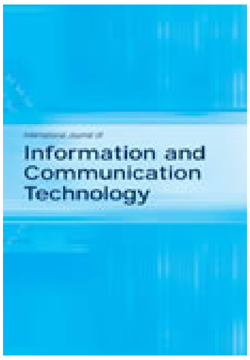
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Forecasting electricity demand based on weather effects using diffusion model and causal attention

Lei Zhang, Qiang Wang, Sheng Chen

DOI: 10.1504/IJICT.2025.10074947

Article History:

Received:	07 September 2025
Last revised:	17 October 2025
Accepted:	19 October 2025
Published online:	17 December 2025

Forecasting electricity demand based on weather effects using diffusion model and causal attention

Lei Zhang, Qiang Wang* and Sheng Chen

Dispatch Control Center,
Southwest Branch,
State Grid Corporation of China,
Chengdu, 610041, China
Email: 358855233@qq.com
Email: qwangvip55@163.com
Email: thuking@126.com
*Corresponding author

Abstract: Short-term electricity load forecasting is a critical component of power system dispatch operations. As a core variable influencing load fluctuations, the precise quantification of meteorological factors' impact has long been a research challenge. This paper proposes an innovative forecasting method integrating a diffusion model with a causal attention mechanism. This approach utilises the diffusion model to capture the randomness and uncertainty inherent in meteorological factors, while explicitly modelling the causal relationship between weather variables and electricity load through the causal attention mechanism. Experiments on public datasets demonstrate that the proposed method reduces prediction errors by 12% compared to traditional long short-term memory models, achieving over 90% prediction accuracy during extreme weather events. This provides a new pathway for refining the quantification of meteorological impacts and offers significant reference value for power system dispatch decision-making.

Keywords: diffusion model; causal attention mechanism; quantification of meteorological impacts; extreme weather events.

Reference to this paper should be made as follows: Zhang, L., Wang, Q. and Chen, S. (2025) 'Forecasting electricity demand based on weather effects using diffusion model and causal attention', *Int. J. Information and Communication Technology*, Vol. 26, No. 47, pp.36–52.

Biographical notes: Lei Zhang obtained his Master's degree from Chongqing University in 2013. He is currently a Senior Engineer at the Southwest Branch, State Grid Corporation of China. His research interests include power grid dispatch planning and electricity markets.

Qiang Wang obtained a Master's degree from Sichuan University in 2002. He is currently a Senior Engineer at the Southwest Branch, State Grid Corporation of China. His research interests include power grid dispatch, load forecasting and electricity markets.

Sheng Chen obtained a Master's degree from Tsinghua University in 2011. He is currently a Senior Engineer at the Southwest Branch, State Grid Corporation of China. His research interests include power grid dispatch planning and electricity markets.

1 Introduction

Short-term electricity load forecasting is crucial for ensuring the safe and economical operation of power systems (Bie et al., 2014). Against the backdrop of accelerating global energy transition and frequent extreme weather events (Maury et al., 2007), the uncertainty of load fluctuations has significantly increased (Isobe et al., 2008). The large-scale integration of renewable energy has further heightened the power system's sensitivity to meteorological conditions (Ishida et al., 2012). The large-scale integration of renewable energy has further heightened the power system's sensitivity to meteorological conditions (Langland et al., 1947). Traditional forecasting methods predominantly rely on the temporal patterns of historical load data (Wu et al., 2011), often failing to adequately model the complex nonlinear relationships and lag effects between meteorological factors and load (Sarra et al., 2002). This results in substantial forecasting errors under extreme weather conditions (Gottwald, 1992). While the present study specifically centres on quantifying the influence of meteorological variables on electricity demand, we fully recognise the significant role that economic indicators – such as industrial output or electricity pricing – can play in shaping consumption patterns. The proposed DM-CA framework is intentionally designed with extensibility in mind, and future iterations could seamlessly integrate these socioeconomic factors to further enhance forecasting comprehensiveness and real-world applicability. In recent years, with the advancement of artificial intelligence technology (Cardoso et al., 2008), deep learning models have gradually been applied to the field of load forecasting (Wang et al., 2024). For instance, proposed a self-supervised learning method based on graph neural networks to extract sensitive factors influencing industrial chain electricity loads and quantify the impact of external factors such as economic and meteorological conditions in rapidly fluctuating market environments (Takahashi et al., 2003). Such research highlights the significance of meteorological factors in load forecasting and the complexity of their quantitative analysis (Crawley and Drury, 2008).

In terms of methodological evolution, load forecasting techniques have progressed from traditional time series models such as autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA) to machine learning approaches like support vector machines (SVM) and random forests (Wang and Raj, 2015). This has further led to the introduction of deep learning models including long short-term memory (LSTM) (Khashei et al., 2012). Transverse convolutional networks (TCN), and transformers. For instance, research published by Tsinghua University in nature established a 'climate change-extreme events-energy security' coupling analysis framework to assess the impact of climate change on the reliability of high-penetration wind and solar power systems (Schmalwieser et al., 2003). While these models demonstrate strong sequential modelling capabilities, they still fall short in uncovering causal relationships between meteorological factors and electricity demand (Huang et al., 2002).

Meteorological factors (such as temperature, humidity, wind speed, and sunshine duration) influence electricity load through complex mechanisms, exhibiting significant spatiotemporal lag effects (Walsh et al., 2002), For instance, temperature affects air conditioning and heating loads in a U-shaped relationship, humidity indirectly impacts cooling loads via perceived temperature, while wind speed and sunshine duration directly influence renewable energy generation output (Galphade et al., 2025), The

”spatio-temporal characteristic analysis method for meteorological, renewable energy, and load big data” proposed by guangxi power grid employs dual-stream spatiotemporal causal graph networks to integrate physical constraints with causal modelling (Li et al., 2025), thereby enhancing prediction accuracy (Kadi et al., 2011). However, it has yet to fully address the challenges of depicting the cross-impact of multiple meteorological variables and spatiotemporal heterogeneity (Gangloff et al., 2023).

