

Article

MIRA-Mamba: Multi-Scale Interactive Recalibrated Mamba for Multi-Step Indoor Temperature Forecasting in HVAC Systems

Mengyuan Jing ^{1,2} , Daiwei Ruan ³, Tiejun Sun ³, Chunlei Wu ^{1,2,*} and Leiquan Wang ^{1,2}

¹ Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), No. 66, Changjiang West Road, Qingdao 266580, China; z24070128@s.upc.edu.cn (M.J.); wangleiquan@upc.edu.cn (L.W.)

² Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, China University of Petroleum (East China), No. 66, Changjiang West Road, Qingdao 266580, China

³ R&D Center, Qingdao Hisense Hitachi Air-conditioning Systems Co., Ltd., Qingdao 266510, China; ruandaiwei@hisensehitachi.com (D.R.); suntiejun@hisensehitachi.com (T.S.)

* Correspondence: wuchunlei@upc.edu.cn

Abstract

For multivariate-input-driven multi-step temperature forecasting in air-conditioning systems, existing models often rely on a single temporal receptive field, limiting the amount of temporal information captured and hindering the effective characterization of multi-scale dynamic features. Moreover, accurately capturing condition-dependent variations in the contributions of different variables to the prediction target remains challenging, resulting in persistent error accumulation over long forecasting horizons. To address these issues, this paper proposes a Multi-scale Interactive Recalibrated Mamba (MIRA-Mamba) forecasting framework. The multi-scale temporal interaction embedding module extracts and fuses dynamic features from different temporal receptive fields, enhancing the representation of multivariate coupling relationships and alleviating information loss caused by single-scale representations. The bidirectional state-space recalibration encoder employs bidirectional Mamba to capture long-range temporal dependencies and incorporates a feature contribution recalibration mechanism to adaptively adjust variable weights, thereby strengthening high-contribution variables and suppressing interference from low-contribution variables. In addition, a private dataset, Hisense-AC, is constructed from real air-conditioning operation logs collected from 15 cities over half a year. Experimental results show that MIRA-Mamba achieves the lowest forecasting errors across all four operating modes on the Hisense-AC dataset, with a maximum MSE reduction of 5.4% compared with the best-performing baseline. On six public time-series datasets, MIRA-Mamba obtains the best MSE in 17 out of 24 forecasting settings, demonstrating its effectiveness in real HVAC scenarios and its generalization capability on public benchmarks.



Academic Editor: Constantinos A. Balaras

Received: 5 June 2026

Revised: 1 July 2026

Accepted: 5 July 2026

Published: 8 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: air-conditioning system; feature contribution recalibration; time series forecasting; Mamba

1. Introduction

Building operations account for approximately 30% of global final energy consumption [1]. Among building subsystems, heating, ventilation, and air conditioning (HVAC) systems are responsible for a major portion of heating and cooling loads, making them an important target for energy-efficiency optimization. Intelligent HVAC operation often requires anticipating future indoor temperature variations, and reliable future temperature

trajectories can support downstream control decision-making [2,3]. Motivated by this requirement, this study aims to reduce multi-step indoor temperature forecasting errors compared with competitive baselines while maintaining stable performance across different HVAC operating modes.

Compared with general time series forecasting, HVAC indoor temperature forecasting involves stronger fast–slow coupling and operating-condition dependence. Indoor temperature is jointly affected by weather, user settings, equipment states, and historical thermal conditions, leading to both short-term fluctuations and delayed thermal responses. These coupled effects result in multi-scale temperature dynamics and mode-dependent variable contributions.

Existing studies have explored various forecasting methods for building and HVAC applications. Traditional statistical models usually rely on linearity and stationarity assumptions, limiting their ability to capture nonlinear thermal dynamics, operating-mode transitions, and multi-source disturbances [4,5]. To improve nonlinear temporal modeling, data-driven methods have been increasingly applied to building temperature and energy forecasting. LSTM-based models have been used for multi-step or multi-zone indoor temperature prediction [6–8], while CNN-LSTM architectures and attention mechanisms further improve forecasting performance by combining local dynamic feature extraction with salient temporal information modeling [9,10]. Although these methods improve nonlinear representation, recurrent compression, local extraction, or fixed mappings still weaken distant historical information, limiting the characterization of delayed indoor temperature responses.

