Applied Energy 401 (2025) 126771

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apen

Robust photovoltaic forecasting under severe data missingness via multi-domain collaboration and covariate interaction

Ke Yan [a], Jian Liu [a], Jiazhen Zhang [a], Fan Yang [b], Yuan Gao [c,*], Yang Du [d]

[a] Mechanical and Vehicle Engineering, Hunan University, Changsha, 410082, China
[b] School of Software Technology, Zhejiang University, Hangzhou, 310027, China
[c] The Center for Energy Systems Design (CESD), International Institute for Carbon-Neutral Energy Research (WPI-I2CNER), Kyushu University, Fukuoka, 819-0395, Japan
[d] College of Science and Engineering, James Cook University, Cairns, Australia

HIGHLIGHTS

- Evaluating whether covariate inclusion enhances model forecasting performance.
- Maintaining prediction accuracy and robustness under extensive input-data missingness.
- Showing through experiments that covariates improve solar irradiance forecasts.
- First examining photovoltaic forecasting under data-missing conditions.

ARTICLE INFO

Keywords:
Photovoltaic forecasting
Covariate information
Multi-source domain collaboration
Transfer learning fine-tuning
Shapley additive explanations

ABSTRACT

High-quality photovoltaic (PV) power forecasting is essential for efficient energy management and reliable grid integration, yet real-world data are often plagued by extensive missingness in both target and auxiliary variables. To address this challenge, we propose MDCTL-MCI, a missingness-aware forecasting framework that jointly leverages signal decomposition, multi-scale covariate interaction, and multi-domain collaborative transfer learning. First, multivariate singular spectrum analysis (MSSA) denoises and reconstructs incomplete time series, enhancing underlying temporal structures without explicit imputation. Next, a lightweight multiscale covariate interaction (MCI) module models interactions among reconstructed PV power, global horizontal irradiance, direct normal irradiance, and total solar irradiance at varying temporal resolutions, capturing both local fluctuations and global trends. Finally, a multi-source domain collaborative transfer learning strategy aggregates knowledge from multiple PV sites to form a global model, which is then fine-tuned on a small set of high-quality, MSSA-processed samples at each site. By freezing all but the output layer during fine-tuning, MDCTL-MCI adapts efficiently to local data heterogeneity. Extensive experiments on four Chinese PV installations reveal that, compared to baseline methods, the proposed method improves average accuracy by 10.5 % under complete data conditions and by 15.3 % under various missing data scenarios.

1. Introduction

1.1. Background and literature review

Renewable energy technologies have attracted wide attention across the world in the context of developing low-carbon economy and sustainable societies [1]. Among the various renewable energy technologies, solar photovoltaics (PVs) are widely regarded as one of the most promising technologies due to their cleanliness, sustainability, and low carbon emissions [2,3]. In 2023, the newly installed global PV capacity reached a record high of 346 GW, representing 75 % of all newly added renewable energy capacity [4]. This underscores the pivotal role of solar PV in shaping the future energy systems [5].

The solar PV generation is influenced by various external factors, such as the intermittency of solar irradiance, leading to considerable volatile PV power output [6]. Accurate PV power forecasting algorithms for PV systems can effectively identify the intermittency and alleviate

* Corresponding author.
Email address: gao.yuan.564@m.kyushu-u.ac.jp (Y. Gao).

https://doi.org/10.1016/j.apenergy.2025.126771
Received 24 April 2025; Received in revised form 21 August 2025; Accepted 14 September 2025
Available online 23 September 2025

Nomenclature

Abbreviations
PV Photovoltaic
DPV Distributed photovoltaic power
LSTM Long short-term memory
MSSA Multivariate singular spectrum analysis
MCI Multiscale covariate interaction
SHAP Shapley additive explanation
MF Multilevel data fusion
NBEA Neural basis expansion analysis
CNN Convolution neural network
DeepCNN Deep convolution neural network
MSMixer Multiscale network with mixed features
BiLSTM Bidirectional long short-term memory
MLP Multi-layer perceptron
FL Federal learning
BTM Behind the meter
FedSign Sign federated learning
DP Differential privacy
FFARM Feature filtering and adaptive reconstruction module
VMD Variational mode decomposition
ACCNet An innovative capsule CNN
PFedGAN Personalized federated generative adversarial network
MSCM Multi-scale convolution module
SVD Singular value decomposition
TMM Temporal mixing module
FMM Feature mixing module
MDCTL Multi-domain collaborative transfer learning
AMP Automatic mixed precision
Float16 Half precision floating-point numbers
GHI Global horizontal irradiance
DNI Direct normal irradiance
TSI Total solar irradiance
MAE Mean absolute error
MSE Mean squared error
CV Coefficient of variation

volatility in PV power output, which is important for smart grid integration and applied energy management [7,8]. Conventional photovoltaic power forecasting methods typically rely on dynamic physical models of the PV systems. However, due to the structural diversity of the PV systems and the significant variability in external environmental factors, physical modeling-based methods exhibit notable limitations in practical applications. In contrast, data-driven deep learning based PV power generation forecasting algorithms offer more stable and reliable results [9–11]. In recent years, a number of novel deep learning model architectures have been proposed and extensively applied in PV power forecasting research across varying temporal and spatial scales. Table 1 summarizes the most related studies mentioned above. The term ‘Input covariate included’ denotes whether the forecasting model incorporates covariates as input features. The term ‘Covariate impact analysis’ denotes whether the study quantifies the impact of covariate inclusion on predictive accuracy, thereby verifying its positive contribution. The term ‘Covariate missing evaluation’ denotes whether the study accounts for scenarios with missing or low-quality covariates, and evaluates the stability of model predictions under such conditions.

Based on the literature study listed in Table 1, the performance of deep learning-based PV power forecasting methods is predominantly dependent on the availability of sufficient and high-quality input information [24,25]. Most existing research adopts a covariate input strategy, integrating multiple related time series data as auxiliary information to improve the forecasting accuracy (Table 1). For example, Hu et al. [12] combine long short-term memory (LSTM) networks with self-attention mechanisms to effectively utilize weather forecast data, enhancing model forecasting accuracy. Almaghrabi et al. [13] propose a method that integrates multi-level data fusion and neural-based extension analysis for multivariate solar power time series forecasting, leading to improved accuracy at the regional level. Zhang et al. [16] introduce a multi-scale network with hybrid features to enhance short-term PV power forecasting accuracy, effectively integrating multiple input data sources and evaluating the effectiveness of regional meteorological forecasts.

1.2. Research gap and our contribution

As summarized above, the incorporation of covariates is theoretically beneficial for improving predictive performance. Nevertheless, existing studies primarily integrate them directly into models without systematically investigating their underlying mechanisms of influence or their marginal contributions. Furthermore, these models depend critically on high-quality covariate data to achieve satisfactory forecasting accuracy. However, photovoltaic power forecasting often faces constraints of limited data availability and suboptimal data quality in practice. For instance, solar irradiance, which is widely acknowledged as a key covariate, is recorded at only about 5 % of meteorological stations in China, owing to the substantial costs of instrument acquisition, maintenance, and calibration [26]. In addition, data collection is susceptible to obstructions and environmental interferences, leading to measurement noise and missing values, which further undermine data reliability.

Most existing studies concentrate on imputing raw observational data to mitigate the issue of low-quality measurements. For instance, Li et al. employed two classes of neural network algorithms to enhance estimation accuracy and applicability, effectively addressing the challenges of missing values and poor-quality measurements in hourly solar irradiance data collection [26]. Bacsakin et al. introduced a novel technique that combines the implementation of machine-learning algorithms with a metaheuristic optimization algorithm to estimate and impute missing values in solar-irradiance measurements [27]. Alternatively, learning directly from the available data and generating synthetic samples for model training and evaluation is also feasible. For instance, Zhang et al. proposed a diffusion-model-based restoration forecasting method that enables load-sequence generation, quantile prediction, and missing-data imputation [28]. However, data-imputation models introduce additional methodological overhead and make it difficult to quantify how errors in the imputed data may impact the final model’s predictive performance.

Based on the current research landscape, this study considers how covariate information can be effectively utilized to enhance predictive performance, and whether the inherent generalization capacity and robustness of deep learning algorithms can be leveraged to directly forecast solar irradiance in the presence of substantial missing input features, without performing additional imputation, and to conduct a thorough analysis of the various influencing factors and the underlying predictive mechanisms.

Broadly speaking, we summarize the current research gaps as follows:

(1) How does the covariate information effectively and positively influence the performance in the PV power forecasting scenarios? Whether the inclusion of covariates can improve the model’s predictive performance.

Table 1
A literature review of most recent studies on data-driven deep learning based PV power generation forecasting algorithms.

References	Model	Input covariate included	Covariate impact analysis	Covariate missing evaluation	Contribution	Limitation
Proposed	MDCTL-MCI	✓	✓	✓	A multi-source domain collaborative learning framework based on covariate interactions is proposed for robust PV forecasting	Due to the limitations of the dataset, the selected covariates require further expansion.
[12]	LSTM-Attention	✓	✓	×	The proposed method effectively leverages weather forecasts and historical data to improve the forecasting accuracy	The impact of weather forecast data errors on model performance has not been considered.
[13]	MF-NBEA	✓	×	×	MF-NBEA is employed to forecast multi-dimensional solar irradiance, providing accurate and interpretable predictions	The model incurs high computational costs when handling high-dimensional data.
[14]	WNPS-LSTM-Informer	✓	×	×	It addresses the challenges posed by the randomness, intermittency, and uncertainty in large-scale PV systems	Forecasting accuracy is significantly affected by environmental constraints
[15]	DeepCNN-SHAP	✓	✓	×	A CNN-based PV forecasting method with shading model integration significantly improves forecasting accuracy.	The proposed PV shading model lacks generalization, failing to adapt to obstacles of various shapes.
[16]	MSMixer	✓	×	×	MSMixer effectively integrates meteorological forecasts to capture valuable feature information and improves forecasting stability.	The study has not yet considered uncertainty analysis.
[17]	CNN-BiLSTM-Attention	✓	×	×	The framework takes account of the uncertainties and shows advanced forecast reliability and sharpness.	The model lacks sufficient comparison with advanced models and ablation studies.
[18]	MLP-FL	✓	×	×	The application of FL in BTM energy forecasting has been explored, achieve the trade-off between accuracy and privacy.	The algorithm's computational efficiency needs improvement.
[19]	FedSign-DP	×	×	×	Forecasting accuracy is improved while preserving user privacy, with a 32-fold reduction in bandwidth consumption	The algorithm offers improved speed but delivers slightly lower accuracy than the local model.
[20]	Timeset-FFARM	✓	×	×	The framework introduces FFARM for improved key feature extraction and proposes an AL loss to enhance forecasting accuracy.	As the forecasting horizon extends, the model's advantages diminish, and training incurs higher time costs.
[21]	VMD-ACCNet	✓	×	×	A two-stage feature fusion method is proposed to enable more accurate and efficient short-term PV power forecasting	The signal decomposition methods considered in the study need further optimization.
[22]	TimesNet-DeepAR	✓	×	×	PFedGAN handles DPV data scarcity and heterogeneity, while TNE-DeepAR boosts accuracy via weather inputs.	GAN increases the required training rounds for convergence.
[23]	Physical-Transformer-LSTM	✓	×	×	The physical constrained PV power forecasting network is proposed, achieving higher accuracy with better robustness.	The physical model incurs substantial computational cost and demonstrates poor generalization capability.

(2) How to maintain the model's predictive accuracy and robustness in the presence of substantial missing input data, including missing values in both the target variable and covariates?
(3) Theoretically, what are the main factors for producing quality-guaranteed PV power forecasting results?

To address the above three questions, in this study, a multi-domain collaborative transfer learning combining a multiscale covariate interaction (MDCTL-MCI) framework is proposed for quality-guaranteed PV power forecasting. First, a multivariate singular spectrum analysis (MSSA) method is applied for data noise reduction and data expression enhancement. Second, a lightweight multi-scale covariate interaction approach is utilized for effectively modeling relations among variables and extracting deep temporal patterns. Next, a multi-source domain collaboration strategy is employed to enhance the robustness of the proposed model under conditions of low-quality data. Last, Shapley additive explanation (SHAP) technique is adopted for finding the key factors of the forecasting performance. The main contributions of this study are summarized as follows:

(1) **A novel time series modeling framework integrating signal decomposition with multi-scale covariate interaction is proposed to address the challenges of high noise interference and complex multi-feature coupling.** The framework employs MSSA to denoise and structurally enhance the original multivariate time series, preserving key trends and suppressing random fluctuations. MCI is utilized to capture dynamic dependencies between the target variable and multiple meteorological covariates at different time scales, enabling joint modeling of local variations and global trends. The proposed framework effectively leverages covariate information, significantly improving forecasting accuracy and stability, and provides a generalizable solution for multivariate time series modeling in complex environments.
(2) **A multi-source domain collaborative training framework is developed to address cross-site data distribution disparities and suboptimal data quality in photovoltaic forecasting.** Built upon MCI, the framework aggregates parameters from multiple PV sites through weighted averaging to enable cross-site knowledge sharing and complementary feature extraction. Transfer learning is applied using a small set of high-quality local data processed by MSSA, thereby enhancing adaptability to site-specific conditions while mitigating the adverse effects of low-quality data. In addition, SHAP is incorporated to evaluate feature contributions for both global and local models under varying data quality, uncovering underlying predictive mechanisms and patterns of information utilization. This collaborative framework demonstrates strong