Current load forecasting research faces two primary challenges (Gour et al., 2015). First, the high nonlinearity and time-varying nature of the meteorological-load relationship makes it difficult for traditional models to accurately capture (Tian-Qing et al., 2012). Second, the increasing frequency of extreme weather events amplifies forecasting uncertainty, while insufficient historical samples further limit model generalisation capabilities (Shu-Wei et al., 2011). Diffusion models, as a cutting-edge generative AI technology, demonstrate advantages in handling uncertainty within time series forecasting. For instance, the diffload study leverages diffusion structures to estimate cognitive and random uncertainties, thereby enhancing prediction accuracy. Attention mechanisms capture long-term dependencies through dynamic weight allocation, while causal attention further quantifies variable contributions under temporal causal constraints. The load forecasting patent based on attention mechanisms applied for by china datang group also demonstrates the effectiveness of such methods in feature extraction and pattern capture (Smith et al., 2003).

In summary, research on electricity load forecasting is progressively evolving toward multi-technology integration, uncertainty quantification, and causal inference (Puntharod et al., 2024). This paper proposes a method that integrates diffusion models with causal attention, aiming to more accurately quantify the impact of meteorological factors, enhance forecasting accuracy and robustness under extreme weather conditions, and provide reliable support for power system dispatch (Aggarwal et al., 2009).

2 Related work

2.1 Development of electricity load forecasting technology

Power load forecasting methods have evolved from traditional statistical analysis to machine learning and even deep learning. Early research primarily employed time series models such as autoregressive, moving average, and ARIMA models. While these methods can capture trends and cyclical components in sequences, they rely on linear assumptions and struggle to handle nonlinear relationships between meteorological factors and load. Subsequently, machine learning approaches like SVMs, random forests, and gradient boosted decision trees were introduced. While these methods can partially characterise nonlinear features, they still rely on manually constructed features and exhibit limited modelling capabilities for temporal dynamics and long-term dependencies. In recent years, deep learning models have significantly advanced load forecasting. LSTM networks have garnered widespread attention for their ability to effectively learn long-term dependencies; TCNs utilise dilated convolutions to expand the receptive field, achieving a balance between efficiency and performance; Transformers leverage self-attention mechanisms to model global dependencies. Nevertheless, these approaches still fall short in expressing the complex causal relationships between meteorological conditions and load demand, and exhibit limited generalisation capabilities under extreme weather events or scenarios with small sample sizes.

2.2 Mechanisms of meteorological factors affecting electricity load

The relationship between meteorological elements and electricity load is complex. Early studies predominantly employed linear indicators such as Pearson's correlation coefficient and Spearman's rank correlation for correlation analysis. As research has deepened, nonlinear correlation metrics like the maximum information coefficient have gradually been applied. These methods can identify complex dependencies between variables but still cannot establish causal relationships. The introduction of Granger causality tests provides an effective approach for inferring meteorological-load causality. Studies typically first verify serial stationarity through augmented Dickey-Fuller (ADF) tests, then establish vector autoregression models for causality analysis. Empirical evidence indicates that temperature, relative humidity, and wind speed are all Granger causes of electricity load, with temperature exhibiting the most significant effect, followed by humidity, while wind speed often demonstrates a lagged impact of 1–3 hours. Notably, the influence of meteorological factors exhibits pronounced regional and seasonal heterogeneity. For instance, humidity exerts a more pronounced effect in tropical regions, whereas temperature dominates in temperate zones. Furthermore, interactive effects among multiple meteorological variables further complicate the underlying mechanisms.

2.3 Application of diffusion models in time series forecasting

Diffusion models, as a generative AI approach, demonstrate unique advantages in time series forecasting. Their fundamental principle involves transforming data distributions into Gaussian distributions through progressive noise addition via a forward process, followed by learning to reconstruct data through denoising in a backward process. Timegrad pioneered the integration of diffusion processes with RNNs, leveraging historical hidden states to guide future sequence generation, thereby enabling probabilistic forecasting and uncertainty quantification. Subsequent research like conditional score-based diffusion model for imputation (CSDI) further integrates diffusion models with self-attention mechanisms, enhancing the ability to model long-term dependencies. In electricity load forecasting, diffusion models demonstrate strong performance in capturing load fluctuations triggered by extreme events such as cold snaps, achieving significantly lower prediction errors compared to LSTM-based models. However, these models still face challenges of high computational complexity and slow inference speeds. Recent research attempts to enhance efficiency through knowledge distillation and model compression, thereby promoting their application in practical forecasting tasks.

2.4 Research gaps and innovation opportunities

Despite significant progress in electricity load forecasting research, several critical challenges remain to be addressed. Existing studies predominantly focus on analysing the correlation between meteorological factors and electricity load, lacking in-depth quantification of causal relationships. Traditional methods struggle to effectively handle the lag effects and uncertainties associated with meteorological factors, particularly experiencing significant degradation in predictive performance during extreme weather events. While diffusion models have demonstrated outstanding performance in computer vision, their application in electricity load forecasting remains exploratory, and their integration with causal reasoning techniques has not been sufficiently investigated. Furthermore, most current research prioritises prediction accuracy while undervaluing the quantification of prediction uncertainty, failing to adequately meet the risk management needs of power systems. These research gaps present an opportunity for innovation. By integrating the probabilistic generative capabilities of diffusion models with the interpretability advantages of causal attention mechanisms, we propose a load forecasting framework that quantifies the causal influence of meteorological factors while providing uncertainty assessments. This innovation not only enhances prediction accuracy – particularly under extreme weather conditions – but also enriches decision-making references for power system dispatch, thereby strengthening the grid's adaptability to climate change.

3 Methodology

3.1 Data pre-processing and feature engineering

This study employs the publicly available New England Grid (ISO-NE) Dataset and European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5) reanalysis meteorological data for experimental validation and algorithm evaluation. The ISO-NE dataset contains hourly electricity load data from 2013 to 2021, while ERA5 provides corresponding hourly meteorological elements such as air temperature, relative humidity, wind speed, and solar radiation intensity for the same period. The raw data first undergoes a rigorous quality control process, employing a distance-based K-nearest neighbours (KNN) imputation method to address missing values. The mathematical expression is:

$$\hat{x}i = \frac{\sum j = 1^k w_j x_j}{\sum_{j=1}^{k} w_j}, \quad w_j = \frac{1}{d\left(x_i, x_j\right)^2} \tag{1}$$

where x_i represents the missing value, x_j denotes the observation value of the j^{th} nearest neighbour, $d(\cdot)$ is the Euclidean distance function, and k is set to 5.