To further address long-range dependency retention and error accumulation, general long-sequence forecasting models have introduced efficient attention and decomposition-based architectures to improve long-term temporal modeling [11,12], while frequency-domain and linear forecasting designs provide alternative ways to capture periodic and trend patterns [13,14]. More recent studies further enhance complex temporal pattern modeling through structured temporal representations and cross-scale or cross-variable interaction mechanisms [15–17]. These methods improve global temporal modeling in general forecasting tasks. However, they are mainly designed for generic time series and usually lack targeted modeling of HVAC event-driven operation, fast-slow thermal response superposition, and operating-mode-dependent variable contributions. Therefore, directly applying them to HVAC multi-step temperature forecasting may still lead to insufficient historical dependency utilization and fixed variable-contribution representation.

Selective state–space models provide a promising direction for efficient long-sequence modeling. Mamba employs an input-dependent selection mechanism to capture long-range dependencies while maintaining linear computational complexity [18]. This property makes Mamba a suitable foundation for modeling delayed indoor temperature responses. However, the original Mamba mainly performs selective state updates on a single sequence representation, without explicitly separating short-term control responses from slow-varying thermal dynamics. Its relatively fixed feature and channel-relationship modeling also limit adaptation to operation-dependent variable contributions.

To address these limitations, this paper proposes a Multi-scale Interactive Recalibrated Mamba framework, termed MIRA-Mamba, for indoor temperature prediction of HVAC systems. The framework embeds HVAC response characteristics into Mamba-based forecasting by considering control-induced short-term variations, historical delayed thermal effects, and operating-condition-dependent variable relevance. By integrating multi-scale temporal interaction, bidirectional state–space dependency modeling, and feature contribution recalibration, MIRA-Mamba links efficient sequence modeling with indoor temperature evolution characteristics and provides a targeted solution for robust forecasting under diverse HVAC operating modes. Our contributions lie in three aspects:

- A multi-scale temporal interaction embedding module is designed to extract, align, and fuse features from different temporal receptive fields, thereby enhancing the representation of rapid control-induced temperature variations, slow-varying delayed thermal responses, and cross-variable coupling relationships.
- A bidirectional state-space recalibration encoder is developed to capture long-range temporal dependencies using Mamba and strengthen the modeling of delayed indoor temperature responses through bidirectional encoding. A feature contribution recalibration mechanism is further introduced to adaptively adjust the contributions of control, environmental, and thermal-state variables under different operating conditions.
- Extensive experiments are conducted on the real-world Hisense-AC dataset and six public time-series benchmarks. Results show that MIRA-Mamba consistently reduces forecasting errors across different HVAC operating modes, with MSE reductions of 4.0% and 5.4% in Heating_on and Heating_off, respectively, and achieves the best MSE in 17 out of 24 public forecasting settings.

2. Related Work

Existing studies on building and HVAC forecasting have explored statistical models, machine-learning methods, and hybrid physical-data models to approximate building thermal dynamics. However, these methods often rely on predefined structures, handcrafted features, or simplified assumptions, making them less effective for nonlinear and delayed temperature responses under complex operating conditions [4,5].

Deep learning methods have been introduced to improve thermal dynamics modeling. Xu et al. improved indoor temperature prediction in public buildings, while Mtibaa et al. and Fang et al. explored LSTM-based multi-step and multi-zone prediction frameworks [6–8]. Although LSTM-based models can capture nonlinear temporal dependencies, they compress historical information into recurrent hidden states, which weakens the representation of long-lasting delayed effects and short-term fluctuations. Recent studies further considered practical deployment issues such as data availability, prediction uncertainty, and computational efficiency [19–21]. Other works introduced physical constraints, transfer learning, or graph structures to enhance generalization and spatial-temporal coupling [22–25].

Beyond HVAC-specific studies, long-term time series forecasting models provide useful references. Informer reduces attention complexity, while Autoformer and FEDformer introduce decomposition, auto-correlation, and frequency-domain modeling to capture long-term patterns [11–13]. Later studies reconsidered forecasting structures: DLinear showed the effectiveness of decomposed linear models, and PatchTST improved local context modeling through patch representation [14,15]. Recent models further emphasize multi-scale patterns and cross-variable interactions. iTransformer strengthens variate-centric modeling, while TimeMixer improves forecasting through decomposable multi-scale mixing [16,17]. TimesNet and Crossformer enhance temporal representation through period modeling and cross-dimension dependencies [26,27], and MICN and SCINet further explore multi-scale context modeling and hierarchical interaction [28,29].