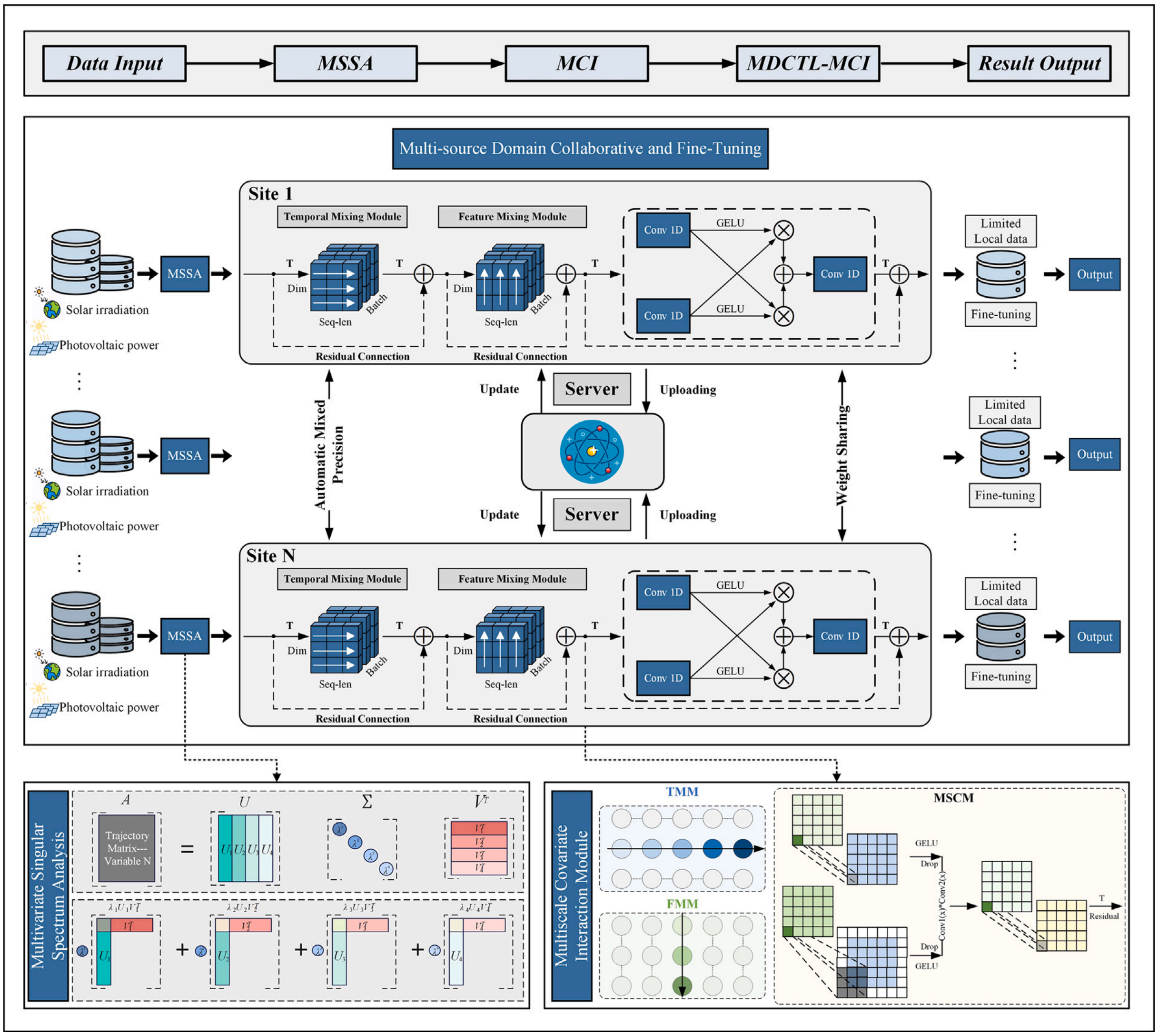

Fig. 1. The overall flowchart of the proposed forecasting framework.

applicability for multi-source forecasting and provides methodological support for building robust cross-regional photovoltaic forecasting models.

2. Methodology

This section provides a detailed description of the proposed methodology. As illustrated in Fig. 1, the overall architecture of the proposed method comprises four main components: First, MSSA is employed to capture more informative patterns while effectively filtering out noise. Second, MCI captures multi-scale temporal dependencies by modeling the interactions among feature variables. Third, the multi-source domain collaborative learning approach is adopted to improve the model's generalization ability and mitigate the impact of low-quality data on overall forecasting accuracy. Last, transfer learning fine-tuning is applied to better adapt the global model to local data distributions and mitigate the effects of data heterogeneity. The implementation details of each component are provided in the subsequent subsections.

2.1. Multivariate singular spectrum analysis

Multivariate singular spectrum analysis (MSSA) is a non-parametric technique for analyzing multivariate time series. It constructs a trajectory matrix from the input data, performs singular value decomposition (SVD), and then groups and reconstructs the informative components to effectively uncover latent patterns within the data. This process enhances the model's capacity to capture the dynamic characteristics of time series. In the domain of PV forecasting, common signal decomposition methods are primarily applied during the noise removal stage. However, their potential for deep feature extraction and signal reconstruction has yet to be fully explored, which limits the exploitation of the data's intrinsic value. Compared with traditional methods, MSSA demonstrates enhanced signal decomposition capabilities. It not only effectively filters noise components from the input data but also adaptively extracts valuable trends and periodic features from low-quality and complex signals. In this study, MSSA is utilized to fully leverage its capability to extract critical information from low-quality data, thereby enhancing both the predictive performance and robustness of the model. The workflow of MSSA is depicted in Fig. 2.

The detailed procedure of MSSA is described as follows: Define the input multivariate time series X with length N and p variables, where x_{ij} denotes the value of the i-th variable at the j-th time point. Taking the subsequence X_i as an example, the first step of MSSA involves constructing the trajectory matrix. By choosing suitable embedding dimensions and delay parameters, the time series X_i is transformed into the trajectory matrix X_i' using the sliding window method, as shown

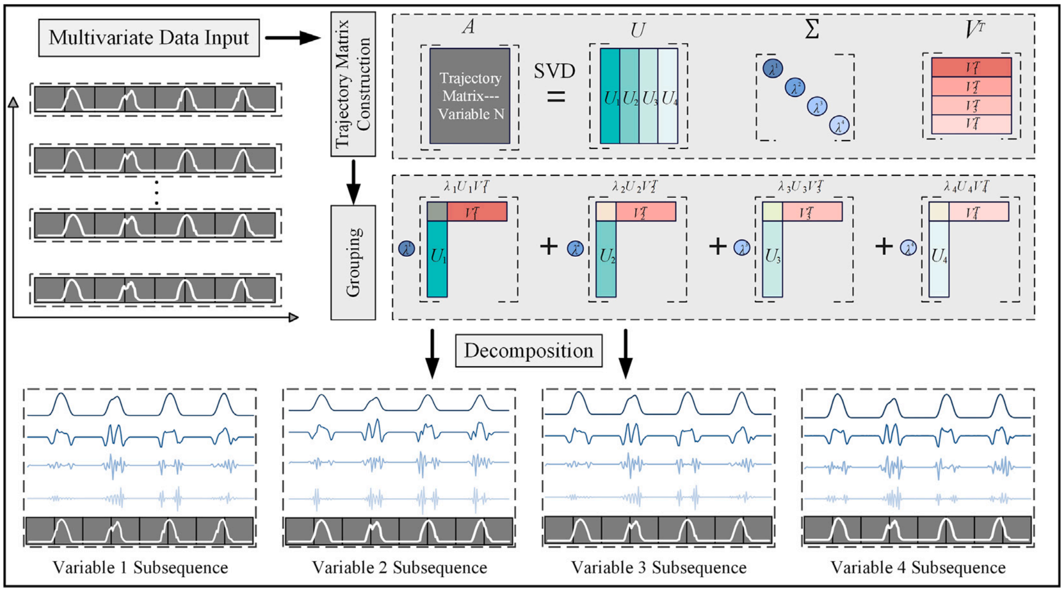

Fig. 2. Multivariate singular spectrum analysis decomposition process.

in Eq. (1). Here, L represents the embedding window size, and the dimensions of the embedding matrix are $(N - L + 1) \times L$. The second step involves applying SVD to the matrix X_i', resulting in three matrices U, Σ, and V^T. In this decomposition, U and V are the left and right singular vector matrices of X_i', respectively, which represent the principal components of the row space and the column space of X_i'. Σ is the diagonal matrix containing the singular values, as shown in Eq. (2), where σ_i denotes the singular values. The singular values are arranged in descending order, with larger singular values corresponding to the dominant signal components and smaller singular values often corresponding to noise. The third step consists of selecting the top k singular values for each variable and retaining the corresponding eigenvectors to preserve the primary information. To avoid redundancy and reduce computational complexity, k is fixed at 4 in the experimental setup used in this study. The fourth step involves signal reconstruction based on the selected singular value vectors, enabling the extraction of effective components from the denoised time series X_i^r, as shown in Eq. (3). After repeating this procedure for each variable subsequence, the results are aggregated to obtain the reconstructed data X^r, as shown in Eq. (4).

$$X_i' = \begin{bmatrix} x_{i1} & x_{i2} & \cdots & x_{iL} \\ x_{i2} & x_{i3} & \cdots & x_{iL+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{iN-L+1} & x_{iN-L+2} & \cdots & x_{iN} \end{bmatrix} \tag{1}$$

$$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_i) \tag{2}$$

$$X_i^r = U_k \Sigma_k V_k^T \tag{3}$$

$$X^r = [U_1\Sigma_1V_1^T, U_2\Sigma_2V_2^T, \ldots, U_p\Sigma_pV_p^T] \tag{4}$$

2.2. Multiscale covariate interaction module

To effectively leverage the multi-dimensional covariate information generated by MSSA and improve the predictive accuracy of the target variable, this study proposes a time series modeling approach based on multi-scale covariate interaction. In contrast to most existing research that focuses on increasing model complexity to improve temporal feature extraction, this method emphasizes a lightweight model design that aims to achieve strong generalization with limited parameters.

Numerous studies have demonstrated that simple linear methods exhibit promising performance in time series forecasting tasks [29–31]. Due to their lightweight nature, these methods are particularly well-suited for real-time scheduling applications. Compared to attention-based methods, these models are more effective in reducing overfitting risk. Consequently, this research omits the commonly used attention mechanism in favor of an architecture based on multi-layer perceptrons and one-dimensional convolutional neural networks. The proposed MCI framework consists of three main components: the temporal mixing module (TMM), the feature mixing module (FMM), and the multi-scale convolution module (MSCM). These modules are connected using residual connections to preserve transmitted information, enhancing model stability and improving forecasting accuracy, as illustrated in Fig. 3.

TMM captures global dynamic dependencies across time steps by applying an MLP with shared weights along the temporal dimension. Specifically, the input data $X_{\text{in}} \in \mathbb{R}^{B\times C\times L}$ is first transposed to obtain $X_{\text{in}}^T \in \mathbb{R}^{B\times L\times C}$, then projected into a higher-dimensional feature space via a fully connected layer, and passed through a nonlinear transformation using the ReLU activation function. The transformed features are then mapped back to the original sequence length, resulting in the output of TMM, as shown in Eq. (5). Similarly, the FMM operates along the variable dimension and is designed to capture the intricate interactions between the multi-dimensional covariates derived from MSSA and the target variable, thereby extracting essential information for precise forecasting of the target variable, as shown in Eq. (6). To further improve the model's capacity for multi-scale cross-variable modeling, a parallel MSCM is introduced, which simultaneously extracts local features and global patterns in time series data. Specifically, the input data is processed through two convolution operations, Conv_1 and Conv_2, each with different receptive fields, to generate feature maps X_1 and X_2. Conv_1 uses a 1×1 convolutional kernel to capture short-term fluctuations in the time series, such as periodic components, while Conv_2 employs a larger 3×3 convolutional kernel to extract long-term global trends, thus strengthening its ability to model long-range dependencies. Subsequently, X_1 and X_2 interact across different scales via Hadamard product-based fusion.

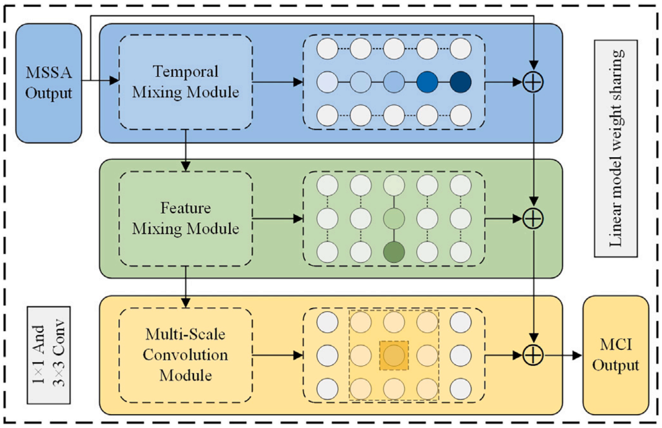

Fig. 3. The workflow of multi-scale covariate interaction module.

Finally, the integrated features are further refined through a third convolution operation, Conv_3, producing the final output features $X_{\mathrm{ICB}}^{\mathrm{out}}$, as shown in Eq. (7).

$$\mathbf{X}_{\mathrm{temp}}^{\mathrm{out}} = \mathrm{Dropout}\left(\mathrm{Linear}_2\left(\mathrm{ReLU}\left(\mathrm{Linear}_1\left(\mathbf{X}_{\mathrm{in}}^{T}\right)\right)\right)\right) + \mathbf{X}_{\mathrm{in}} \tag{5}$$

$$\mathbf{X}_{\mathrm{temp}}^{\mathrm{out}} = \mathrm{Dropout}\left(\mathrm{Linear}_2\left(\mathrm{ReLU}\left(\mathrm{Linear}_1\left(\mathbf{X}_{\mathrm{in}}^{T}\right)\right)\right)\right) + \mathbf{X}_{\mathrm{in}} \tag{6}$$

$$\mathbf{X}_{\mathrm{ICB}}^{\mathrm{out}} = \mathrm{Conv}_3\left(\mathbf{X}_1 \odot \mathbf{X}_2\right) + \mathbf{X}_{\mathrm{channel}}^{\mathrm{out}} \tag{7}$$

2.3. Multi-source domain collaborative learning framework with global knowledge guidance and transfer learning fine-tuning

This section introduces a novel multi-source domain collaborative learning framework, termed MDCTL-MCI, which is built upon the MCI framework. The proposed approach is designed to improve the generalization capability of the model, increase robustness against low-quality data, and address performance degradation caused by data heterogeneity across domains. MDCTL-MCI comprises three main components: (1) multi-source domain collaborative learning, (2) transfer learning fine-tuning guided by global knowledge, and (3) optimization using automatic mixed precision (AMP). These components are depicted in Fig. 4.