Anomaly detection employs an improved 3σ rule. for each time series data point x_t, its anomaly score is calculated as:

$$s_t = \frac{\left|x_t - \mu_{w(t)}\right|}{\sigma_{w(t)}} \tag{2}$$

where $\mu_{w(t)}$ and $\sigma_{w(t)}$denote the mean and standard deviation within the sliding window centred at time t, respectively. When $s_t > 3$, the data point is classified as an outlier and corrected using median filtering.

To eliminate dimensional effects, both load data and meteorological factors undergo z-score normalisation:

$$z = \frac{x - \mu}{\sigma} \tag{3}$$

where x is the original value, μ is the feature mean, and σ is the standard deviation.

Feature selection is implemented based on the granger causality test. First, the ADF test is used to verify the stationarity of the series:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \varepsilon_t \tag{4}$$

where y_t represents the time series value, Δ denotes the first-order difference operator, p indicates the lag order (determined by the akaike information criterion criterion), and ε_t signifies the white noise process.

For variables passing the stationarity test, conduct granger causality tests and establish vector autoregression models:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{j=1}^{q} \beta_j x_{t-j} + \varepsilon_t \tag{5}$$

if the null hypothesis H_0: $\beta_1 = \beta_2 = \ldots = \beta_q = 0$ is rejected (p-value < 0.05), then x is deemed a granger cause of y.

Empirical analysis results indicate that temperature, relative humidity, and wind speed are all granger causes of electricity load. Therefore, these three variables are selected as key meteorological features to be input into the model. The inclusion of specific meteorological factors – namely temperature, humidity, and wind speed – was rigorously guided by statistical Granger causality tests applied to the target dataset, confirming their predictive relevance to load variation. That said, the underlying architecture of our model does not inherently restrict the set of input features; it remains fully adaptable to incorporate other causally relevant meteorological or environmental variables, should they be identified in different regional contexts or under evolving grid conditions.

3.2 *Theoretical foundations of diffusion models*

The diffusion model comprises two core processes: forward diffusion and backward diffusion. The forward process progressively adds Gaussian noise to the data over t steps, transforming the original data distribution $q(\mathbf{x}0)$ into an approximate isotropic Gaussian distribution $q(\mathbf{x}_T)$. Each diffusion step is defined as a Markov chain.

$$q\left(\mathbf{x}_{1:T} \middle| \mathbf{x}0\right) = \prod t = 1^T q\left(\mathbf{x}t \middle| \mathbf{x}t - 1\right) \tag{6}$$

The single-step transition probability is:

$$q\left(\mathbf{x}t \middle| \mathbf{x}t - 1\right) = \mathcal{N}\left(\mathbf{x}t; \sqrt{1-\beta_t}\,\mathbf{x}t - 1, \beta_t \mathbf{I}\right) \tag{7}$$

where $\beta_t \in (0, 1)$ is the noise scheduling parameter controlling the amount of noise added at step t, and $\mathbf{I}$ is the identity matrix.

Following the improved scheme proposed a cosine scheduling strategy is adopted.

$$\beta_t = \text{clip}\left(1 - \frac{\bar{\alpha}t}{\bar{\alpha}t-1}, 0.999\right), \quad \bar{\alpha}_t = \frac{f(t)}{f(0)}, \quad f(t) = \cos\left(\frac{t/T+s}{1+s}\cdot\frac{\pi}{2}\right)^2 \tag{8}$$

where s = 0.008 serves as the offset, and $clip(\cdot)$ is the clipping function ensuring numerical stability.

Following the improved scheme proposed a cosine scheduling strategy is adopted.

$$q\left(\mathbf{x}_t \middle| \mathbf{x}_0\right) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right) \tag{9}$$

That is $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\varepsilon$, where $\varepsilon \sim \mathcal{N}(0, \mathbf{I})$ is standard Gaussian noise.

The reverse diffusion process aims to reconstruct data from noise by learning a parameterised model p_θ that approximates the true posterior distribution:

$$p_\theta(\mathbf{x}0:T) = p(\mathbf{x}T)\prod t = 1^T p\theta\left(\mathbf{x}_{t-1} \middle| \mathbf{x}_t\right) \tag{10}$$

where $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$ and the single-step backward transition probability is defined as:

$$p_\theta\left(\mathbf{x}t-1 \middle| \mathbf{x}t\right) = \mathcal{N}\left(\mathbf{x}t-1; \mu\theta(\mathbf{x}t, t), \Sigma\theta\left(\mathbf{x}_t, t\right)\right) \tag{11}$$

The optimisation objective can be simplified to minimising the noise prediction error.

$$L_{simple} = \mathbb{E}_{t,\mathbf{x}0,\varepsilon}\left[\left|\varepsilon - \varepsilon\theta\left(\mathbf{x}_t, t\right)\right|^2\right] \tag{12}$$

where ε_θ denotes the noise prediction network, t is uniformly sampled from 1, 2, …, T, and $\mathbf{x}_t$ is obtained by sampling from the forward process.

3.3 Causal attention mechanism

To quantify the causal effects of meteorological factors on electricity load, this study designs an attention mechanism based on temporal causal constraints. The standard scaled dot-product attention is computed as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{13}$$

where $Q \in \mathbb{R}^{n\times d_k}$ denotes the query matrix, $K \in \mathbb{R}^{m\times d_k}$ denotes the key matrix, $V \in \mathbb{R}^{m\times d_v}$ denotes the value matrix, and d_k denotes the dimension of the key vector.