Structured state-space models offer another efficient route for long-sequence modeling. S4 and S5 enhance long-range dependency modeling through state-space parameterization [30,31], while Mamba introduces input-dependent selective state transitions with linear complexity [18]. Recent studies have also introduced Mamba into time series forecasting to exploit its selective state modeling ability for multivariate long-sequence prediction [32–34]. Further efforts have extended Mamba-based forecasting from the perspectives of training optimization and multi-scale temporal modeling [35,36]. Overall, these methods provide useful foundations for efficient long-sequence forecasting, but their adaptation to

HVAC event-driven temperature forecasting remains limited, especially in jointly modeling multi-scale thermal responses and operating-condition-dependent variable relevance.

3. Materials and Methods

3.1. Task Definition

HVAC temperature trajectory forecasting aims to predict the future target temperature sequence from historical multivariate observations. This task can support downstream HVAC operation analysis, while this study focuses on improving forecasting accuracy and stability under different operating modes.

Let the observation vector at time step t be denoted as $\mathbf{x}_t \in \mathbb{R}^N$, where N represents the dimensionality of system sensor readings, control variables, environmental variables, and other relevant factors. Given a historical window length L , the input sequence is defined as

$$\mathbf{X}_t = [\mathbf{x}_{t-L+1}, \mathbf{x}_{t-L+2}, \dots, \mathbf{x}_t] \in \mathbb{R}^{L \times N}. \quad (1)$$

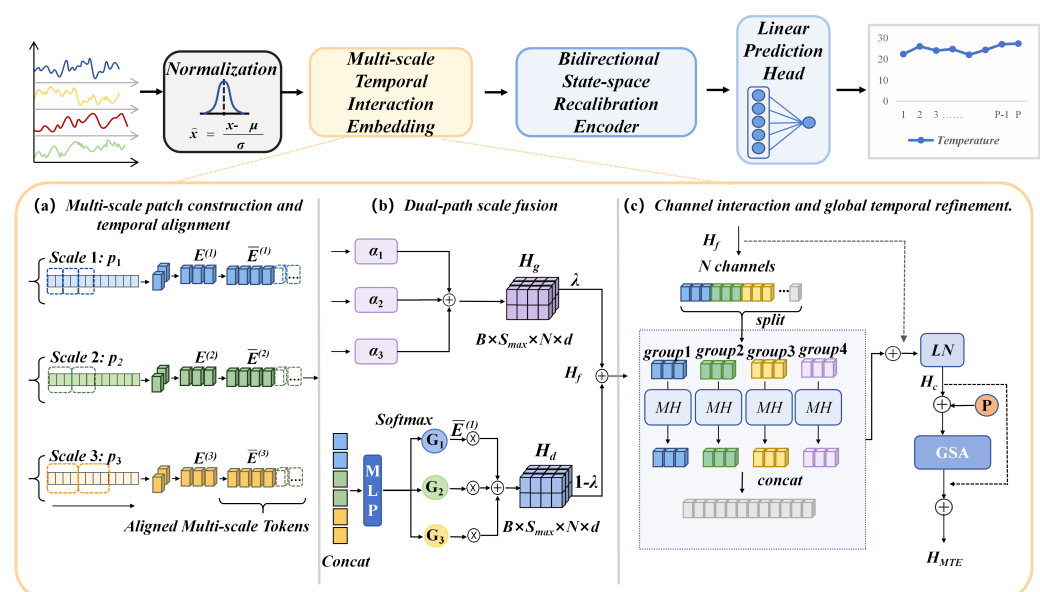
The forecasting task aims to predict the future trajectories of one or more target variables from historical multivariate observations. Let N_y denote the number of forecasting target variables. The forecasting objective is to predict the future P -step sequence:

$$\hat{\mathbf{Y}}_t = [\hat{\mathbf{y}}_{t+1}, \hat{\mathbf{y}}_{t+2}, \dots, \hat{\mathbf{y}}_{t+P}] \in \mathbb{R}^{P \times N_y}, \quad (2)$$

where $\hat{\mathbf{y}}_{t+i} \in \mathbb{R}^{N_y}$ denotes the predicted target vector at the i -th forecasting step. For the Hisense-AC dataset, $N_y = 1$, whereas for the public benchmark datasets, $N_y = N$.