2.3.1. Multi-source domain collaborative learning

Due to substantial differences in data distribution, data quality, and patterns of missing data across PV sites, the predictive performance of MCI exhibits significant variability across different PV sites. Relying solely on locally trained models often fails to yield satisfactory performance. To address these issues and enhance model generalization, a multi-source domain collaborative learning strategy is proposed. By sharing model parameters across different PV sites and enabling knowledge fusion at the feature level, this approach facilitates information complementarity in the presence of missing data, thereby enhancing the availability of richer and more reliable auxiliary information for low-performing PV sites. Furthermore, a weighted averaging strategy is employed to aggregate model parameters from all PV sites during the global model update process. This allows the global model to leverage high-performing PV site models to compensate for the limitations of low-performing PV site models, ultimately improving their performance compared to models trained solely on local data.

The multi-source domain collaborative learning strategy is composed of five key stages: model initialization, local training, parameter uploading, global parameter aggregation, and parameter broadcasting. In each communication round, each local PV site k trains its local model θ_k independently using its own dataset D_k. The local loss function is defined in Eq. (8), where $f_\theta(x_i)$ and y_i denote the predicted and true values of sample i, respectively, and $l(\cdot)$ represents the loss function applied to each individual sample. The local model updates its parameters via gradient descent, and the updated weights θ_k^t are then transmitted to the central server. The server applies a weighted averaging strategy to compute the new global model parameters as defined in Eq. (9), where K denotes the total number of PV sites, and the weighting factor reflects the contribution of each PV site to the global update. After aggregation, the server broadcasts the updated global model θ^{t+1} to each local PV site for the next training round. Through multiple communication

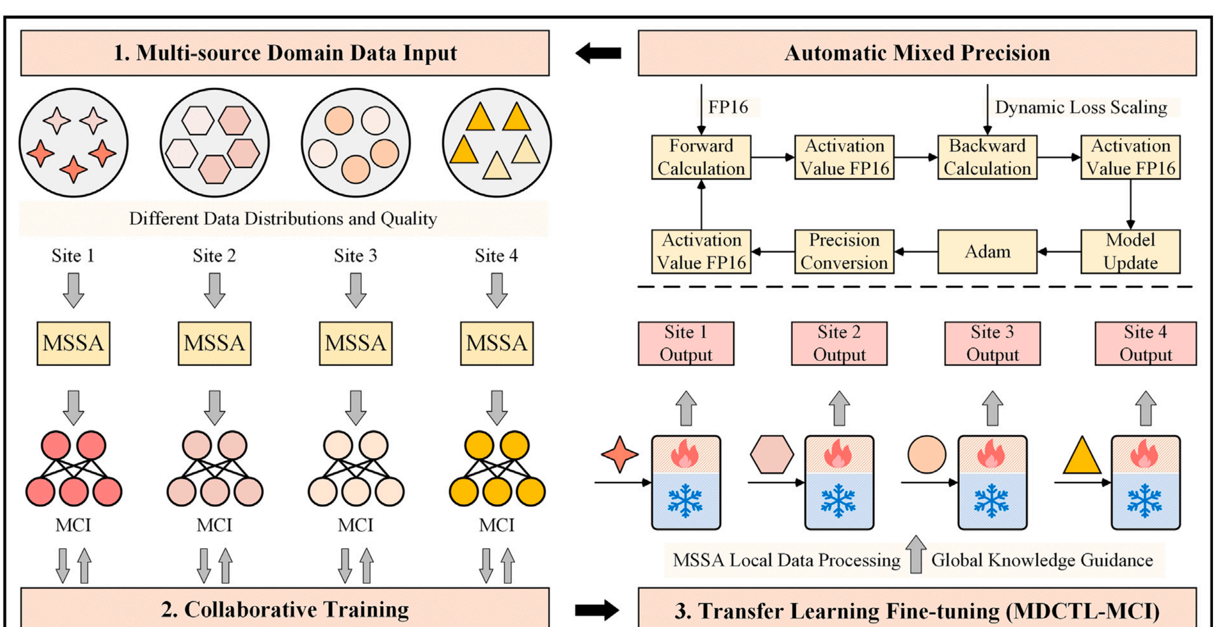

Fig. 4. The diagram of multi-source domain collaborative learning framework with global knowledge guidance and transfer learning fine-tuning.

rounds and iterative updates, a globally consistent model is eventually obtained.

$$L_k(\theta) = \frac{1}{|D_k|} \sum_{(x_i y_i) \in D_k} l\left(f_\theta\left(x_i\right), y_i\right) \tag{8}$$

$$\theta^{t+1} = \sum_{k=1}^{K} \frac{|D_k|}{\sum_{j=1}^{K} |D_k|} \theta_k^t \tag{9}$$

2.3.2. Transfer learning fine-tuning method with global knowledge guidance

The Multi-source domain collaborative learning strategy enhances overall generalization capabilities by integrating model information from multiple PV sites and improves the predictive performance of PV sites with low performance. However, during the global parameter aggregation process, interference from low-quality data and the heterogeneity of data distributions across PV sites may negatively impact high-performing PV sites, reducing the overall forecasting accuracy of the global model. To address this issue, we propose a globally guided transfer learning fine-tuning method. Specifically, MSSA is employed to preprocess the fine-tuning data, effectively retaining the main trend components while suppressing noise interference. A few-shot learning paradigm is adopted, in which the global model functions as a source of knowledge, and fine-tuning is performed using only a small set of high-quality MSSA-processed samples.

To fully leverage global knowledge and prevent overfitting, all parameters except those in the output layer are kept frozen during the fine-tuning stage, and only the weights and biases of the output layer are optimized using the Adam optimizer with a small learning rate. The output layer receives the time-series hidden representations generated by the backbone network as input. It then applies a fully connected mapping along the temporal dimension, transforming the sequence length into the prediction horizon and ensuring temporal and dimensional alignment with the target labels. This design preserves globally shared features while re-establishing the mapping between global representations and site-specific statistical characteristics using local data, enhancing adaptability to local distributions and mitigating the adverse effects of data heterogeneity. The update process of the fine-tuned local model at PV site k is formulated in Eq. (10), where θ^{frozen} are global model parameters kept frozen to retain shared knowledge, and θ^{out} are trainable output layer parameters adapted to local data.

$$\theta^*_{k,\text{out}} = \arg\min_{\theta^{\text{out}}} \frac{1}{|D_k|} \sum_{(x_i, y_i) \in D_k} l\left(f(x_i; \theta^{\text{frozen}}, \theta^{\text{out}}), y_i\right) \tag{10}$$

2.3.3. Automatic mixed precision

To reduce the computational overhead associated with multi-source domain collaborative learning, Automatic Mixed Precision (AMP) is integrated into local training. AMP aims to enhance model deployment flexibility and training efficiency by accelerating computation and decreasing GPU memory consumption while preserving model accuracy. During the forward propagation phase, AMP utilizes half-precision floating-point numbers (float16) for tensor computations, substantially reducing memory usage and improving computational throughput. In the backpropagation phase, to mitigate gradient underflow caused by the limited numerical precision of float16, a dynamic loss scaling mechanism is introduced. This mechanism amplifies the loss value adaptively to ensure the numerical stability of gradients within the representable range of float16.

3. Case study

3.1. Dataset

This study employs the dataset released by Chen et al. [32], which contains operational data from multiple solar PV and wind power plants. Specifically, one year of continuous operational data from four solar PV stations, recorded at a temporal resolution of 30 min, is selected. The data are partitioned into training, validation, and test sets in a 6:1:1 ratio. The selected PV stations vary considerably in both technical specifications and geographic location, with rated output capacities ranging from 30 MW to 130 MW. These stations are located in the northern, central, and northwestern regions of China, spanning diverse terrain types, including deserts, mountains, and plains.

The dataset exhibits significant data quality issues. While the PV power output data is relatively complete, covariates such as solar irradiance and weather conditions exhibit missing rates ranging from 0 % to 80 % across different stations. Consequently, Chen et al. [32] provided the original dataset with missing values and an anomaly-corrected version. In this study, the anomaly-corrected version of the dataset is adopted for subsequent experiments.

Given the critical role of covariate types in determining model accuracy, both Pearson correlation analysis (for linear relationships) and Spearman correlation analysis (for nonlinear relationships) are conducted on six variables. Global horizontal irradiance (GHI), direct normal irradiance (DNI), and total solar irradiance (TSI), which show the strongest correlation with PV power output, are selected as input variables for subsequent experiments. To better understand the data distribution, marginal histograms are plotted to depict the relationship between each selected variable and PV power output, as shown in Fig. 5.

3.2. Experimental setup and evaluation

This study employs Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Variation (CV) as evaluation metrics to comprehensively assess the model's predictive accuracy and stability, as defined in Eqs. (11)–(13). Specifically, y_i denotes the actual value of the i-th sample, $\hat{y}_i$ represents the predicted value of the i-th sample, $\bar{y}$ denotes the sample mean, σ represents the standard deviation of forecasting error, and μ denotes the mean forecasting error. MAE provides a direct measure of the average forecasting error, while MSE penalizes larger errors more heavily, emphasizing significant deviations and making it more sensitive to outliers. CV is used to quantify the degree of dispersion of the error relative to the average error, thereby reflecting the sensitivity of forecasting error under different levels of covariate missingness.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|. \tag{11}$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \tag{12}$$

$$\text{CV} = \frac{\sigma}{\mu}. \tag{13}$$

To comprehensively evaluate the effectiveness of the proposed method, seven cutting-edge time series forecasting methods are selected for comparative evaluation. These include attention-based models such as Pyraformer [33], Transformer [34], Informer [35], TimeXer [36], iTransformer [37], and PatchTST [38], and MLP-based models such as LightTS [39], TSMixer [40], and MCI. The implementations of all baseline models can be obtained from Wu et al. [41].

A complete experimental framework is constructed. To ensure the rigor and fairness of the experiments, all methods are configured with identical hyperparameters across all settings. Each model takes 48 historical time steps as input and performs multi-step forecasting for the next 48 time steps in a single forward pass. The experiments are designed around the following aspects to thoroughly evaluate the performance of the proposed method: (1) Comparison of the forecasting performance between single target variable input and joint covariate input in evaluating the impact of covariate inclusion on forecasting accuracy. (2) Investigation of the impact of covariate quality on forecasting accuracy and robustness. The various forecasting tasks are illustrated in Fig. 6. (3)

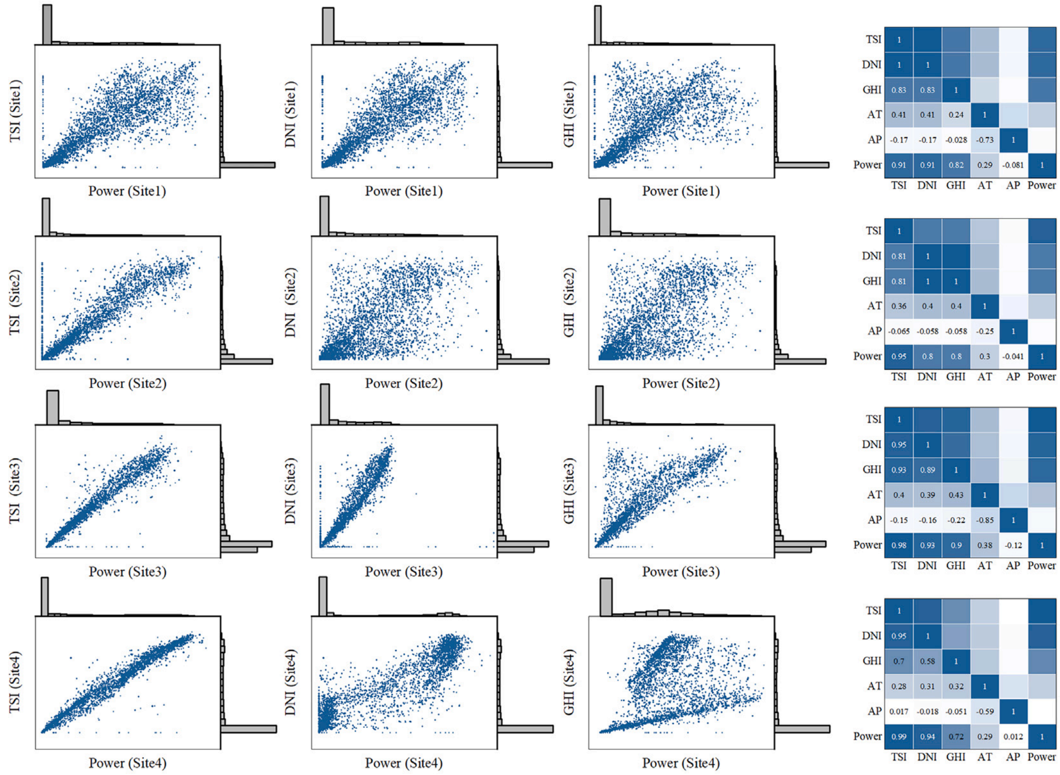

Fig. 5. The marginal distribution histograms of covariates and target variable, along with the Pearson correlation matrix.