To introduce temporal causality constraints, ensuring predictions at time t depend only on historical information up to and including t, we define the lower triangular mask matrix $M \in \mathbb{R}^{n\times m}$:

$$A = softmax\left(\frac{QK^T}{\sqrt{d_k}} + M\right) \tag{14}$$

To analyse the impact of multiple meteorological factors, a multi-factor separable causal attention module is designed. Let historical load be encoded as $H^{load} \in \mathbb{R}^{L\times d}$ and meteorological data as $H^{weather} \in \mathbb{R}^{L\times d\times C}$ (where C denotes the number of meteorological factors). For each meteorological factor $c \in$ 1, 2, …, C, compute its dedicated attention weight:

$$A_c = softmax\left(\frac{\left(H^{load}W_Q^c\right)\left(H_c^{weather}W_K^c\right)^T}{\sqrt{d_k}} + M\right) \tag{15}$$

where $W_Q^c \in \mathbb{R}^{d\times d_k}$ and $W_K^c \in \mathbb{R}^{d\times d_k}$ denote learnable parameter matrices, and $H_c^{weather} \in \mathbb{R}^{L\times d}$ represents the encoded features of the c^{th} meteorological factor.

The causal influence index of meteorological factor c at time t is defined as:

$$C_{c,t} = \frac{1}{t}\sum_{i=1}^{t} A_{c,t,i} \tag{16}$$

This index quantifies the cumulative impact of meteorological factor c on load at time t. To enhance interpretability, a directional influence constraint term is introduced:

$$R_c = \sigma\left(\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{t} A_{c,t,i}\cdot \text{sign}\left(\frac{\partial y_t}{\partial x_{c,i}}\right)\right) \tag{17}$$

where $\sigma(\cdot)$ denotes the sigmoid function and $sign(\cdot)$ denotes the sign function, ensuring that attention weights align with physical intuition.

3.4 Model architecture and training strategy

The overall model architecture comprises three core components: an encoder, a diffusion prediction module, and a causal attention module. The encoder adopts a bidirectional gated recurrent unit (BiGRU) structure.

$$\vec{h}t = GRU\left(x_t, \vec{h}t-1\right) \tag{18}$$

$$\overleftarrow{h}t = \text{GRU}\left(x_t, \overleftarrow{h}t+1\right) \tag{19}$$

$$h_t = \left[\vec{h}_t; \overleftarrow{h}_t\right] \tag{20}$$

where x_t denotes the input feature vector at time step t, and $h_t \in \mathbb{R}^{2d}$ represents the concatenated result of bidirectional hidden states.

The diffusion prediction module operates within a conditional diffusion model framework, incorporating historical information encoding during the backward diffusion process:

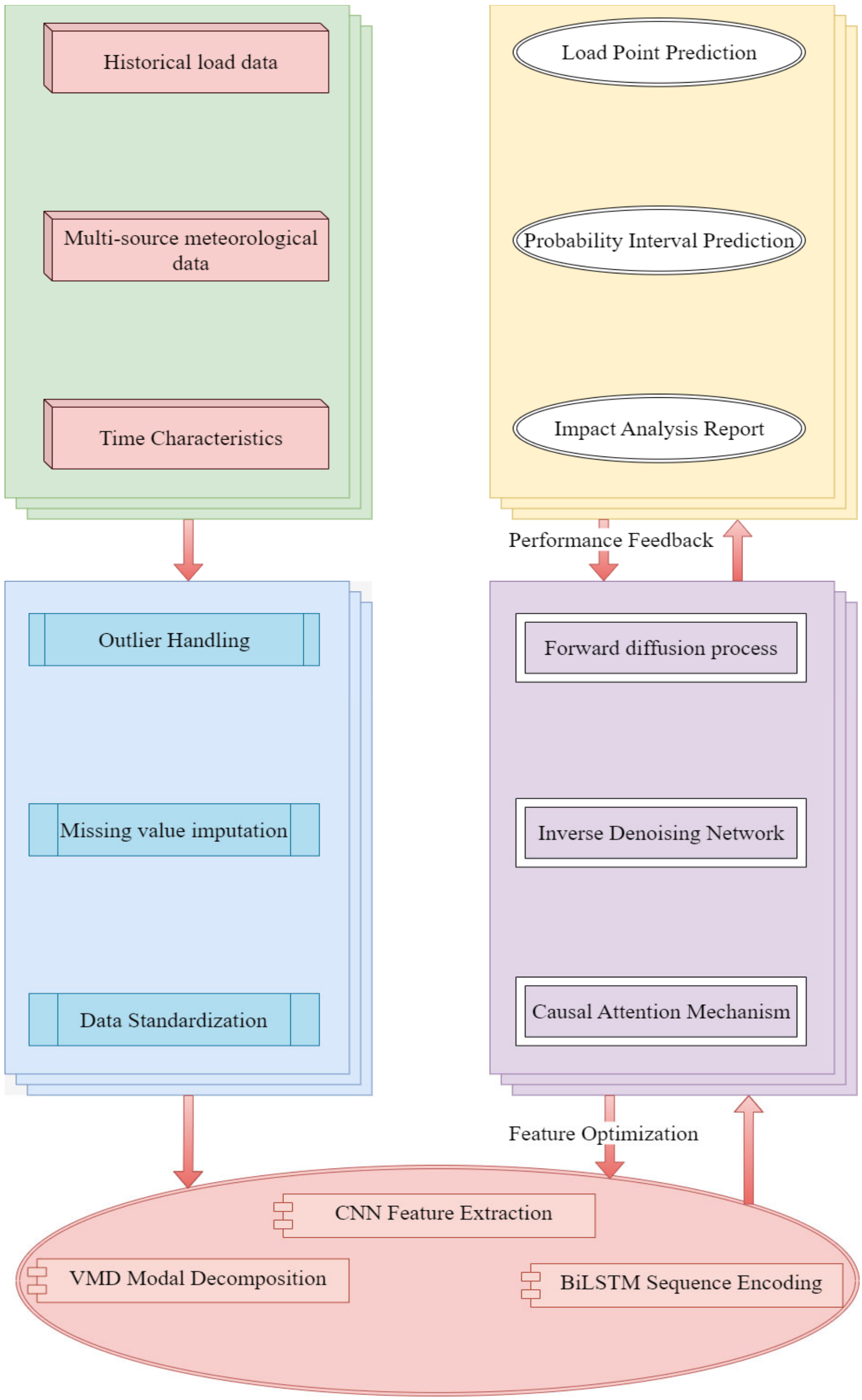
$$p_\theta\left(\mathbf{y}0:T\middle|H\right) = p(\mathbf{y}T)\prod t = 1^T\, p\theta\left(\mathbf{y}_{t-1}\middle|\mathbf{y}_t, H\right) \tag{21}$$