3.2. Model Architecture

As illustrated in Figure 1, MIRA-Mamba takes historical multivariate time series as input and extracts multi-scale temporal representations through the Multi-scale Temporal Interaction Embedding (MTE) module, as described in Section 3.3. The resulting embeddings are then processed by the bidirectional state-space recalibration encoder to model long-range temporal dependencies and dynamically recalibrate variable contributions across operating conditions, as described in Section 3.4. Finally, a linear prediction head generates the future indoor temperature trajectory.



3.3. Multi-Scale Temporal Interaction Embedding (MTE)

HVAC temperature sequences exhibit short-term fluctuations, medium-scale operational variations, and long-term thermal inertia. A single temporal scale is therefore insufficient to capture the complex dynamics jointly influenced by environmental disturbances and indoor thermal states. To address this limitation, we design a Multi-scale Temporal Interaction Embedding (MTE) module to learn complementary representations across different temporal receptive fields and enhance the modeling of scale variations and cross-variable dependencies.

Multi-scale patch construction and temporal alignment. Motivated by the multi-resolution modeling of dynamic thermal processes [15,36], we construct multi-scale patch sequences along the temporal dimension from the batched HVAC input tensor $\mathbf{X} \in \mathbb{R}^{B \times L \times N}$, where B , L , and N denote the batch size, historical window length, and number of variables, respectively. For the k -th temporal scale, a sliding window with patch length p_k and stride s is applied to unfold the input sequence:

$$\mathbf{U}^{(k)} = \text{Unfold}_{p_k, s}(\mathbf{X}) \in \mathbb{R}^{B \times S_k \times N \times p_k}, \quad (3)$$

where $\mathbf{U}^{(k)}$ denotes the patch sequence at the k -th temporal scale, and $\text{Unfold}_{p_k, s}(\cdot)$ represents sliding-window unfolding along the temporal dimension. The resulting patches capture local variations across different temporal receptive fields, and their number is

$$S_k = \left\lfloor \frac{L - p_k}{s} \right\rfloor + 1, \quad (4)$$

where $\lfloor \cdot \rfloor$ denotes the floor operation.

Since different patch lengths p_k produce different numbers of patches S_k , the resulting scale-specific representations have inconsistent token lengths and cannot be directly fused. Therefore, each patch sequence is independently projected into a unified hidden dimension d :

$$\mathbf{E}^{(k)} = \mathbf{U}^{(k)} \mathbf{W}^{(k)} + \mathbf{b}^{(k)} \in \mathbb{R}^{B \times S_k \times N \times d}, \quad (5)$$

where $\mathbf{W}^{(k)} \in \mathbb{R}^{p_k \times d}$ and $\mathbf{b}^{(k)} \in \mathbb{R}^d$ are the projection weight and bias for the k -th temporal scale, respectively.

Then, to keep the token length consistent across different scales, all scale-specific patch sequences are padded to a unified length $S_{\max}$ by repeating the last token:

$$\bar{\mathbf{E}}^{(k)} = \text{PadLast}_{S_{\max}}(\mathbf{E}^{(k)}) \in \mathbb{R}^{B \times S_{\max} \times N \times d}, \quad (6)$$

where $S_{\max} = \max_{1 \leq k \leq K} S_k$, K denotes the number of temporal scales, and $\text{PadLast}_{S_{\max}}(\cdot)$ represents token-length alignment by repeating the last valid token. Unlike zero padding, this strategy preserves the final valid local representation and avoids introducing artificial zero-valued tokens, which reduce discontinuities during multi-scale fusion. Through scale-specific projection and token-length alignment, the patch sequences are transformed into consistent multi-scale embeddings for subsequent scale fusion and cross-variable interaction.