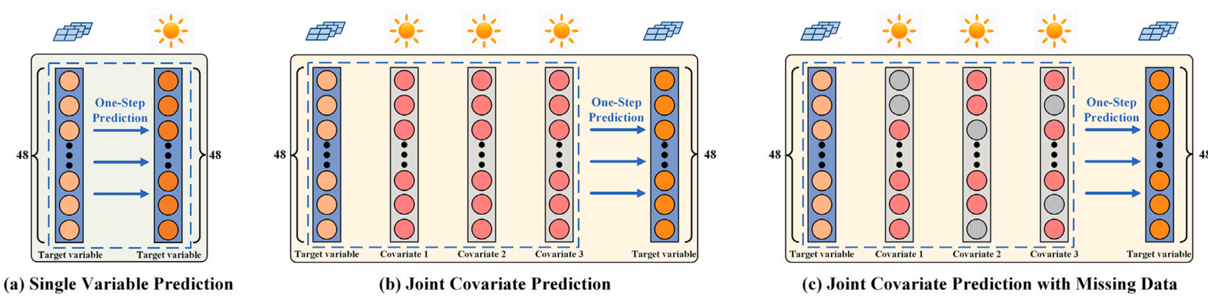

Fig. 6. Diagram of three forecasting tasks: (a) single target variable input; (b) joint covariate input; (c) joint covariate input with missing data.

Analysis of model interpretability using SHAP and elucidation of the factors contributing to its robust forecasting performance. (4) Evaluation of the computational efficiency of the proposed model and ablation studies to assess the contribution of individual components.

4. Experimental results analysis

4.1. *The impact of covariate inclusion on forecasting performance*

This section investigates the impact of covariate information on the forecasting accuracy of PV power output. The experiment compares the forecasting results using single-target variable input versus joint covariate input. For MCI and MDCTL-MCI, single-variable forecasts are obtained by freezing the MSSA signal decomposition module. To ensure the reliability of the results and minimize the influence of random variations, the experiment is repeated five times using different random seeds. The final results are averaged across all repetitions, as presented in Table 2. Fig. 7 presents the average values of MAE and MSE for both single-target variable input and joint covariate input, as well as the change in forecasting accuracy resulting from covariate inclusion, as shown in Eq. (14). In this equation, E_{multi} represents the results from joint covariate input, while E_{single} denotes the results from single-target variable input. $\Delta P > 0$ indicates that the inclusion of covariates leads to improved forecasting accuracy, as highlighted in red. Additionally, Fig. 7 illustrates the forecasting curves for the four stations under covariate inclusion.

In addition, we assessed the computational efficiency of the proposed framework for real-time PV power forecasting. Experimental results show that the lightweight model and AMP training strategy proposed in this study have competitive performance in terms of parameter count, model complexity, iteration time, and memory consumption. This advantage enables the proposed method to meet the real-time demands of PV systems. The results can be found in the appendix.

$$\Delta P = \frac{E_{\text{singal}} - E_{\text{multi}}}{E_{\text{singal}}} \times 100\ \% \tag{14}$$

The experimental results show that significant variations exist in the performance of different models after covariate inclusion, and the models' ability to utilize covariate information is also influenced by the

Table 2
Prediction results of different methods under single target variable input and joint covariate input.

SITE1	Univariate		Covariate		SITE3	Univariate		Covariate	
	MSE	MAE	MSE	MAE		MSE	MAE	MSE	MAE
MCI	0.1290	0.2001	0.1219	0.1884	MCI	0.2406	0.2773	0.2172	0.2610
MDCTL-MCI	0.1310	0.2053	0.1183	0.1914	MDCTL-MCI	0.2413	0.2761	0.2169	0.2605
LightTS	0.1450	0.2193	0.1310	0.1960	LightTS	0.2412	0.2719	0.2435	0.2743
TSMixer	0.1285	0.1976	0.1251	0.1939	TSMixer	0.2417	0.2781	0.2305	0.2734
TimeXer	0.1438	0.2082	0.1449	0.2088	TimeXer	0.2797	0.2631	0.2824	0.2676
PatchTST	0.1466	0.2124	0.1489	0.2138	PatchTST	0.2869	0.2755	0.2871	0.2764
Pyraformer	0.1559	0.2338	0.1484	0.2256	Pyraformer	0.2658	0.2986	0.2601	0.3013
iTransformer	0.1401	0.1974	0.1415	0.1983	iTransformer	0.2814	0.2686	0.2861	0.2716
Informer	0.1455	0.2096	0.1495	0.2094	Informer	0.2572	0.2794	0.2604	0.2805
Transformer	0.1460	0.2071	0.1516	0.2046	Transformer	0.2577	0.2672	0.2554	0.2625
SITE2	Univariate		Covariate		SITE4	Univariate		Covariate	
	MSE	MAE	MSE	MAE		MSE	MAE	MSE	MAE
MCI	0.2493	0.2812	0.2315	0.2724	MCI	0.1353	0.2079	0.1321	0.1999
MDCTL-MCI	0.2481	0.2785	0.2339	0.2679	MDCTL-MCI	0.1326	0.2001	0.1264	0.1843
LightTS	0.2521	0.2765	0.2480	0.2710	LightTS	0.1440	0.2193	0.1398	0.2030
TSMixer	0.2495	0.2814	0.2468	0.2811	TSMixer	0.1436	0.2226	0.1406	0.2169
TimeXer	0.3025	0.2751	0.3015	0.2768	TimeXer	0.1423	0.2003	0.1450	0.2075
PatchTST	0.3037	0.2854	0.3047	0.2818	PatchTST	0.1433	0.2057	0.1531	0.2168
Pyraformer	0.2720	0.3022	0.2610	0.2972	Pyraformer	0.1604	0.2231	0.1467	0.2065
iTransformer	0.3105	0.2865	0.3118	0.2835	iTransformer	0.1441	0.1975	0.1474	0.2015
Informer	0.2733	0.2895	0.2809	0.2903	Informer	0.1513	0.2196	0.1470	0.2146
Transformer	0.2757	0.2825	0.2864	0.2813	Transformer	0.1467	0.2242	0.1517	0.2214

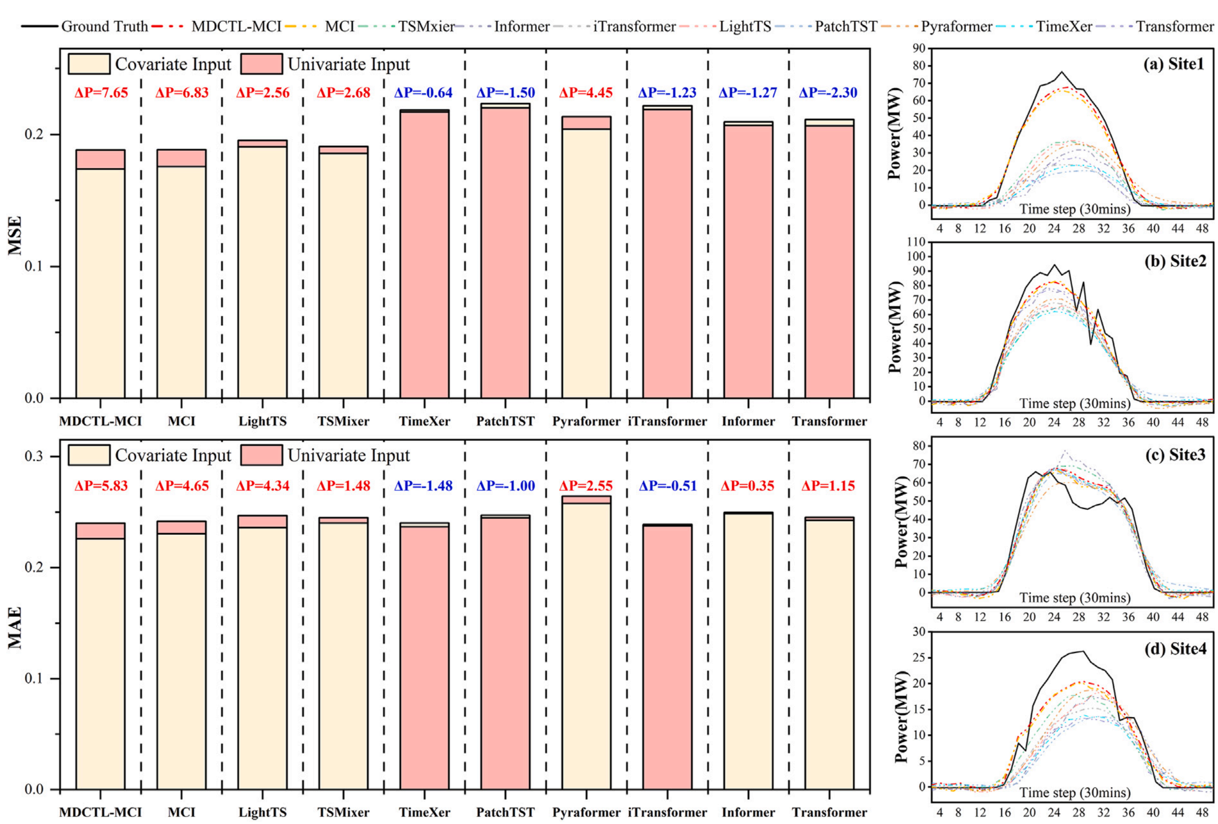

Fig. 7. Impact of covariate inclusion on MAE, MSE, and forecasting result curves across four stations after covariate inclusion.

characteristics of the dataset. Specifically, MCI and MDCTL-MCI consistently achieved notable improvements in forecasting accuracy across different PV sites. This indicates that the covariate interaction module proposed in this study effectively leverages covariate information, enhancing the model's predictive capability for the target variable. Among the baseline methods, the MLP-based approach demonstrated an overall improvement in forecasting accuracy after covariate inclusion. In contrast, attention-based methods exhibited either a decrease in accuracy or no significant change, with their performance generally lower than that of the MLP-based approach. This finding aligns with the conclusions drawn in [40,42]. Notably, certain attention-based methods, such as Pyraformer, exhibited improved forecasting accuracy after incorporating covariates. This enhancement may be attributed to its unique multi-scale attention mechanism, which effectively integrates covariate

Table 3
Prediction results of different methods under varying covariate quality conditions.