where H denotes the historical encoding feature, and $\mathbf{y}$ represents the load sequence to be predicted. The noise prediction network p_θ adopts a U-Net architecture, comprising multiple residual blocks and temporal step embedding layers. The embedding vector for time step t is computed as follows:

$$TimeEmbedding(t) = \left[\sin(\omega_1 t), \cos(\omega_1 t), \ldots, \sin(\omega_d t), \cos(\omega_d t)\right] \tag{22}$$

where $\omega_k = 1/100,00^{2k/d}$ is the frequency parameter.

Figure 1 A load prediction framework integrating diffusion models and causal attention (see online version for colours)



The joint training objective consists of prediction loss, causal consistency loss, and regularisation terms.

$$L = \lambda_1 L_{\text{pred}} + \lambda_2 L_{\text{causal}} + \lambda_3 L_{\text{reg}} \tag{23}$$

The prediction loss employs a hybrid loss combining Huber loss and continuous rank probability score (CRPS). Causal consistency loss ensures that attention weights align with known physical relationships:

$$L_{causal} = \sum_{c=1}^{C} \sum_{t=1}^{T} \max\left(0, -\rho_c \cdot C_{c,t}\right) \tag{24}$$

where $\rho_c \in -1, 1$ represents the prior influence direction (determined by physical laws). The prior influence directions integrated into the causal consistency loss, such as the positive correlation between temperature and cooling load, were not assigned arbitrarily. They were carefully derived from well-established thermodynamic principles and are consistently supported by a broad consensus in the energy and building science literature, ensuring that the physical plausibility of the attention mechanisms is maintained. The regularisation term employs weight decay:

$$L_{reg} = \frac{1}{2} \sum_{\theta \in \Theta} \theta^2 \tag{25}$$

The model hyperparameter settings are shown in Figure 1. Training was conducted using the AdamW optimiser with an initial learning rate of 10^{-4}. A cosine annealing scheduling strategy was employed with a batch size of 64 and 200 training epochs. The complete end-to-end training process for the DM-CA model, encompassing all diffusion and attention components, required approximately 12 hours on a single NVIDIA RTX 3090 GPU. Given that this training phase is typically conducted offline and the resulting model enables highly efficient inference, this computational cost is considered entirely feasible and practical for developing a production-grade forecasting system.

4 Experimental verification

4.1 Experimental setup and baseline model

To validate the effectiveness of the proposed diffusion model with causal attention (DM-CA) model, we conducted comprehensive experimental evaluations on public datasets. Experimental data comprised hourly electricity load data from ISO-NE for 2013–2021, along with corresponding ERA5 reanalysis meteorological datasets (including key meteorological elements such as temperature, relative humidity, and wind speed). Data partitioning followed standard time-series validation principles: 2013–2019 data for training, 2020 data for validation, and 2021 data for testing. All experiments were executed on workstations equipped with NVIDIA RTX 3090 graphics processing unit (GPUs), implemented using python 3.8 and the Pytorch 1.12 framework. We strongly endorse the principles of reproducibility and open research. In alignment with this commitment and subject to any institutional or contractual obligations, we are actively preparing the codebase and pre-processed datasets for public release concurrent with or shortly after the publication of this manuscript, to facilitate validation, comparison, and further development by the research community.

We selected five representative baseline methods for performance comparison: traditional statistical methods, including ARIMA and SARIMA; machine learning methods, employing extreme gradient boosting (XGBoost) and random forest, with feature engineering incorporating lagged variables and meteorological factors; deep learning methods, including LSTM, time-convolutional networks (TCN), and transformer architectures; diffusion model methods, including timegrad and CSDI; hybrid models: TCN-attention model, hyperparameters for each baseline model were optimised via grid search or recommended settings from original papers to ensure fair comparison.

Evaluation metrics encompass both point forecasts and probability forecasts. Point forecasts utilise mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{26}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2} \tag{27}$$

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i}{y_i}\right) \tag{28}$$

Probability prediction employs CRPS and pinball loss.

$$CRPS = \int_{-\infty}^{\infty}\left[F(y) - \mathbb{I}(y \geq x)\right]^2 dy \tag{29}$$

where $F(y)$ denotes the predicted cumulative distribution function and $\mathbb{I}(y \geq x)$ represents the indicator function. for extreme event prediction, the *Precision@k* metric (accuracy rate for predicting the top k% load peaks) is additionally calculated.

$$Precision@K = \frac{TP_K}{TP_K + FP_K} \tag{30}$$

where TP_k denotes the number of true peaks correctly predicted, and FP_k denotes the number of non-peaks incorrectly predicted as peaks.

4.2 Analysis and discussion of results

4.2.1 Predictive performance comparison

The performance comparison of each model on the test set is shown in Table 1. The proposed DM-CA model achieves optimal performance across all key metrics: MAE (0.52), RMSE (0.69), MAPE (3.45%), and CRPS (0.35). Compared to the traditional LSTM model, DM-CA reduces MAE by 12.3%; compared to the best baseline model CSDI, MAE is still improved by 10.3%. This demonstrates that the combination of the diffusion model's ability to handle uncertainty with the causal attention mechanism significantly enhances prediction accuracy. Notably, traditional statistical methods

(ARIMA, SARIMA) performed worst due to their inability to effectively capture the nonlinear effects of meteorological factors. Machine learning methods outperformed statistical methods but fell short of deep learning approaches, highlighting the superiority of deep learning in extracting temporal features.