Dual-path scale fusion. Although multi-scale patches provide local representations across different temporal receptive fields, the relative importance of these scales varies with operating states, control strategies, and environmental disturbances. Naive averaging or direct concatenation cannot adequately capture global scale preferences while adapting to local dynamic changes. Therefore, we introduce a dual-path scale fusion strategy that combines global scale weighting with dynamic gating to model both stable scale preferences and locally adaptive scale contributions.

First, learnable scale logits are used to estimate the global importance of different temporal scales, and the scale-specific features are fused as

$$\alpha_k = \frac{\exp(a_k)}{\sum_{j=1}^K \exp(a_j)}, \quad \mathbf{H}_g = \sum_{k=1}^K \alpha_k \bar{\mathbf{E}}^{(k)}, \quad (7)$$

where a_k is the learnable logit for the k -th temporal scale, α_k is the corresponding normalized global scale weight, and $\mathbf{H}_g$ denotes the globally fused representation. This branch learns relatively stable scale preferences from the training distribution and provides a global prior for subsequent multi-scale fusion.

However, the global scale weights are shared across all samples, temporal positions, and variables, which limits their ability to capture scale differences under local operating states. To enhance adaptive scale selection at different temporal and variable positions, we further concatenate the scale-specific features along the hidden dimension:

$$\mathbf{C} = \text{Concat}(\bar{\mathbf{E}}^{(1)}, \bar{\mathbf{E}}^{(2)}, \dots, \bar{\mathbf{E}}^{(K)}), \quad (8)$$

where $\mathbf{C}$ denotes the concatenated multi-scale feature tensor, and $\text{Concat}(\cdot)$ denotes the concatenation operation along the hidden dimension.

Based on the concatenated multi-scale representation, a gating network is employed to generate dynamic scale weights:

$$\mathbf{G} = \text{Softmax}(\text{Gate}(\mathbf{C})), \quad (9)$$

where $\mathbf{G} \in \mathbb{R}^{B \times S_{\max} \times N \times K}$ denotes the dynamic scale-weight tensor, $\text{Gate}(\cdot)$ is the gating network, and $\text{Softmax}(\cdot)$ is applied along the scale dimension. Unlike the global scale weights, the dynamic scale weights adaptively assign scale contributions for each sample, temporal token, and variable position.

The dynamic fusion result is then obtained by

$$\mathbf{H}_d = \sum_{k=1}^K \mathbf{G}^{(k)} \odot \bar{\mathbf{E}}^{(k)}, \quad (10)$$

where $\mathbf{H}_d$ denotes the dynamically gated representation, $\mathbf{G}^{(k)}$ is the dynamic weight of the k -th scale for different samples, temporal tokens, and variables, and $\odot$ denotes element-wise multiplication with broadcasting along the hidden dimension.

Finally, the global and dynamic fusion results are combined to obtain the dual-path scale fusion output:

$$\mathbf{H}_f = \text{LN}(\lambda \mathbf{H}_g + (1 - \lambda) \mathbf{H}_d), \quad (11)$$

where $\mathbf{H}_f$ denotes the fused representation, $\text{LN}(\cdot)$ is LayerNorm, and λ is a fixed trade-off coefficient balancing the contributions of the two fusion branches. In all experiments, λ is empirically set to 0.6 and remains unchanged during training. Through the cooperation of global scale weighting and dynamic gating, this strategy preserves the overall scale preference while enabling local scale adaptation, thereby improving the stability and flexibility of multi-scale feature fusion.

Channel interaction and global temporal refinement. After dual-path scale fusion, the model obtains a unified multi-scale temporal representation. However, sensor readings, control variables, and environmental variables in HVAC systems usually exhibit strong coupling relationships. Modeling only temporal multi-scale patterns is therefore insufficient to fully capture inter-variable dependencies. To enhance cross-variable relationship modeling, we introduce grouped channel attention based on the fused representation:

$$\mathbf{H}_c = \text{LN}(\mathbf{H}_f + \text{GCA}(\mathbf{H}_f)), \quad (12)$$

where $\mathbf{H}_c$ denotes the representation after grouped channel attention, and $\text{GCA}(\cdot)$ represents the grouped channel attention operation. Specifically, the variable dimension is divided into several groups, and multi-head attention is applied independently within each group. The resulting group-wise representations are then concatenated in the original variable order. This design captures inter-variable dependencies within each group while maintaining computational efficiency.