SITE1	RATE0		RATE0.1		RATE0.3		RATE0.5		RATE0.7		RATE0.9		CV		RANK	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MCI	0.1199	0.1899	0.1229	0.1955	0.1275	0.2039	0.1421	0.2154	0.1328	0.2119	0.1258	0.2012	5.6554	4.3352	3	5
MDCTL-MCI	0.1193	0.1922	0.1196	0.1938	0.1230	0.1981	0.1248	0.1994	0.1253	0.2032	0.1270	0.2021	2.3232	2.0252	1	1
LightTS	0.1313	0.1969	0.1310	0.1972	0.1316	0.1983	0.1318	0.1988	0.1318	0.1987	0.1321	0.1991	0.2788	0.4229	4	2
TSMixer	0.1254	0.1942	0.1253	0.1965	0.1273	0.2003	0.1281	0.2019	0.1296	0.2047	0.1307	0.2023	1.5529	1.7832	2	4
PatchTST	0.1511	0.2179	0.1565	0.2228	0.1557	0.2237	0.1496	0.2104	0.1505	0.2145	0.1500	0.2109	1.8324	2.4272	7	7
Pyraformer	0.1510	0.2337	0.1666	0.2468	0.1527	0.2341	0.1574	0.2460	0.1671	0.2508	0.1581	0.2429	3.9011	2.6407	8	8
iTransformer	0.1411	0.1974	0.1407	0.1972	0.1402	0.1970	0.1421	0.1981	0.1425	0.1991	0.1448	0.2007	1.0669	0.6462	5	3
Informer	0.1518	0.2115	0.1688	0.2221	0.2127	0.2488	0.2205	0.2460	0.2815	0.2983	0.2718	0.2791	21.986	12.013	9	9
Transformer	0.1521	0.2043	0.1803	0.2201	0.1953	0.2293	0.2434	0.2789	0.3963	0.4054	0.2508	0.2785	33.569	24.869	10	10
TimeXer	0.1449	0.2074	0.1443	0.2064	0.1456	0.2100	0.1442	0.2064	0.1441	0.2076	0.1443	0.2074	0.3641	0.5737	6	6
SITE2	**RATE0**		**RATE0.1**		**RATE0.3**		**RATE0.5**		**RATE0.7**		**RATE0.9**		**CV**		**RANK**	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MCI	0.2305	0.2716	0.2530	0.2884	0.2407	0.2908	0.2559	0.2950	0.2617	0.2905	0.2542	0.2853	4.2135	2.7216	2	6
MDCTL-MCI	0.2349	0.2670	0.2350	0.2683	0.2341	0.2711	0.2365	0.2723	0.2354	0.2684	0.2371	0.2676	0.4357	0.6742	1	1
LightTS	0.2457	0.2738	0.2471	0.2739	0.2500	0.2750	0.2496	0.2762	0.2506	0.2758	0.2556	0.3012	1.2494	3.5226	3	3
TSMixer	0.2467	0.2803	0.2449	0.2817	0.2457	0.2821	0.2491	0.2858	0.2550	0.2824	0.2580	0.2817	1.9701	0.5980	4	4
PatchTST	0.3050	0.2817	0.3104	0.2865	0.3086	0.2890	0.3055	0.2841	0.3069	0.2850	0.3131	0.2925	0.9191	1.2254	6	5
Pyraformer	0.2623	0.2999	0.2690	0.3038	0.2792	0.3212	0.3309	0.3535	0.4072	0.4228	0.3971	0.4119	18.309	14.001	10	10
iTransformer	0.3129	0.2876	0.3105	0.2839	0.3102	0.2864	0.3112	0.2893	0.3099	0.2863	0.3106	0.2888	0.3147	0.6178	7	7
Informer	0.2798	0.2890	0.2936	0.3020	0.3016	0.2958	0.3203	0.2988	0.3146	0.2953	0.3644	0.3183	8.5703	3.0487	8	8
Transformer	0.2942	0.2752	0.2918	0.3031	0.3247	0.3016	0.3202	0.3178	0.3122	0.3231	0.3470	0.3510	5.9704	7.4211	9	9
TimeXer	0.3006	0.2758	0.3011	0.2777	0.3013	0.2793	0.3025	0.2788	0.2988	0.2769	0.3021	0.2793	0.3889	0.4626	5	2
SITE3	**RATE0**		**RATE0.1**		**RATE0.3**		**RATE0.5**		**RATE0.7**		**RATE0.9**		**CV**		**RANK**	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MCI	0.2161	0.2610	0.2171	0.2684	0.2275	0.2736	0.2407	0.2845	0.2457	0.2926	0.2400	0.2790	5.0483	3.7496	2	5
MDCTL-MCI	0.2172	0.2592	0.2163	0.2631	0.2180	0.2693	0.2214	0.2701	0.2237	0.2691	0.2275	0.2654	1.7915	1.4797	1	1
LightTS	0.2407	0.2721	0.2397	0.2709	0.2407	0.2731	0.2482	0.2866	0.2436	0.2883	0.2487	0.3063	1.4938	4.4366	4	7
TSMixer	0.2295	0.2721	0.2278	0.2712	0.2324	0.2735	0.2400	0.2724	0.2448	0.2724	0.2754	0.2836	6.7026	1.5538	3	4
PatchTST	0.2850	0.2722	0.2944	0.2800	0.2959	0.2826	0.2894	0.2745	0.2912	0.2777	0.2927	0.2757	1.2246	1.2415	8	6
Pyraformer	0.2544	0.2979	0.2621	0.3023	0.2655	0.3080	0.2630	0.3075	0.2778	0.3144	0.2921	0.3266	4.6018	2.9812	5	10
iTransformer	0.2857	0.2719	0.2843	0.2717	0.2872	0.2715	0.2859	0.2747	0.2860	0.2735	0.2846	0.2721	0.3423	0.4209	7	3
Informer	0.2604	0.2771	0.2779	0.2967	0.2882	0.3147	0.3295	0.2975	0.3844	0.3120	0.3570	0.3082	14.050	4.2085	10	9
Transformer	0.2556	0.2635	0.2655	0.2753	0.2666	0.2806	0.2963	0.2910	0.3645	0.3067	0.3049	0.2976	12.582	5.0182	9	8
TimeXer	0.2820	0.2674	0.2830	0.2695	0.2830	0.2655	0.2864	0.2708	0.2858	0.2703	0.2850	0.2683	0.5683	0.6754	6	2
SITE4	**RATE0**		**RATE0.1**		**RATE0.3**		**RATE0.5**		**RATE0.7**		**RATE0.9**		**CV**		**RANK**	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MCI	0.1317	0.1989	0.1357	0.2079	0.1380	0.2147	0.1360	0.2066	0.1354	0.2160	0.1353	0.2125	1.3777	2.7688	2	3
MDCTL-MCI	0.1274	0.1889	0.1292	0.1927	0.1281	0.1966	0.1295	0.2037	0.1271	0.2031	0.1283	0.2044	0.6933	2.9954	1	1
LightTS	0.1443	0.2118	0.1452	0.2143	0.1431	0.2067	0.1437	0.2151	0.1435	0.2099	0.1435	0.2118	0.4907	1.3256	3	4
TSMixer	0.1414	0.2224	0.1437	0.2258	0.1425	0.2240	0.1435	0.2284	0.1460	0.2438	0.1492	0.2513	1.7864	4.7006	4	7
PatchTST	0.1516	0.2149	0.1491	0.2105	0.1547	0.2192	0.1493	0.2091	0.1541	0.2205	0.1529	0.2163	1.4317	1.9346	7	5
Pyraformer	0.1527	0.2149	0.1605	0.2262	0.1867	0.2507	0.1825	0.2500	0.1648	0.2256	0.1666	0.2376	7.0738	5.6311	8	8
iTransformer	0.1378	0.1886	0.1396	0.1903	0.1398	0.1928	0.1526	0.2088	0.1537	0.2073	0.1527	0.2050	4.8061	4.2174	5	2
Informer	0.1474	0.2105	0.1594	0.2209	0.1675	0.2284	0.1948	0.2452	0.2077	0.2540	0.2301	0.2891	15.674	10.686	9	9
Transformer	0.1403	0.2109	0.1598	0.2245	0.2091	0.3018	0.1835	0.2846	0.2445	0.3230	0.3295	0.3375	29.674	16.922	10	10
TimeXer	0.1488	0.2177	0.1544	0.2217	0.1495	0.2123	0.1507	0.2175	0.1499	0.2175	0.1518	0.2151	1.2166	1.3201	6	6

information across various time scales. The remaining baseline methods did not exhibit consistent performance improvements across different PV sites.

4.2. *The impact of covariate quality on forecasting performance*

In practical applications, covariates are often affected by missing values and outliers, which can degrade model performance. Therefore, this section further investigates the impact of covariate quality on forecasting accuracy and robustness. In the experiment, a fixed random seed is employed, and the covariate missing rates are set to 0 %, 10 %, 30 %, 50 %, 70 %, and 90 % while preserving the completeness of the target variable in order to simulate realistic data conditions. The experiment is analyzed in terms of forecasting accuracy and forecasting stability, with the results presented in Table 3.

4.2.1. *Accuracy analysis*

As shown in Table 3, although model performance varies across PV sites, the MDCTL-MCI introduced in this study consistently achieves the highest average forecasting accuracy. Specifically, compared to the average forecasting accuracy of the baseline methods under conditions of complete covariate information, the MAE and MSE of MDCTL-MCI are reduced by 6.51 % and 14.5 %. Under varying levels of covariate missing rates, MDCTL-MCI exhibits a more pronounced advantage, with MAE and MSE showing reductions of 9.7 % and 20.9 %. Among the baseline methods, attention-based models generally exhibit lower accuracy than MLP-based models and are more sensitive to variations in covariate quality. Specifically, LightTS outperforms the other baseline models in forecasting performance, while TimeXer exhibits comparatively superior accuracy among attention-based methods. However, compared to MDCTL-MCI, the average MAE and MSE of LightTS increase by 4.16 % and 7.97 %, respectively, whereas TimeXer shows increases of

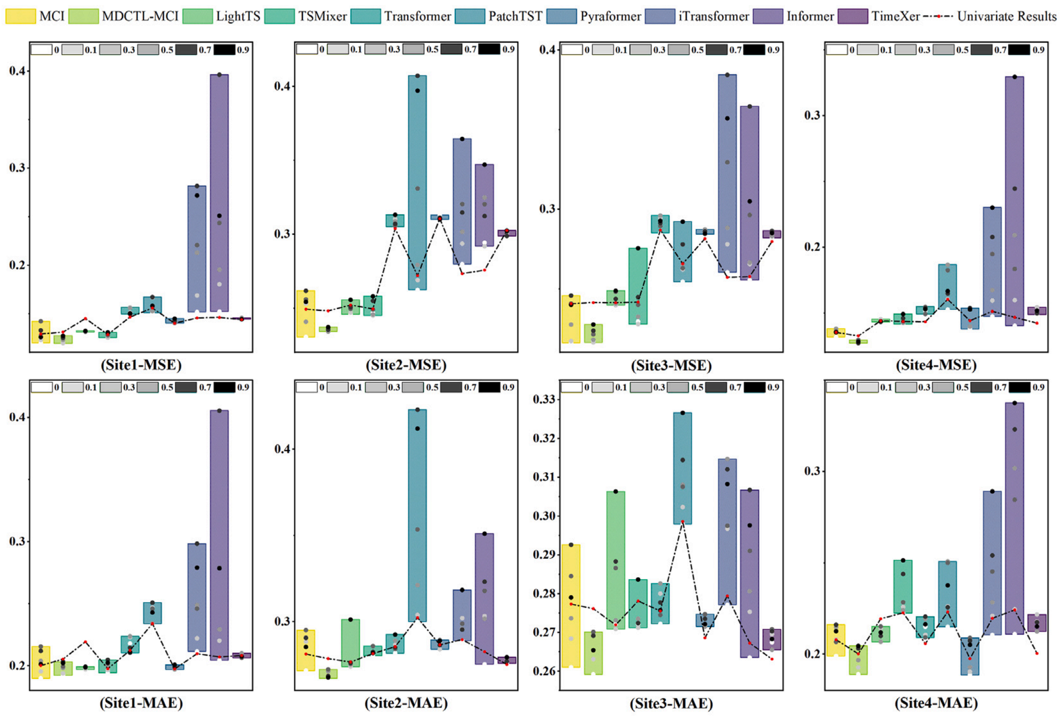

Fig. 8. Error distribution boxplots of different methods under varying covariate quality conditions.

4.07 % and 19.7 %. Further analysis reveals that MCI achieves forecasting accuracy comparable to MDCTL-MCI under conditions of complete covariate information, which can be explained by its strong ability to utilize covariate data, thereby enhancing forecasting accuracy. However, as the covariate missing rate increases, MCI fails to maintain stable forecasting accuracy, leading to a significant increase in forecasting errors.

4.2.2. Robustness analysis

Fig. 8 shows boxplots of the error distributions for each method under different covariate quality conditions. The height of each box reflects the distribution range of forecasting errors and indirectly indicates the magnitude of the CV metric. The color of the points within the boxes indicates the forecasting results corresponding to varying covariate missing rates. Additionally, a line chart depicting the forecasting performance under single-target variable input is included as a reference curve for comparison. From Fig. 8, it can be observed that MDCTL-MCI maintains relatively small box ranges across all PV sites, with the boxes generally positioned below the baseline curve, indicating that its forecasting performance remains stable under most covariate missing rate conditions and consistently outperforms the single-target input approach. Among the baseline methods, TimeXer and PatchTST also demonstrate relatively stable performance under high covariate missing rates. The stability of PatchTST can be attributed to its channel-independent mechanism, which partially mitigates the impact of missing covariate information on the target variable. However, this mechanism also limits inter-variable interactions, which may hinder further improvements in forecasting accuracy. TimeXer effectively extracts covariate information by leveraging its distinctive covariate encoding and cross-attention mechanism between internal and external variables, thereby exhibiting strong robustness. However, its accuracy is still lower than that of MDCTL-MCI. In contrast, Pyraformer, Transformer, and Informer exhibit the greatest variability in error, with most of their boxplots located above the baseline curve. This suggests that low-quality covariate information negatively impacts most Transformer-based models, resulting in forecasting performance that falls below the single-target input baseline. MCI, TSMixer, and LightTS exhibit robust performance across the majority of PV sites, with the exception of Site 3, where the MAE values show notably higher variability.

Fig. 9 illustrates the curves of observed and predicted values of each method under varying covariate quality conditions, using Site 3 as a representative example. The associated error bands are also depicted. Red upward arrows and gray downward arrows indicate improvements or deteriorations in forecasting performance resulting from incorporating covariate information. As shown in Fig. 9, the MDCTL-MCI consistently improves forecasting accuracy under all levels of covariate missingness at Site 3. In fact, only at Site 4 does the MAE metric exhibit a marginal increase compared to the single-target input setting when the missing rate reaches 70 % and 90 %. In contrast, the MCI demonstrates robustness when the covariate missing rate is below 30 %, and the inclusion of covariate information positively contributes to its predictive performance. Pyraformer yields accuracy improvement only under missing rates below 10 %, while the remaining methods demonstrate no performance improvement under any level of covariate missingness.

4.3. Interpretability analysis

This section employs Shapley additive explanation (SHAP) to analyze the model's interpretability. By generating heatmaps of feature importance at various historical steps, this approach reveals the contribution of each variable in both MCI and MDCTL-MCI models to forecasting outcomes, thereby explaining the model's accuracy and stability. SHAP is a model-agnostic interpretation method that calculates the marginal contribution of features to the model's output using Shapley values to construct an additive explanation model. The precise Shapley value ϕ_j, corresponding to feature j, is determined using the following Eq. (15), where X represents the set of input features, $S \subseteq X_j$ denotes a subset of features excluding feature j, $f(x_j)$ is the model's forecast for the input sample x_j, and $f_S(x_j)$ is the model's forecast for the input sample x_j with the features in S replaced by their expected values.