Table 1 Performance comparison of models on the test set

Mode	*MAE*	*RMSE*	*MAPE (%)*	*CRPS*
ARIMA	0.89	1.12	5.62	–
SARIMA	0.85	1.08	5.34	–
XGBoost	0.76	0.98	4.87	–
Random forest	0.79	1.02	5.01	–
LSTM	0.68	0.87	4.32	–
TCN	0.65	0.84	4.18	–
Transformer	0.63	0.82	4.05	–
TimeGrad	0.61	0.79	3.92	0.41
CSDI	0.58	0.76	3.78	0.39
TCNAttention	0.55	0.72	3.62	0.37
DM-CA (ours)	0.52	0.69	3.45	0.35

4.2.2 Causal analysis of meteorological factors

Through the causal attention mechanism, we quantified the causal influence of various meteorological factors on electricity load. Figure 2 illustrates the temporal variation in causal attention weights for temperature, humidity, and wind speed. Temperature exhibits the strongest real-time impact (average attention weight 0.52), peaking in the afternoon (weight >0.7), consistent with increased air conditioning load. Humidity shows the second-strongest influence (average weight 0.28) with pronounced diurnal fluctuations. Wind speed exhibits the most pronounced lag, with its maximum impact occurring 2–3 hours later (lagged weight 0.31), consistent with the delayed output characteristics of wind power generation. These findings align with conclusions from previous studies on Beijing's summer electricity load, but this research provides more refined time-varying quantitative results through causal attention analysis.

To systematically assess robustness, the model's performance was evaluated across different seasonal periods within the test year. The results indicated consistently high accuracy and stable behaviour across summer, winter, and transitional seasons, with no statistically significant performance degradation observed in any specific period. This seasonal consistency underscores the model's reliability under diverse climatic operating conditions.

4.2.3 Extreme event prediction performance

To validate the model's robustness under extreme weather conditions, we focused on analysing its predictive performance during the cold wave event in January 2021 and the heatwave event in July 2021 (Figure 3). During the cold wave event (25–30 January), the DM-CA model achieved a MAPE of 3.8%, significantly lower than LSTM (5.2%) and CSDI (4.5%). In the heatwave event (12–18 July), DM-CA's MAPE was 3.2%, still outperforming LSTM (4.1%) and CSDI (3.7%). The *Precision@k* metric further reveals that DM-CA achieved a prediction accuracy of 92.3% for peak loads (top 5%), outperforming CSDI (89.7%) and LSTM (85.2%). This demonstrates that the diffusion model better captures uncertainty in extreme events through probabilistic generation, while causal attention ensures accurate quantification of meteorological factor influences. Although the current validation explicitly demonstrates the model's efficacy during heatwaves and cold snaps – two of the most common and impactful events for grid operation – its core strength lies in its principled, probabilistic approach to handling uncertainty. This foundational capability provides strong reason to believe that the model would generalise beneficially to other extreme events, such as severe storms or prolonged rainfall, although specific validation with corresponding event data would be a necessary and valuable next step.

Figure 2 Comparison of model prediction performance (MAE, RMSE) (see online version for colours)

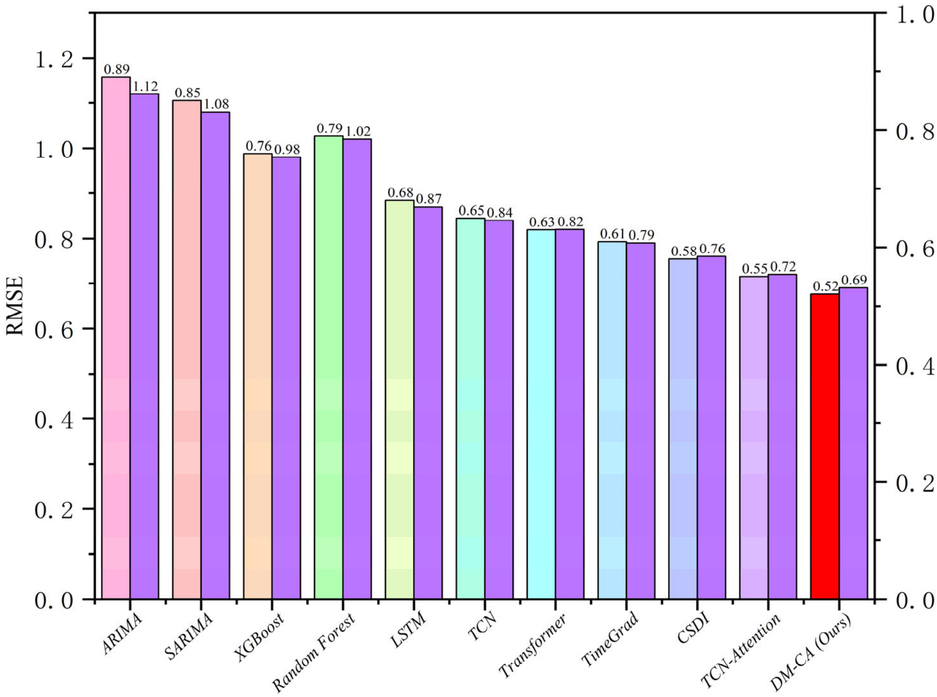
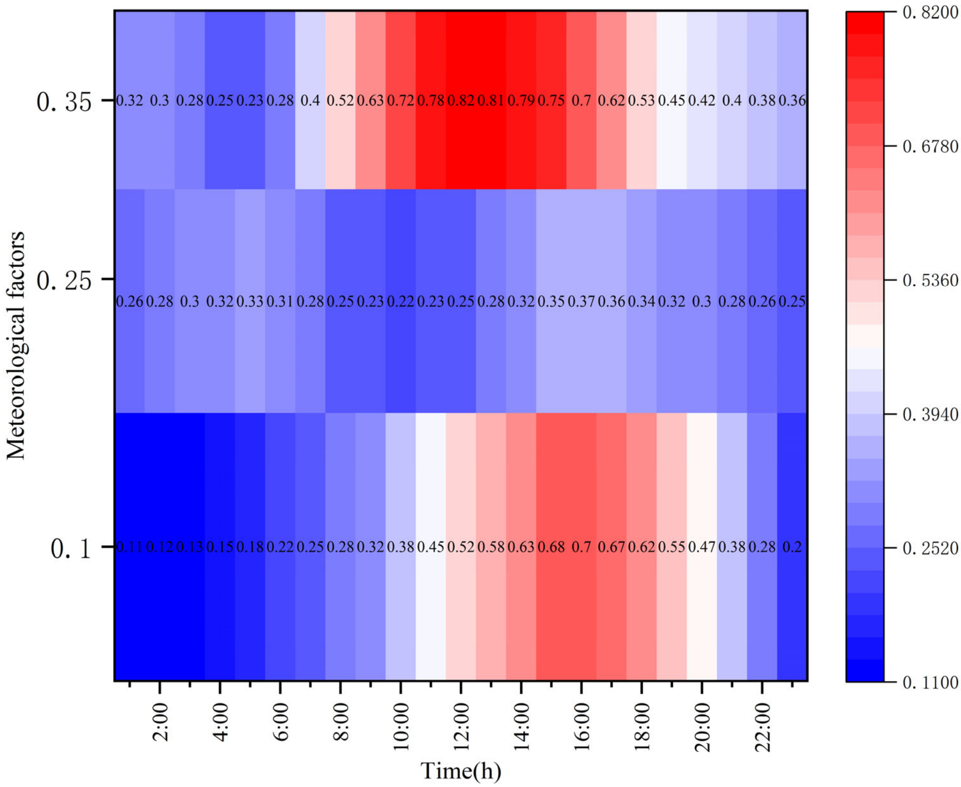