Although channel interaction enhances cross-variable dependency modeling, semantic discrepancies still exist among local representations formed at different temporal scales. To further mitigate scale-level representation inconsistency and exploit complementary information across different temporal receptive fields, we introduce a global self-attention mechanism with learnable positional encoding to supplement temporal positional information in the patch sequence:

$$\mathbf{H}_{\text{MTE}} = \text{LN}(\mathbf{H}_c + \beta \cdot \text{GSA}(\mathbf{H}_c + \mathbf{P})), \quad (13)$$

where $\text{GSA}(\cdot)$ denotes the global self-attention operation, $\mathbf{P}$ is the learnable positional encoding, β is the residual weight of the global temporal refinement branch, and $\mathbf{H}_{\text{MTE}}$ is the final output of the multi-scale temporal interaction embedding module.

3.4. Bidirectional State-Space Recalibration Encoder

HVAC temperature dynamics often exhibit delayed responses, as environmental disturbances and historical thermal states not only affect the current temperature but also influence future trajectories over extended periods. Local convolution and short-term temporal modeling are therefore insufficient to capture such long-range dependencies. Moreover, the contributions of input variables vary across operating stages. Treating all variables equally amplifies low-contribution or less informative features during multi-step forecasting and reduces prediction stability.

To address these issues, we design a bidirectional state-space recalibration encoder that combines long-range temporal modeling with adaptive feature recalibration. The encoder captures historical dependencies through bidirectional state evolution and adjusts variable contributions through a sequential feature contribution recalibration process consisting of variable-importance gating and group-level affinity modeling. This design enhances high-contribution features, attenuates low-contribution features, and captures structured dependencies among variable groups.

As shown in Figure 2, the Multi-scale Temporal Interaction Embedding (MTE) module extracts local dynamics from multiple temporal receptive fields and models cross-variable interactions. To further capture long-delayed thermal responses in HVAC temperature dynamics, an interactive bidirectional Mamba-GRU encoder is adopted as the temporal backbone for the multi-scale embedded representations.

Given the output of the MTE module $\mathbf{H}_{\text{MTE}} \in \mathbb{R}^{B \times S_{\max} \times N \times d}$, it is reshaped into $\tilde{\mathbf{H}} \in \mathbb{R}^{(BN) \times S_{\max} \times d}$ to enable variable-wise temporal encoding along the aligned patch dimension.

The forward Mamba branch processes the aligned patch sequence along the chronological direction:

$$\vec{\mathbf{m}}_\tau = \text{Mamba}_f(\tilde{\mathbf{H}}_\tau), \quad \tau = 1, 2, \dots, S_{\max}, \quad (14)$$

where $\tilde{\mathbf{H}}_\tau \in \mathbb{R}^{(BN) \times d}$ is the input representation at the τ -th aligned patch token, and $\vec{\mathbf{m}}_\tau \in \mathbb{R}^{(BN) \times d}$ denotes the forward Mamba state.

