-scene reconstruction and correction. Figure S4. North-south load 5–8 winter cross-scene reconstruction and correction. Table S1. Algorithm parameters and numerical settings in this paper. Table S2. Algorithm metrics under different data missing rates from 10–50%. Table S3. Impact of algorithm structure on accuracy. Table S4. Comparison of different source-scene selection strategies. Table S5. Threshold sensitivity table. Table S6. Cross-scene reconstruction performance under different unseen conditions. Table S7. Statistical robustness analysis under repeated experiments. Table S8. Algorithm pseudocode. Table S9. Comparison of classification methods. Table S10. Detailed information of baseline methods. Table S11. Hyperparameter settings of baseline methods.

Author Contributions: Conceptualization, data curation, J.C.; methodology, original draft preparation, Y.Z. (Yuxiu Zang); supervision, resources, Y.Z. (Yan Zhao); investigation, visualization, W.G.; project administration, software, review and editing, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors declare that this study received funding from "Research on the Evaluation of the Value of New-Type Multi-Form Typical Load Interaction and the Dynamic Optimization of Marketization Combination", State Grid Liaoning Electric Power Company—Liaoyang Power Supply Company, grant number: 2025-0-43-215. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Materials. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors sincerely thank the anonymous reviewers for their critical comments and suggestions for improving the manuscript.

Conflicts of Interest: The authors declare that this study received funding from "Research on the Evaluation of the Value of New-Type Multi-Form Typical Load Interaction and the Dynamic Optimization of Marketization Combination", State Grid Liaoning Provincial Electric Power Company, Liaoyang Power Supply Company. Grant number: 2025-0-43-215. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

Abbreviations

Variables	Definition	Variables	Definition	Variables	Definition
DRL	Deep Reinforcement Learning	DDPM	Denoising Diffusion Probabilistic Model	b_1^{time}, b_1^{fre}	The bias vectors for the first layer
LSTM	Long Short-Term Memory	TimeGNN	Temporal Graph Neural Network	ω_1^{time}, ω_1^{fre}	The weight matrix for the second layer
VMD	Variational Mode Decomposition	MAE	Mean Absolute Error	b_1^{time}, b_1^{fre}	The bias vectors for the second layer
CNNs	Convolutional Neural Networks	RMSE	Root Mean Squared Error	$SI^{(i,j)}$	The similarity parameter in scene i and j
GRUs	Gated Recurrent Units	KS distance	Kolmogorov-Smirnov distance	$\Sigma(i)$ $\Sigma(j)$	The i and j load scenarios
P_t	Load power data	$\lvert X[m] \rvert$	The signal amplitude	n	The sampling time span

T	The scheduling cycle of time	$\left\|X[m]\right\|^2$	The energy of the signal	R_2	The persistent anomaly judgment mask
$\max P_t$ $\min P_t$	The maximum power. The minimum power.	f_m	The frequency index in the discrete Fourier transform	σ_Δ	The historical load variation standard deviation
R	The extreme value difference	FC	The center of mass frequency	$P_{0.5}$	The median of the load sequence
S	The standardized measure of short-term volatility	MSF	The frequency variance	L_{re}	The denoising loss
R/S	The relabeled extreme difference	H	The spectral entropy	L_{GAN}	The adversarial loss
$\widetilde{P}(f_m)$	The normalized spectral energy ratio	R_1	The mutation anomaly judgment mask	$\hat{p}_{0,t}^{\text{miss},\Omega_0}$	The reconstruction load value output
ϕ^{time}	Temporal-domian conditional encoder	P_t P_{t-1}	The power values at time at time points t and $t-1$	$L_{\text{re}}^{\Omega_0}(\mathcal{G}_\theta)$	The reconstruction loss of the missing region
ϕ^{fre}	Frequency-domain conditional encoder	$\Delta\overline{P}_t$	The average historical variation	ΔP_t^{obs} $\nabla^2 P_t^{\text{obs}}$	first-order difference second-order difference
ω_1^{time}, ω_1^{fre}	The weight matrices for the first layer	$\lambda_{\mathcal{G}}$	The generator weight	P_t^{obs}	The data collection value.
$\left\|h^{\Sigma(i)}\right\|_2$, $\left\|h^{\Sigma(j)}\right\|_2$	L2 norm of the two feature vectors	λ_{GAN}	The adversarial loss weight	$P_0^{\text{miss},\Omega_0}$	The actual load data distribution
A	The missing mask	$P_x^{\text{miss},\Omega_0}$	The true value of the missing region	P_0	The original complete load data
		P_t^{fact}	The actual data value	R_1	The mutation anomaly judgment mask