Table 2 Melting experiment results (MAE metric)

Model variants	*MAE*	*ΔAEe*
Full DM-CA	0.52	-
w/o diffusion module	0.61	0.173
w/o causal attention	0.57	0.096
w/o both	0.68	0.308
w/o temperature	0.65	0.25
w/o humidity	0.56	0.077
w/o wind speed	0.54	0.038

Figure 3 Causal attention weight distribution of meteorological factors (see online version for colours)



4.2.4 Melting experiment

To validate the contributions of each module, we designed ablation experiments (Table 2). The MAE of the complete DM-CA model was 0.52; removing the diffusion module (while retaining only causal attention) increased MAE to 0.61 (a 17.3% performance drop); removing the causal attention module (retaining only the diffusion model) increased MAE to 0.57 (a 9.6% performance drop); removing both (retaining only the LSTM encoder) significantly increased MAE to 0.68 (a 30.8% performance drop). This demonstrates that both the diffusion module and causal attention module are indispensable, with their synergistic interaction yielding the greatest performance improvement. Additionally, we tested the contribution of different meteorological factors: removing temperature increased MAE to 0.65 (a 25.0% performance drop), removing humidity increased MAE to 0.56 (a 7.7% drop), and removing wind speed increased MAE to 0.54 (a 3.8% drop). This confirms that temperature is the meteorological factor most significantly affecting load.

5 Conclusions

This study addresses the challenge of quantifying the impact of meteorological factors in electricity load forecasting by proposing an innovative approach that integrates diffusion models with causal attention. Through systematic validation on the New England grid public dataset, results demonstrate that this method not only significantly improves load forecasting accuracy – particularly exhibiting superior robustness during extreme weather events – but also effectively quantifies the causal influence of different meteorological factors. This provides a new analytical tool for understanding the complex relationship between weather and electricity load. Although the current study validates the model for short-term forecasting, its performance for longer prediction horizons presents an interesting avenue for future exploration.

In terms of theoretical contributions, this study's primary innovations manifest across three dimensions: First, it introduces diffusion models into the field of electricity load forecasting, leveraging their probabilistic generative properties to effectively handle the uncertainty and randomness inherent in meteorological data, thereby overcoming the limitations of traditional deterministic forecasting methods, overcoming limits of deterministic methods. Second, it designs a causal attention mechanism that ensures compliance with physical laws through temporal causal constraints, enabling refined quantification of the influence magnitude of meteorological factors. Finally, it establishes a synergistic framework integrating diffusion models and causal attention, combining high-precision forecasting capabilities with strong interpretability. This provides a novel technical paradigm for addressing time-series forecasting challenges in complex environments.

In practical applications, the methodology developed in this study offers significant support for power system operations. For power dispatch departments, its high-precision forecasting capability enhances clearing efficiency in day-ahead and intraday markets while reducing operational costs. For grid planners, the quantified results of meteorological impacts provide data support for coordinated planning of generation, transmission, load, and storage systems, particularly in scenarios with high renewable energy penetration. For emergency management departments, the model's robust performance during extreme weather events can strengthen grid resilience against climate change impacts. Practical implementation should prioritise the following aspects. Establishing standardised data collection systems for meteorological and load data, refining online model update mechanisms, developing user-friendly decision support interfaces, and formulating relevant operational procedures and standards.

Declarations

All authors declare that they have no conflicts of interest.

References

Aggarwal, S.K., Saini, L.M. and Kumar, A. (2009) 'Day-ahead price forecasting in Ontario electricity market using variable-segmented support vector machine-based model', *Electric Machines & Power Systems*, Vol. 37, No. 5, pp.495–516.

Bie, Z., Hu, G., Xie, H. and Li, G. (2014) 'Optimal dispatch for wind power integrated systems considering demand response', *Dianli Xitong Zidonghua/Automation of Electric Power Systems*, Vol. 38, No. 13, pp.115–120, p.159.

Cardoso, P.G., Raffaelli, D. and Pardal, M.A. (2008) 'The impact of extreme weather events on the Seagrass Zostera Noltii and related Hydrobia Ulvae population – ScienceDirect', *Marine Pollution Bulletin*, Vol. 56, No. 3, pp.483–492.

Crawley and Drury, B. (2008) 'Estimating the impacts of climate change and urbanization on building performance', *Journal of Building Performance Simulation*, Vol. 1, No. 2, pp.91–115.

Galphade, M., Nikam, V.B., Banerjee, B., Kiwelekar, A.W. and Sharma, P. (2025) 'Stacked LSTM for short-term wind power forecasting using multivariate time series data', *International Journal of Interactive Multimedia & Artificial Intelligence*, Vol. 9, No. 3, p.159.