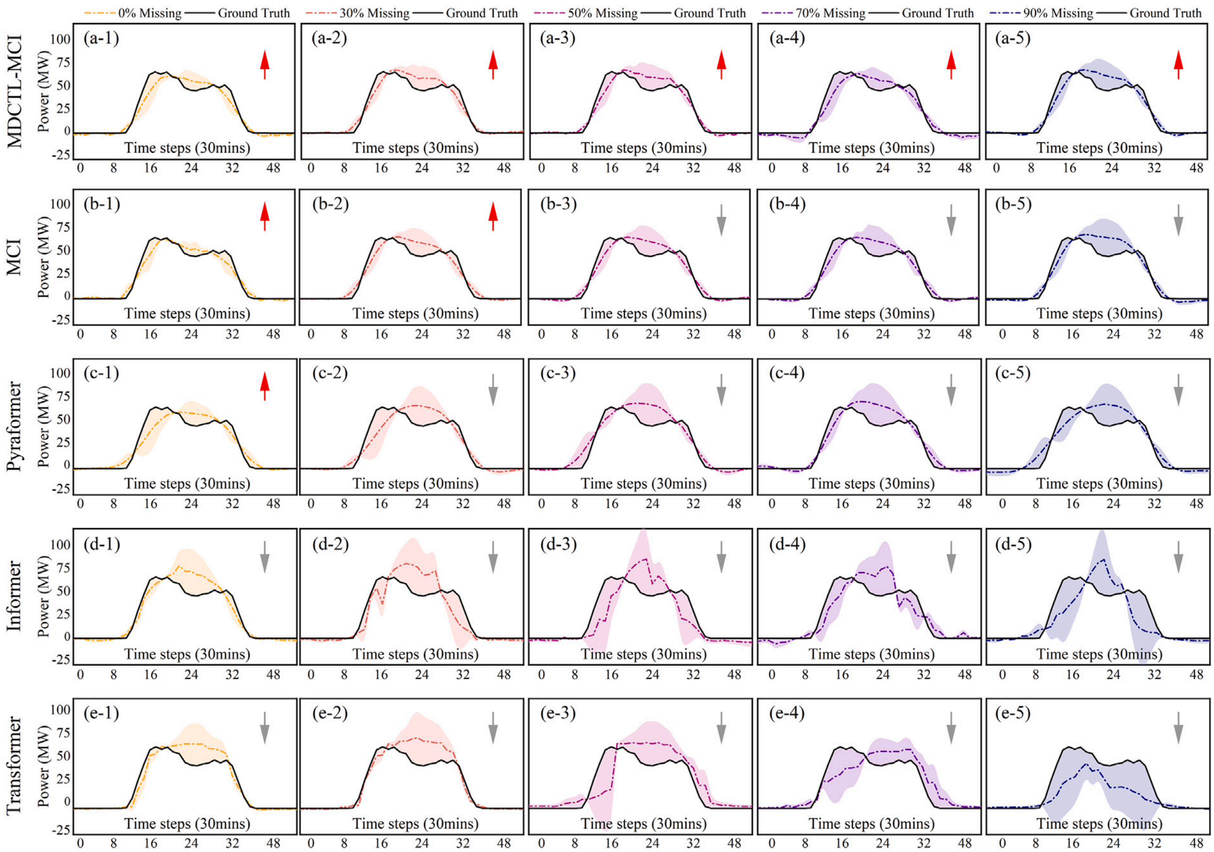

Fig. 9. Observed and predicted value curves of different methods under varying covariate quality conditions.

$$\phi_j = \sum_{S \subseteq X_j} \frac{|S|!(|X| - |S| - 1)!}{|X|!} \left[f(x_j) - f_S(x_j) \right] \tag{15}$$

In this study, one hundred time series samples were randomly selected from the test set to construct the feature heatmap. Owing to space constraints, Site3 was chosen as a case study, and two forecasting steps, T + 17 and T + 27, were randomly selected for detailed analysis. The results are shown in Fig. 10. The x-axis denotes the historical time steps, ranging from T-48 to T-1, while the y-axis displays each input feature and its corresponding sub-features, with POWER as the target variable and the remaining as covariates. SHAP values greater than zero are visualized in red, indicating a positive contribution to the forecasting, whereas values less than zero are shown in blue, signifying a negative influence.

From Fig. 10, it can be inferred that: (1) The target variable POWER has the most pronounced effect on the forecasting outcomes across all variables. The covariates GHI, DNI, and TSI also exhibit considerable effects. Regarding the sub-features obtained from MSSA decomposition, those associated with lower decomposition levels, such as sub-features F1 and F2, make more substantial contributions to the forecasting results. As the level of decomposition increases, the significance of the sub-features gradually diminishes. (2) The MCI and MDCTL-MCI models attribute varying levels of importance to variables across multiple historical time steps during the forecasting process, indicating their ability to capture both short-term dynamics and long-term dependencies. Further analysis reveals that for the forecasting horizons of T + 17 and T + 27, the models assign the highest weights to variables around T-30 and T-24, respectively. This suggests that the models prefer using historical data from the same time period on the previous day for forecasting. (3) In scenarios without covariate missingness, both MDCTL-MCI and MCI models effectively leverage variables at different historical time steps, enhancing forecasting performance. However, the SHAP values of sub-features in the MCI model are marginally higher than those in the MDCTL-MCI model. This finding accounts for the slightly inferior performance of MDCTL-MCI compared to MCI at certain sites under conditions without covariate missingness. (4) As the covariate missing rate increases, the MDCTL-MCI model becomes more reliant on the target variable, with the SHAP values of covariates gradually approaching 0, indicating that MDCTL-MCI can adaptively adjust feature importance to maintain stable forecasting performance. In contrast, the MCI model tends to rely on low-quality covariate data for forecasting, which fails to provide effective temporal information and introduces additional noise.

Based on the above analysis, the stable predictive performance of MDCTL-MCI can be explained by the following factors: First, the incorporation of the MSSA effectively suppresses noise introduced by low-quality data and extracts trend-related information from incomplete covariates. Second, the multi-source collaborative learning strategy alleviates performance degradation caused by data heterogeneity and improves the robustness of the global model against local data quality fluctuations. Third, the information exchange across PV sites allows the model to adaptively decrease its reliance on covariate features when the missing rate is high, and instead emphasize the use of the complete target variable for forecasting, thereby alleviating the negative effects of low-quality covariate information on predictive stability.

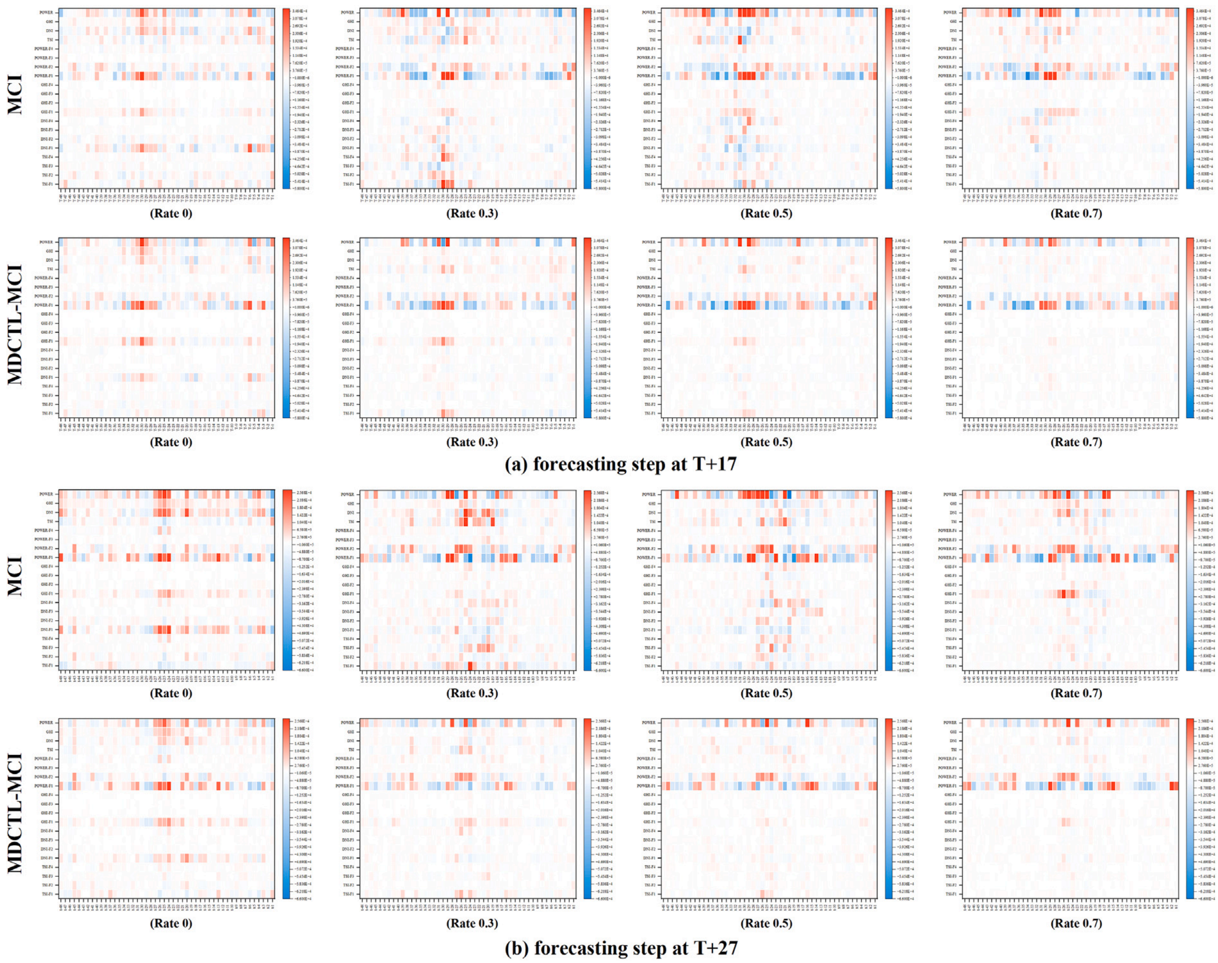

Fig. 10. Heatmap of feature importance across historical time steps under different covariate missing rates.

5. Conclusion

This study introduces a novel algorithm for PV power generation forecasting, referred to as MDCTL-MCI, which combines multi-scale feature interaction with a multi-source domain collaborative learning framework enhanced by transfer learning. It effectively addresses the challenges of performance degradation induced by low-quality data. Experimental results demonstrate that the proposed method exhibits dual advantages of forecasting accuracy and robustness, offering a quality-guaranteed PV power forecasting performance for real-time energy resource coordination. Further discussion of the experimental results is provided below.

Impact of utilizing covariates on forecasting performance: The contribution of covariates to forecasting performance largely depends on the characteristics of the dataset, the quality of the covariates, and the model's ability to leverage covariate information. Experimental results show that even in the absence of data quality issues, incorporating covariate information does not necessarily enhance the predictive accuracy of the target variable. In general, the MLP-based framework, along with multi-scale information interaction across variable dimensions, facilitates more effective utilization of covariate features and demonstrates superior performance to transformer-based methods in the application scenario investigated in this study. The underlying reasons can be summarized as follows:

(1) Models based on self-attention mechanisms excel at capturing global feature patterns, however, in this study's short-term photovoltaic forecasting task, most predictive signals are concentrated in recent local patterns, and therefore the advantage of global modeling is not fully realized.
(2) Transformer-based methods employ data-dependent weight modeling, making them more sensitive to noise and missing values in the inputs, which may result in high attention weights being assigned to irrelevant or erroneous time steps, thereby increasing the risk of overfitting.

It is worth noting that transformer-based methods, such as TimeXer, can still achieve strong predictive performance when the forecasting task involves pronounced long-range dependencies and appropriate data preprocessing and architectural adaptations are applied. Future research should focus not only on expanding input feature sets but also on designing effective modules for covariate feature extraction and integration to ensure the positive contribution of covariates to target variable forecasting performance. Otherwise, this could lead to degraded forecasting accuracy and increased computational costs. In contrast, the multi-scale covariate interaction method proposed in this study effectively exploits covariate information and enhances forecasting accuracy across different PV sites.

Forecasting quality under poor data quality: PV systems often face significant data quality issues, which pose challenges to the quality of forecasting results. Quality-guaranteed forecasting performance under low-quality data is crucial for practical applications. Experimental results show that the channel-independent mechanism, cross-attention mechanism for internal and external variables, and diversified data sampling mechanism all maintain error fluctuations within a narrow range under varying covariate missing rates. When combined with strategies such as signal decomposition and multi-scale feature extraction, model stability is further improved. However, these methods still exhibit relatively low overall accuracy and do not achieve an optimal trade-off between accuracy and stability. In contrast, the proposed MDCTL-MCI method optimizes information input through MSSA signal decomposition, reduces interference from low-quality data using multi-source domain collaborative learning, and improves forecasting accuracy through transfer learning fine-tuning. This approach consistently achieves stable forecasting performance under low data quality conditions without additional data preprocessing.

Interpretability study: This section employs Shapley additive explanation (SHAP) to analyze the model's interpretability. The following conclusions are drawn: the incorporation of the MSSA effectively suppresses noise introduced by low-quality data and extracts trend-related information from incomplete covariates. The multi-source collaborative learning strategy alleviates performance degradation caused by data heterogeneity and improves the robustness of the global model against local data quality fluctuations. The information exchange across PV sites allows the model to adaptively decrease its reliance on covariate features when the missing rate is high, and instead emphasizes the use of the complete target variable for forecasting, guaranteeing the forecasting performance of the proposed framework.

In summary, the MDCTL-MCI method proposed in this study effectively tackles the limitations of covariate underutilization and the instability and inaccuracy of forecasts under poor data quality conditions, which remain common in existing research. The proposed model establishes a solid foundation for the deployment of PV systems in complex environments and offers significant contributions to the development of PV technology.

Limitation

Although the method proposed in this study demonstrates significant advantages in PV power forecasting tasks, it still has several limitations that warrant further exploration. While the role of covariate information in forecasting accuracy has been explored in detail, the types of covariates considered in this study are relatively limited due to limitations in the diversity of available variables in the dataset. Future research should consider incorporating additional environmental factors that may influence PV power generation to provide valuable insights for future studies.

CRediT authorship contribution statement

Ke Yan: Writing – review & editing, Writing – original draft, Validation, Supervision. **Jian Liu:** Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Data curation. **Jiazhen Zhang:** Writing – review & editing, Project administration. **Fan Yang:** Writing – review & editing. **Yuan Gao:** Writing – review & editing, Validation, Supervision. **Yang Du:** Writing – review & editing.

Declaration of competing interest

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

Acknowledgement

This work is supported by National Science Foundation of China Excellent Young Overseas Talents Programme under Grant No. Z202304392531 and The Science and Technology Innovation Program (Overseas Expert Program) of Hunan Province under Grant No. 2024RC9006.