References

1. Hou, Z.; Fu, Q.; Li, W.; Wang, Y.; Dong, Z.; Ye, X.; Chen, X.; Zhang, F. Symmetry-Aware interpretable anomaly alarm optimization method for power monitoring systems based on hierarchical attention deep reinforcement learning. *Symmetry* **2026**, *18*, 216. [CrossRef]
2. Xiong, R.; Sun, X.; Meng, X.; Shen, W.; Sun, F. Advancing fault diagnosis in next-generation smart battery with multidimensional sensors. *Appl. Energy* **2024**, *364*, 123202. [CrossRef]
3. He, C.; Lu, Y.; Wang, W. Adaptive instantaneous frequency synchrosqueezing transform and enhanced AdaBoost for power quality disturbance detection. *Symmetry* **2026**, *18*, 475. [CrossRef]
4. Jin, Z.Y.; Zhao, J.B.; Ding, L.; Chakrabarti, S.; Gryazina, E.; Terzija, V. Power system anomaly detection using innovation reduction properties of iterated extended Kalman filter. *Int. J. Electr. Power Energy Syst.* **2022**, *136*, 107613. [CrossRef]
5. Zang, T.L.; Wang, S.J.; Li, C.Z.; Liu, Y.; Xiao, Y.; Wang, Z.; Yu, X. Anomaly detection method for cyber physical power system based on bilateral data fusion. *Int. J. Electr. Power Energy Syst.* **2025**, *169*, 110813. [CrossRef]
6. Qiu, Z.Q.; Wang, L.Z.; Pan, H.L.; Zhang, B.; Wang, L.; Hong, Y. Research on fault sensor data reconstruction for 10MW offshore wind turbines based on LSTM. *Acta Sol. Energy* **2024**, *45*, 342–349. [CrossRef]
7. Zhao, W.S.; Wang, J.Y.; Li, X.J. Splicing and reconstruction of energy storage battery operational data. *High Volt. Technol.* **2024**, *50*, 5594–5605. [CrossRef]
8. Liu, Q.C.; Zhong, Y.; Lin, C. Electricity data cleaning and interpolation based on robust nonnegative matrix factorization. *Power Syst. Technol.* **2024**, *48*, 2103–2112.
9. Ge, J.X.; Cai, G.W.; Liu, C.; Yuan, F.; Li, Y. A virtual power plant load data reconstruction method based on federated learning and multimodal BERT. *Trans. Chin. Soc. Electr. Eng.* **2025**, 1–16.
10. Zhou, X.; Zhang, X.; Dai, J.; Zhang, T. Photovoltaic power prediction technology based on multi-source feature fusion. *Symmetry* **2025**, *17*, 414. [CrossRef]
11. Yin, X.Y.; Cui, H.; Chen, X.Y.; Ding, Y.; Song, R.; Li, C. Generalized load modeling and parameter discrimination for distribution networks considering distributed PV fault Ride-Through capability. *Autom. Electr. Power Syst.* **2025**, *49*, 55–64.
12. Li, Y.M.; Li, H.J.; Li, X.J.; Cao, Y.; Xie, Y.; Zhang, C. Short-Term electricity load forecasting method based on the Power Time Mixer model. *Power Syst. Technol.* **2025**, *49*, 4216–4227. [CrossRef]
13. Fan, C.; Peng, Y.F.; Shen, Y.P.; Guo, Y.; Zhao, S.; Zhou, J.; Li, S. Variable scale multilayer perceptron for helicopter transmission system vibration data abnormity beyond efficient recovery. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108184. [CrossRef]
14. Afzal, S.; Ziapour, B.M.; Shokri, A.; Shakibi, H.; Sobhani, B. Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms. *Energy* **2023**, *282*, 128446. [CrossRef]

15. Sun, L.; Dai, W.Z.; Muhammad, G. Gating memory network multilayer perceptron for traffic forecasting in internet of vehicles systems. *IEEE Internet Things J.* **2025**, *12*, 4783–4793. [CrossRef]
16. Mi, Y.; Lu, C.K.; Shen, J.; Yang, X.; Ge, L. Generation of extreme scenarios for wind power output based on conditional generative adversarial networks. *High Volt. Eng.* **2023**, *49*, 2253–2263. [CrossRef]
17. Wang, K.; Wu, F.; Cao, R.L.; Yao, J.G.; Huang, Q.F.; Li, Y.P. A review of demand response behavior characteristics and modeling in market environments. *Autom. Electr. Power Syst.* **2024**, *48*, 1–14.
18. Sun, Y.; Mao, Y.H.; Li, Z.K.; Zhang, X.; Li, F. Integrated clustering method for consumer load characteristics and adjustable potential based on power big data. *Trans. Chin. Soc. Electr. Eng.* **2021**, *41*, 6259–6271. [CrossRef]
19. Kang, M.; Zhu, R.; Chen, D.; Liu, X.; Yu, W. CM-GAN: A cross-modal generative adversarial network for imputing completely missing data in digital industry. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 2917–2926. [CrossRef] [PubMed]
20. Gu, Y.K.; Li, Y.Y.; Fang, C.; Yang, L.; Liu, R.; Zhou, Z.; Yan, Z. Short-Term electric vehicle charging load forecasting based on multi-source data reconstruction. *Power Syst. Technol.* **2025**, 1–14. [CrossRef]
21. Wang, G.; Mao, Y.; Liu, X.; Liang, H.; Li, K. Spatio-Temporal Recursive Method for Traffic Flow Interpolation. *Symmetry* **2025**, *17*, 1577. [CrossRef]
22. Wang, Y.; Ding, H.; Li, H. Multivariate Time-Series Missing Data Imputation with Convolutional Transformer Model. *Symmetry* **2025**, *17*, 686. [CrossRef]
23. Tashiro, Y.; Song, J.M.; Song, Y. CSDI: Conditional score-based diffusion models for probabilistic time series imputation. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 24804–24816.
24. Cao, W.; Wang, D.; Li, J. BRITS: Bidirectional Recurrent Imputation for Time Series. *Adv. Neural Inf. Process. Syst.* **2018**. [CrossRef]
25. Du, W.; Côté, D.; Liu, Y. SAITS: Self-Attention-based Imputation for Time Series. *Expert Syst. Appl.* **2023**, *219*, 119619. [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.