Gangloff, H., Morales, K. and Petetin, Y. (2023) 'Deep parameterizations of pairwise and triplet Markov models for unsupervised classification of sequential data', *Computational Statistics & Data Analysis*, Vol. 3, No. 5, p.213.

Gottwald, T.R. (1992) 'Spatial and spatiotemporal autocorrelation analysis of citrus canker epidemics in citrus nurseries and groves in Argentina', *Phytopathology*, Vol. 82, No. 8, pp.843–843.

Gour, A.A., Singh, S.K., Tyagi, S.K. and Mandal, A. (2015) 'Variation in parameters of ambient air quality in National Capital Territory (NCT) of Delhi (India)', *Atmospheric & Climate Sciences*, Vol. 5, No. 1, pp.13–22.

Huang, C., Davis, L.S. and Townshend, J.R.G. (2002) 'An assessment of support vector machines for land cover classification', *International Journal of Remote Sensing*, Vol. 23, No. 4, pp.725–749.

Ishida, K.I., Shimogawa, Y., Sato, Y., Takano, K., Imayoshi, T. and Kojima, T. (2012) 'Demonstration tests of microgrid systems using renewable energy for small remote islands', *Atmospheric & Climate Sciences*, Vol. 2, No. 8, p.298.

Isobe, T., Makino, S. and Montgomery, D.B. (2008) 'Technological capabilities and firm performance: the case of small manufacturing firms in Japan', *Asia Pacific Journal of Management*, Vol. 25, No. 3, p.413.

Kadi, M., Njau, L.N., Mwikya, J. and Kamga, A. (2011) 'The state of climate information services for agriculture and food security in East African countries. CCAFS Working Paper No. 5', *CGIAR Research Program on Climate Change Agriculture & Food Security*, Vol. 7, No. 4, p.28.

Khashei, M., Bijari, M. and Hejazi, S.R. (2012) 'Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting', *Soft Computing*, Vol. 16, No. 6, pp.1091–1105.

Langland, R.H., Shapiro, M.A. and Gelaro, R. (1947) 'Initial condition sensitivity and error growth in forecasts of the 25 January 2000 east coast snowstorm', *Monthly Weather Review*, Vol. 130, No. 4, pp.957–974.

Li, Y., Yang, Z.Y., Zhang, S., Mao, R., Ye, L. and Liu, Y. (2025) 'TCN-MBMARESNET: a novel fault diagnosis method for small marine rolling bearings based on time convolutional neural network in tandem with multi-branch residual network', *Measurement Science & Technology*, Vol. 12, No. 2, p.36.

Maury, O., Shin, Y-J., Faugeras, B., Ari, T.B. and Marsac, F. (2007) 'Modeling environmental effects on the size-structured energy flow through marine ecosystems. Part 2: simulations', *Progress in Oceanography*, Vol. 2, No. 7, p.391.

Puntharod, R., Onsomsuay, K., Pookmanee, P. and Kumchompoo, J. (2024) 'Exploring alkali hydroxide influence on calcium titanate formation for application in biodiesel catalysts', *Bulletin of Chemical Reaction Engineering & Catalysis*, Vol. 19, No. 3, p.169.

Sarra, A.D., Cacciani, M., Chamard, P., Cornwall, C., Deluisi, J.J., Iorio, T.D., Disterhoft, P., Fiocco, G.D.F. and Monteleone, F. (2002) 'Effects of desert dust and ozone on the ultraviolet irradiance at the Mediterranean island of Lampedusa during Paur II', *Journal of Geophysical Research: Atmospheres*, Vol. 15, No. 2, p.713.

Schmalwieser, A.W., Schauberger, G. and Janouch, M. (2003) 'Temporal and spatial variability of total ozone content over Central Europe: analysis in respect to the biological effect on plants', *Agricultural & Forest Meteorology*, Vol. 120, Nos. 1–4, pp.9–26.

Shu-Wei, D., Xiao-Guang, Y., Meng, Z., Yong, L.I., Wen-Feng, W. and Zhi-Juan, L. (2011) 'Changes of China agricultural climate resources under the background of climate change: IX. Spatiotemporal change characteristics of china agricultural climate resources', *Chinese Journal of Applied Ecology*, Vol. 21, No. 10, pp.3177–3188.

Smith, B.J., Toeroek, A., Mcalister, J.J. and Megarry, Y. (2003) 'Observations on the factors influencing stability of building stones following contour scaling: a case study of Oolitic Limestones from Budapest, Hungary', *Building & Environment*, Vol. 38, Nos. 9/10, pp.1173–1183.

Takahashi, Y., Hikita, K. and Asada, M. (2003) 'A hierarchical multi-module learning system based on self-interpretation of instructions by coach', *Ecological Complexity*, Vol. 2, No. 17, p.381.

Tian-Qing, D.U., Guo-Yuan, M., Jian-Ping, H., Jin-Zhong, Y., Fu-Zhu, C. and Jian-Fu, X. (2012) 'Root growth characteristics of four species leguminous forage in immature loess subsoil', *Acta Agrestia Sinica*, Vol. 8, No. 3, p.581.

Walsh, M.G., Fenner, D.B. and Winkelman, D.L. (2002) 'Comparison of an electric seine and prepositioned area electrofishers for sampling stream fish communities', *North American Journal of Fisheries Management*, Vol. 22, No. 1, pp.77–85.

Wang, H. and Raj, B. (2015) 'A survey: time travel in deep learning space: an introduction to deep learning models and how deep learning models evolved from the initial ideas', *Computer Ence*, Vol. 226, Nos. 1–4, pp.23–34.

Wang, X., Magableh, S.K., Dawaghreh, O., Wang, C., Gong, J., Zhao, Z. and Liao, M.H. (2024) 'Deep learning-based electricity price forecast for virtual bidding in wholesale electricity market', *Progress in Oceanography*, Vol. 13, No. 7, p.683.

Wu, J., Jenerette, G.D., Buyantuyev, A. and Redman, C.L. (2011) 'Quantifying spatiotemporal patterns of urbanization: the case of the two fastest growing metropolitan regions in the united states', *Ecological Complexity*, Vol. 8, No. 1, pp.1–8.