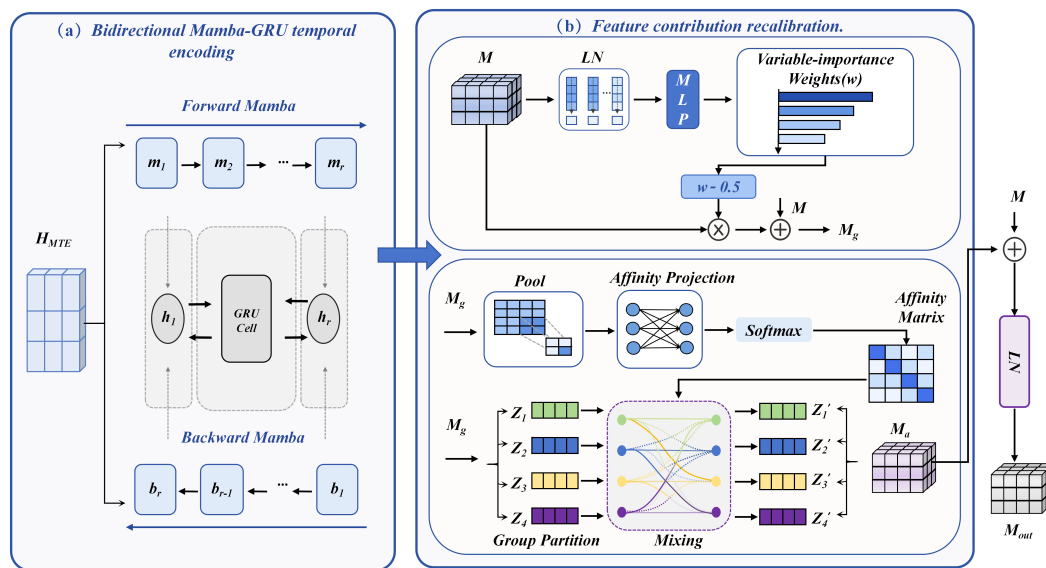


Figure 2. Detailed structure of the bidirectional state-space recalibration encoder.

Similarly, the backward Mamba branch processes the aligned patch sequence in the reverse direction:

$$\overleftarrow{\mathbf{b}}_{\tau} = \text{Mamba}_b(\tilde{\mathbf{H}}_{\tau}), \quad \tau = S_{\max}, S_{\max} - 1, \dots, 1, \quad (15)$$

where $\overleftarrow{\mathbf{b}}_{\tau} \in \mathbb{R}^{(BN) \times d}$ denotes the backward Mamba state at the τ -th aligned patch token. Both directional branches follow the Mamba selective state-space architecture [18], enabling long-range dependency modeling along the aligned patch sequence.

To exchange information between the two directional temporal streams, the corresponding forward and backward Mamba states are updated through an interactive GRU-style gated cell:

$$\begin{bmatrix} \overrightarrow{\mathbf{h}}_{\tau}, \overleftarrow{\mathbf{h}}_{\tau} \end{bmatrix} = \text{IGRU} \left(\begin{bmatrix} \overrightarrow{\mathbf{m}}_{\tau}, \overleftarrow{\mathbf{b}}_{\tau} \end{bmatrix}, \begin{bmatrix} \overrightarrow{\mathbf{h}}_{\tau-1}, \overleftarrow{\mathbf{h}}_{\tau+1} \end{bmatrix} \right), \quad (16)$$

where $\overrightarrow{\mathbf{h}}_{\tau}$ and $\overleftarrow{\mathbf{h}}_{\tau}$ denote the left-to-right and right-to-left hidden states updated at the τ -th aligned patch token, respectively. $\text{IGRU}(\cdot)$ denotes an interactive GRU-style gated updating function that follows the gated recurrent updating mechanism [37] to regulate information exchange between the two directional state-space streams.

After bidirectional temporal encoding, the final bidirectional representation is obtained by combining the last left-to-right hidden state and the first right-to-left hidden state:

$$\mathbf{O} = \text{Reshape} \left(\overrightarrow{\mathbf{h}}_{S_{\max}} + \overleftarrow{\mathbf{h}}_1 \right), \quad (17)$$

where $\mathbf{O} \in \mathbb{R}^{B \times N \times d}$ denotes the bidirectional temporal representation after mapping the encoded sequence back to the variable dimension.

Finally, layer normalization and dropout are applied to obtain the encoder output:

$$\mathbf{M} = \text{Dropout}(\text{LN}(\mathbf{O})), \quad (18)$$

where $\mathbf{M} \in \mathbb{R}^{B \times N \times d}$ is the channel-wise temporal representation generated by the bidirectional Mamba-GRU temporal encoder.

Feature contribution recalibration (FCR). Although bidirectional state–space encoding can extract long-range temporal dependencies, the contributions of different variables still change dynamically across operating stages. To enhance high-contribution variables and attenuate low-contribution features, variable-wise importance weights are first estimated for each sample.

$$\mathbf{w} = \sigma(\text{MLP}_g(\text{LN}(\mathbf{M}))), \quad (19)$$

where $\mathbf{w} \in \mathbb{R}^{B \times N}$ denotes the variable-importance weights, $\sigma(\cdot)$ denotes the sigmoid activation function, and $\text{MLP}_g(\cdot)$ is a lightweight gating network that generates adaptive weights for different variables. This gating network assigns one importance coefficient to each variable. A larger weight indicates that the corresponding variable makes a higher contribution to the current sample.