Appendix A. Computational efficiency analysis and ablation experiments

The computational efficiency of the algorithm directly affects the real-time energy scheduling of PV systems, which is an essential factor in ensuring a stable and continuous power output. In this section, experiments are conducted to evaluate the computational efficiency of each method for individual PV sites. To provide a comprehensive assessment of model performance, the *Torchinfo.summary* function is employed to output the architectural information of each model, including the number of parameters (Params), computational complexity (MACs), memory consumption (Memory), time per iteration (Train-Time), and forecasting accuracy measured by MSE.

Additionally, an ablation study is performed to decompose the components of the proposed model and evaluate the impact of each component on both forecasting accuracy and computational efficiency. Specifically, Ab-T denotes the time-mixing module, Ab-TF further incorporates the variable-dimensional transformation module, and Ab-TFM introduces the multi-scale convolution module. By incorporating MSSA into Ab-TFM, the whole architecture of the MCI is constructed. The experimental results are presented in Table A.1 and Fig. A.1.

In Fig. A.1, the x-axis represents the method names, and the y-axis shows the MSE. The size of the circles reflects the model's memory usage, while the color of the circles represents the number of model parameters. The blue bar chart represents the per-iteration training time, and the red line chart indicates the model's computational complexity. A comprehensive analysis of multiple evaluation metrics suggests that MCI exhibits competitive computational efficiency. It ranks just slightly below the MLP-based models TSMixer and LightTS in terms of parameter count, computational complexity, and memory consumption. However, it achieves more optimal performance with respect to MSE and MAE.

Ablation experiments demonstrate that the simple linear model Ab-T outperforms most comparison models. With the incorporation of the covariate interaction module and multi-scale convolution, forecasting performance is further improved, accompanied by increased computational complexity and memory consumption. The integration of the MSSA introduces notable improvements in forecasting accuracy, though it also leads to higher demands on computational resources. Overall, these findings suggest that MCI achieves a favorable trade-off between forecasting accuracy and computational efficiency while also highlighting the individual contributions of each component in the proposed framework.

Table A.1
Computational efficiency and ablation experimental results.

Method	Params	MACs(M)	Memory (MB)	Train-Time(s/iter)	MSE	MAE
Ab-T	101,728	3.26	1.68	0.0012	0.2504	0.2828
Ab-TF	110,948	3.55	14.35	0.0014	0.2462	0.2833
Ab-TFM	112,368	5.72	16.02	0.0021	0.2455	0.2823
MCI	150,320	14.66	21.35	0.0049	0.2306	0.2717
LightTS	415,878	20.89	5.82	0.0024	0.2457	0.2738
TSMixer	111,584	3.57	13.86	0.0014	0.2467	0.2803
PatchTST	4,648,048	651.22	75.78	0.0057	0.3050	0.2817
Pyraformer	6,066,176	321.76	213.03	0.0215	0.2623	0.2999
iTransformer	4,436,080	171.56	36.97	0.0028	0.3129	0.2876
Informer	12,320,132	926.90	438.52	0.0313	0.2798	0.2890
Transformer	12,910,724	958.58	450.17	0.0332	0.2942	0.2752
TimeXer	8,801,392	294.3	50.87	0.0059	0.3006	0.2758

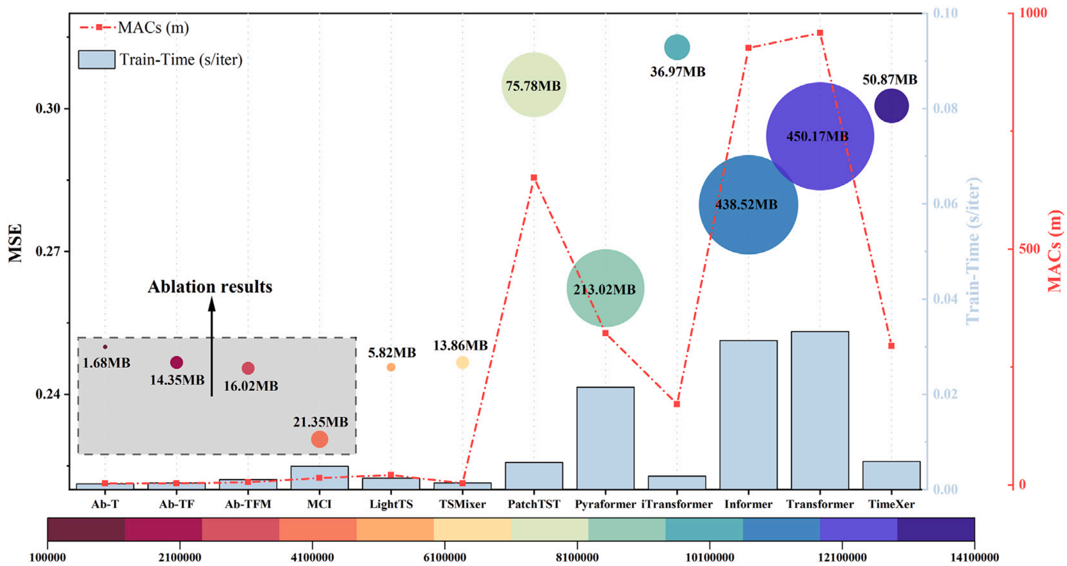

Fig. A.1. Computational efficiency comparison of different methods.

Appendix B. Impact analysis of different covariates on prediction performance

To further examine the influence of different covariates on prediction performance, this study conducted experiments under various scenarios of partial covariate missingness, while ensuring the target variable remained complete. Specifically, based on the proposed MDCTL-MCI method, eight missingness patterns were generated from all possible single-, dual-, and triple-variable combinations of TSI, DNI, and GHI. The hyperparameters used in these experiments were consistent with those described in the main text,

Table B.1
Analysis of the impact of missingness in different types of covariates on forecasting performance.

Missing Covariates	TSI		DNI		GHI		DNI + TSI		TSI + GHI		DNI + GHI	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
RATE0.0	0.1274	0.1889	0.1274	0.1888	0.1274	0.1889	0.1274	0.1888	0.1274	0.1889	0.1274	0.1889
RATE0.1	0.1280	0.1910	0.1279	0.1895	0.1275	0.1895	0.1279	0.1924	0.1273	0.1913	0.1271	0.1899
RATE0.3	0.1274	0.1919	0.1277	0.1888	0.1276	0.1914	0.1266	0.1943	0.1274	0.1939	0.1280	0.1943
RATE0.5	0.1278	0.1936	0.1276	0.1913	0.1277	0.1933	0.1280	0.1986	0.1297	0.2000	0.1293	0.1988
RATE0.7	0.1274	0.1955	0.1269	0.1932	0.1268	0.1951	0.1281	0.2037	0.1290	0.2047	0.1290	0.2074
RATE0.9	0.1275	0.1965	0.1267	0.1927	0.1279	0.1976	0.1278	0.2009	0.1292	0.2071	0.1280	0.2009
Average	0.1276	0.1929	0.1274	0.1907	0.1275	0.1926	0.1276	0.1965	0.1283	0.1977	0.1281	0.1967
CV	0.1914	1.3619	0.3305	0.9374	0.2678	1.5900	0.3934	2.5886	0.7767	3.4337	0.6181	3.2762

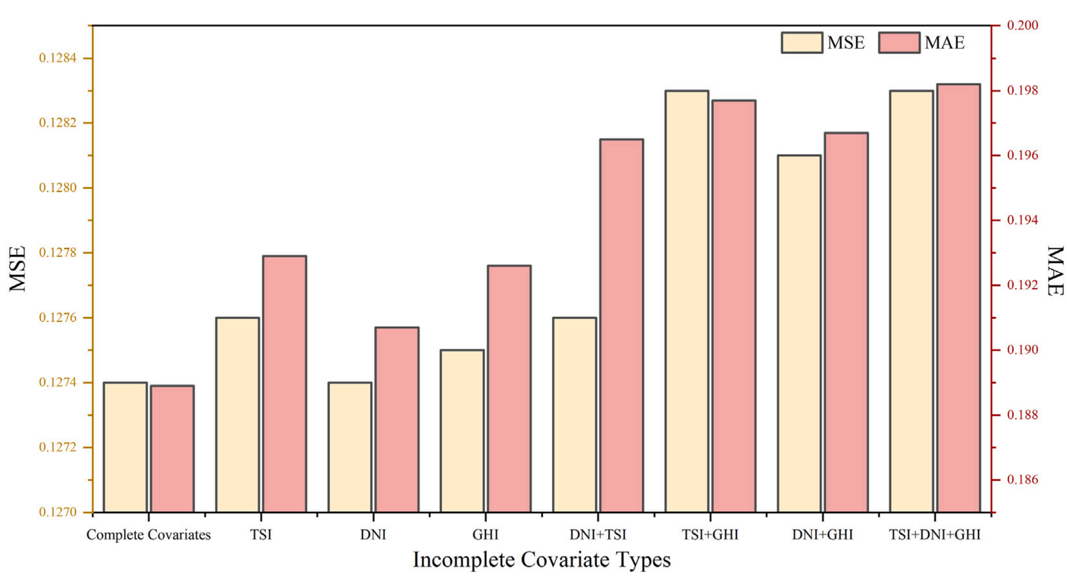

Fig. B.1. Average forecasting performance under missingness of different covariate types.

and the corresponding results are presented in Table B.1 and Fig. B.1.

The results show that all types of partial missingness reduced prediction accuracy. Nevertheless, the proposed MDCTL-MCI method maintained relatively stable predictive performance, with CV values consistently at a low level. The effects of partial missingness in GHI, DNI, or TSI on prediction accuracy were generally comparable, with the overall trend following the order TSI > GHI > DNI. This pattern may be explained by the strong physical correlations among GHI, DNI, and TSI, as well as their similar correlation weights with the target variable. As the number of variables with partial missingness increased, prediction accuracy declined markedly, reaching its lowest point when all three covariates were partially missing. These findings further demonstrate the applicability and robustness of the proposed method in addressing various forms of partial covariate missingness.

Appendix C. Impact of partial missingness in the target variable on predictive performance

To systematically assess the impact of partial missingness in the target variable (Power) on prediction performance, this study compared three partial-missingness scenarios: MS-TC (multivariate input-single target forecasting with partially missing target and covariates), MS-T (multivariate input-single target forecasting with partially missing target), and S-T (single-target-variable input-single target forecasting with partially missing target). The missingness rate was kept identical across all variables, and predictions were evaluated against the complete target variable to ensure comparability and fairness. The proposed MDCTL-MCI model, its locally trained model MCI, and the best-performing baseline method, LightTS, were employed for comparison, with results presented in Fig. C.1.

The results indicate that partial missingness in the target variable has a substantially stronger adverse effect on prediction performance than partial missingness in covariates. In both MS-TC and MS-T scenarios, increasing the missingness rate led to a clear downward trend in prediction accuracy for all models, underscoring the central role of the target variable in the modeling process. Specifically, LightTS exhibited a sharp decline in prediction accuracy once the missingness rate exceeded 0.5, whereas MDCTL-MCI and MCI maintained relatively stable predictive performance, with noticeable error increases only when the missingness rate surpassed 0.7. This demonstrates their greater robustness and adaptability compared to LightTS.

It is worth noting that under low missingness conditions (missingness rate <0.3), the MS-TC scenario yielded slightly higher prediction accuracy for MDCTL-MCI and MCI than MS-T. This suggests that when missingness is mild, the proposed covariate feature extraction method can effectively exploit the temporal information contained in covariates to aid in estimating the target variable, thereby mitigating the performance drop caused by target variable missingness. However, as the missingness rate increased, the concurrent absence of the target variable and covariates in the MS-TC scenario heightened learning difficulty, ultimately leading to lower prediction accuracy than in the MS-T scenario. Overall, MDCTL-MCI consistently demonstrated superior modeling capability across all three missingness scenarios and achieved the highest average prediction accuracy, further confirming its effectiveness and robustness when modeling with missing data.

Appendix D. Exploring model applicability to different covariates

To further assess the proposed model's applicability and scalability to covariates with varying characteristics, air temperature (AT)–which shows relatively weak correlation with the target variable–was introduced as an additional covariate. A systematic evaluation was then performed to examine the model's predictive accuracy and stability under different covariate missingness rates. Using Site3 and Site4 as case studies, we compared the proposed model with MCI and two strong baselines, LightTS and TimeXer. The results are presented in Table D.1, where '(AT)' denotes the experimental setting in which air temperature was added to the original covariate set.

The findings reveal that incorporating AT had minimal impact on the predictive accuracy of MDCTL-MCI, MCI, and TimeXer compared with the setting without AT, whereas LightTS exhibited a marked decline in accuracy under high missingness rates. This effect may be attributed to the weak correlation between AT and the target variable. In this study, AT offered limited additional information for forecasting, and under high missingness, its absence introduced noise that amplified uncertainty in the model inputs, ultimately degrading performance. These results underscore the importance of selecting covariates judiciously to enhance forecasting accuracy while managing computational complexity.

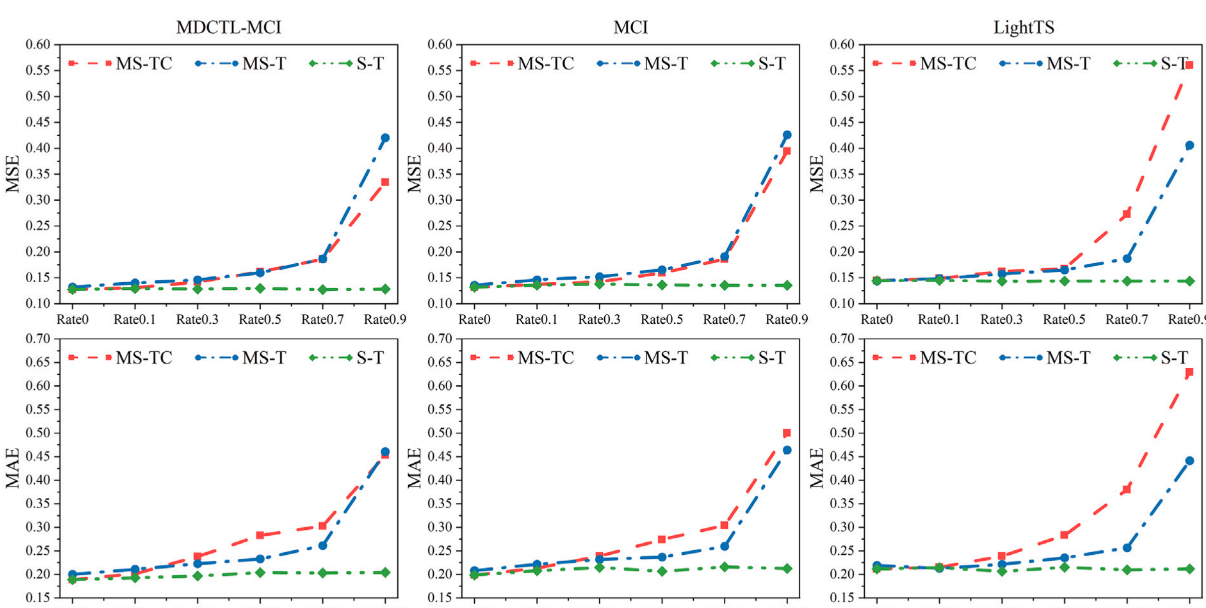

Fig. C.1. Forecasting performance under partial missingness of the target variable.

Table D.1
Comparison of forecasting performance under varying covariate missingness rates with and without the AT covariate.

SITE3	RATE0		RATE0.1		RATE0.3		RATE0.5		RATE0.7		RATE0.9		CV	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MDCTL-MCI	0.2172	0.2592	0.2163	0.2631	0.2180	0.2693	0.2214	0.2701	0.2237	0.2691	0.2275	0.2654	1.962	1.615
MDCTL-MCI(AT)	0.2159	0.2629	0.2170	0.2638	0.2172	0.2637	0.2201	0.2645	0.2238	0.2628	0.2299	0.2701	2.434	1.043
MCI	0.2161	0.2610	0.2171	0.2684	0.2275	0.2736	0.2407	0.2845	0.2457	0.2926	0.2400	0.2790	5.530	4.107
MCI(AT)	0.2148	0.2665	0.2184	0.2698	0.2196	0.2710	0.2372	0.3021	0.2301	0.2837	0.2539	0.3203	6.449	7.521
LightTS	0.2407	0.2721	0.2397	0.2709	0.2407	0.2731	0.2482	0.2866	0.2436	0.2883	0.2487	0.3063	1.636	4.860
LightTS(AT)	0.2387	0.2790	0.2403	0.2797	0.2519	0.2896	0.2610	0.3150	0.2697	0.3332	0.2999	0.3925	8.757	13.863
TimeXer	0.2820	0.2674	0.2830	0.2695	0.2830	0.2655	0.2864	0.2708	0.2858	0.2703	0.2850	0.2683	0.623	0.740
TimeXer(AT)	0.2769	0.2637	0.2807	0.2659	0.2861	0.2707	0.2842	0.2675	0.2885	0.2751	0.2847	0.2679	1.449	1.481
SITE4	RATE0		RATE0.1		RATE0.3		RATE0.5		RATE0.7		RATE0.9		CV	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
MDCTL-MCI	0.1274	0.1889	0.1292	0.1927	0.1281	0.1966	0.1295	0.2037	0.1271	0.2031	0.1283	0.2044	0.759	3.281
MDCTL-MCI(AT)	0.1272	0.1926	0.1281	0.1980	0.1301	0.2138	0.1328	0.2181	0.1319	0.2113	0.1312	0.2084	1.698	4.720
MCI	0.1317	0.1989	0.1357	0.2079	0.1380	0.2147	0.1360	0.2066	0.1354	0.2160	0.1353	0.2125	1.509	3.033
MCI(AT)	0.1273	0.1993	0.1338	0.2093	0.1350	0.2205	0.1348	0.2202	0.1356	0.2202	0.1404	0.2362	3.113	5.716
LightTS	0.1443	0.2118	0.1452	0.2143	0.1431	0.2067	0.1437	0.2151	0.1435	0.2099	0.1435	0.2118	0.538	1.452
LightTS(AT)	0.1406	0.2039	0.1448	0.2121	0.1427	0.2105	0.1481	0.2382	0.1781	0.2978	0.1516	0.2462	9.166	14.962
TimeXer	0.1488	0.2177	0.1544	0.2217	0.1495	0.2123	0.1507	0.2175	0.1499	0.2175	0.1518	0.2151	1.333	1.446
TimeXer(AT)	0.1433	0.2010	0.1510	0.2201	0.1507	0.2165	0.1482	0.2122	0.1523	0.2204	0.1501	0.2158	2.149	3.353

Furthermore, for MCI and LightTS, including the AT covariate resulted in higher coefficients of CV across different missingness rates, leading to increased instability in predictive accuracy. In contrast, MDCTL-MCI and TimeXer consistently maintained lower CV values under the same conditions, indicating more stable predictions across missingness scenarios. Among all experimental settings, MDCTL-MCI attained the highest average prediction accuracy. Overall, MDCTL-MCI maintained superior accuracy and robustness even with the inclusion of a weakly correlated covariate, further confirming its broad applicability and stability under diverse covariate characteristics.

Appendix E. Supplementary data

Supplementary data for this article can be found online at doi:10.1016/j.apenergy.2025.126771.

Data availability

The source code and test data are publicly available at https://github.com/liujian123223/MDCTL-MCI.

References

[1] Truong C, Trueck S, Pitt D, Best R. Seasonality and valuation of renewable energy projects in a two factor model. Appl Energy 2025;389:125669.
[2] Gielen D, Boshell F, Saygin D, Bazilian MD, Wagner N, Gorini R. The role of renewable energy in the global energy transformation. Energy Strateg Rev 2019;24:38–50.
[3] Rai A, Shrivastava A, Jana K. An empirical analysis of machine learning algorithms for solar power forecasting in a high dimensional uncertain environment. IETE Tech Rev 2023;40:558–73.
[4] International Renewable Energy Agency. Renewable capacity statistics 2023. Abu Dhabi: International Renewable Energy Agency; 2023.
[5] Pourasl HH, Barenji RV, Khojastehnezhad VM. Solar energy status in the world: a comprehensive review. Energy Rep 2023;10:3474–93.
[6] Tian Z, Chen Y, Wang G. Enhancing pv power forecasting accuracy through nonlinear weather correction based on multi-task learning. Appl Energy 2025;386:125525.
[7] Espinar B, Aznarte JL, Girard R, Moussa AM, Kariniotakis G. Photovoltaic forecasting: A state of the art., in: 5th European PV-hybrid and mini-grid conference, OTTI-Ostbayerisches Technologie-Transfer-Institut, 2010,pp. 250.
[8] Liu J, Yang F, Yan K. Multi-scale data fusion and adaptiveloss kolmogorov-arnold network for multivariate time series forecasting. Inf Fusion 2026;125:103454. https://doi.org/10.1016/j.inffus.2025.103454
[9] Liu J, Yang F, Yan K, Jiang L. Household energy consumption forecasting based on adaptive signal decomposition enhanced itransformer network. Energy Build 2024;324:114894.
[10] Bi J, Wang H, Yan E, Wang C, Yan K, Jiang L, et al. AI in hvac fault detection and diagnosis: a systematic review. Energy Rev 2024;3:100071. https://www.sciencedirect.com/science/article/pii/S277297022400004X
[11] Yang Y, Liu Y, Zhang Y, Shu S, Zheng J. Dest-gnn: a double-explored spatio-temporal graph neural network for multi-site intra-hour pv power forecasting. Appl Energy 2025;378:124744.
[12] Hu Z, Gao Y, Ji S, Mae M, Imaizumi T. Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data. Appl Energy 2024;359:122709.
[13] Almaghrabi S, Rana M, Hamilton M, Rahaman MS. Multivariate solar power time series forecasting using multilevel data fusion and deep neural networks. Information Fusion 2024;104:102180.
[14] Li Y, Liu G, Cao Y, Chen J, Gang X, Tang J. Wnps-LSTM-informer: a hybrid stacking model for medium-term photovoltaic power forecasting with ranked feature selection. Renew Energy 2025:122687.
[15] Fu J, Sun Y, Li Y, Wang W, Wei W, Ren J, et al. An investigation of photovoltaic power forecasting in buildings considering shadow effects: modeling approach and shap analysis. Renew Energy 2025:122821.
[16] Zhang R, Wu Y, Zhang L, Xu C, Wang Z, Zhang Y, et al. A multiscale network with mixed features and extended regional weather forecasts for predicting short-term photovoltaic power. Energy 2025:134792.
[17] Song Z, Xiao F, Chen Z, Madsen H. Probabilistic ultra-short-term solar photovoltaic power forecasting using natural gradient boosting with attention-enhanced neural networks. Energy AI 2025:100496.
[18] Hosseini P, Taheri S, Akhavan J, Razban A. Privacy-preserving federated learning: application to behind-the-meter solar photovoltaic generation forecasting. Energy Convers Manag 2023;283:116900.
[19] Khalil M, Esseghir M, Boulahia LM. Privacy-preserving federated learning: an application for big data load forecast in buildings. Comput Secur 2023;131:103211.
[20] Yu S, He B, Fang L. Multi-step short-term forecasting of photovoltaic power utilizing timesnet with enhanced feature extraction and a novel loss function. Appl Energy 2025;388:125645.
[21] Liu Y, Liu Y, Cai H, Zhang J. An innovative short-term multihorizon photovoltaic power output forecasting method based on variational mode decomposition and a capsule convolutional neural network. Appl Energy 2023;343:121139.
[22] Deng F, Wang J, Wu L, Gao B, Wei B, Li Z. Distributed photovoltaic power forecasting based on personalized federated adversarial learning. Sustain Energy Grids Netw 2024;40:101537.
[23] Fan S, Geng H, Zhang H, Yang J, Hiroichi K. Photovoltaic power forecasting model employing epoch-dependent adaptive loss weighting and data assimilation. Solar Energy 2025;290:113351.
[24] Gao Y, Ruan Y, Fang C, Yin S. Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. Energy Build 2020;223:110156.
[25] Gao Y, Hu Z, Chen WA, Liu M. Solutions to the insufficiency of label data in renewable energy forecasting: a comparative and integrative analysis of domain adaptation and fine-tuning. Energy 2024;302:131863.
[26] Li H, He X, Hu Y, Lv W, Yang L. Research on the generation method of missing hourly solar radiation data based on multiple neural network algorithm. Energy 2024;287:129650.
[27] Başakín E.E., Ekmekcioğlu Ö., Özger M. Developing a novel approach for missing data imputation of solar radiation: a hybrid differential evolution algorithm based extreme gradient boosting model. Energy Convers Manag 2023;280:116780.
[28] Zhang L, Jiang Z, Ji T, Chen Z. Diffusion-based inpainting approach for multifunctional short-term load forecasting. Appl Energy 2025;377:124442.
[29] Rodríguez F, Genn M, Fontán L, Galarza A. Very short-term temperature forecaster using MLP and n-nearest stations for calculating key control parameters in solar photovoltaic generation. Sustain Energy Technol Assess 2021;45:101085.
[30] Liu J, Huang X, Li Q, Chen Z, Liu G, Tai Y. Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD. Energy Convers Manag 2023;280:116804.

[31] Jain M, Yadav P, Dev S. Holistic and lightweight approach for solar irradiance forecasting. IEEE Trans Geosci Remote Sens 2024.
[32] Chen Y, Xu J. Solar and wind power data from the Chinese state grid renewable energy generation forecasting competition. Sci Data 2022;9:577.
[33] Liu S, Yu H, Liao C, Li J, Lin W, Liu AX, et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting. In: #PLACEHOLDER_PARENT_METADATA_VALUE#; 2022.
[34] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Adv Neural Inf Process Syst 2017;30.
[35] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the AAAI conference on artificial intelligence; 2021. p. 11106–15.
[36] Wang Y, Wu H, Dong J, Qin G, Zhang H, Liu Y, et al. Timexer: empowering transformers for time series forecasting with exogenous variables, arXiv:2402.19072, [preprint]. 2024
[37] Liu Y, Hu T, Zhang H, Wu H, Wang S, Ma L, et al. Itransformer: inverted transformers are effective for time series forecasting. arXiv:2310.06625, [preprint]. 2023
[38] Nie Y, Nguyen NH, Sinthong P, Kalagnanam J. A time series is worth 64 words: long-term forecasting with transformers.arXiv:2211.14730, [preprint]. 2022
[39] Zhang T, Zhang Y, Cao W, Bian J, Yi X, Zheng S, et al. Less is more: fast multivariate time series forecasting with light sampling-oriented mlp structures, arXiv:2207.01186. 2022
[40] Chen SA, Li CL, Yoder N, Arik SO, Pfister T. Tsmixer: an all-mlp architecture for time series forecasting. arXiv:2303.06053, [preprint]. 2023
[41] Wu H, Hu T, Liu Y, Zhou H, Wang J, Long M. Timesnet: temporal 2D-variation modeling for general time series analysis. In: International conference on learning representations; 2023.
[42] Zeng A, Chen M, Zhang L, Xu Q. Are transformers effective for time series forecasting? In: Proceedings of the AAAI conference on artificial intelligence; 2023. p. 11121–8.