Based on the estimated variable-importance weights, residual gating is applied to recalibrate the encoded features:

$$\mathbf{M}_g = \mathbf{M} + \gamma(\mathbf{w} - 0.5) \odot \mathbf{M}, \quad (20)$$

where $\mathbf{M}_g$ denotes the recalibrated representation after variable-importance gating, $\mathbf{w}$ assigns one importance coefficient to each variable and is shared across the hidden dimension when modulating $\mathbf{M} \in \mathbb{R}^{B \times N \times d}$, $\gamma = \sigma(\gamma_{\text{raw}})$ is a bounded gating intensity coefficient, 0.5 is the gating center, and $\odot$ denotes element-wise multiplication. This bounded residual modulation adjusts variable responses around the original feature scale while maintaining a positive scaling factor.

However, variable-wise gating mainly focuses on the importance of individual variables and cannot fully capture structured dependencies among variables. In HVAC systems, variables are usually not independent. For example, control variables, environmental factors, and thermal states exhibit strong group-wise correlations. To further model such structured dependencies, group-level affinity modeling is introduced based on the recalibrated representation $\mathbf{M}_g$.

First, a statistical descriptor is obtained from $\mathbf{M}_g$ through pooling, and a group-level affinity matrix is generated by an affinity network:

$$\mathbf{A} = \text{Softmax}\left(\frac{\text{Aff}(\text{Pool}(\mathbf{M}_g))}{\tau}\right), \quad (21)$$

where $\mathbf{A} \in \mathbb{R}^{N_g \times N_g}$ denotes the group-level affinity matrix of one sample, N_g is the number of variable groups, $\text{Pool}(\cdot)$ is the pooling operation used to obtain global statistical representations, and τ is the temperature coefficient that controls the smoothness of the affinity distribution. The affinity network $\text{Aff}(\cdot)$ maps the pooled descriptor to N_g^2 group-level affinity scores, which are organized as an $N_g \times N_g$ affinity score matrix. The Softmax operation is applied along the last dimension of this matrix, making each row a weight distribution over all input groups, i.e., $\sum_{j=1}^{N_g} \mathbf{A}_{ij} = 1$. The i -th row of $\mathbf{A}$ indicates the relative dependency strengths of different input groups when reconstructing the i -th output group.

Meanwhile, $\mathbf{M}_g$ is divided into N_g variable groups along the variable or channel dimension:

$$\mathcal{Z} = \text{Group}(\mathbf{M}_g) = \{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_{N_g}\}, \quad (22)$$

where $\mathcal{Z}$ denotes the set of grouped variable features, $\mathbf{Z}_i$ denotes the feature representation of the i -th variable group, and $\text{Group}(\cdot)$ denotes the group partition operation. It should be noted that $\mathbf{M}_g$ is used in two parallel paths in this stage. One path generates the affinity

matrix $\mathbf{A}$ through pooling and affinity projection, while the other path provides the grouped features $\{\mathbf{Z}_i\}_{i=1}^{N_g}$ to be mixed.

Based on the affinity matrix $\mathbf{A}$, the i -th reconstructed group is obtained by aggregating all input groups with affinity-guided weights:

$$\mathbf{Z}'_i = \sum_{j=1}^{N_g} \mathbf{A}_{ij} \mathbf{Z}_j, \quad i = 1, 2, \dots, N_g, \quad (23)$$

where $\mathbf{Z}'_i$ denotes the i -th reconstructed variable group, and $\mathbf{A}_{ij}$ represents the affinity weight from the j -th input group to the i -th output group. This formulation indicates that the affinity matrix does not serve as the output feature itself, but acts as the group-wise mixing weights to reconstruct grouped features derived from $\mathbf{M}_g$.

The reconstructed groups are then concatenated to obtain the affinity-modulated representation:

$$\mathbf{M}_\alpha = \text{Concat}(\mathbf{Z}'_1, \mathbf{Z}'_2, \dots, \mathbf{Z}'_{N_g}), \quad (24)$$

where $\mathbf{M}_\alpha$ denotes the reconstructed representation after group-level affinity modeling, and $\text{Concat}(\cdot)$ denotes the concatenation operation along the channel or variable-group dimension. This process supplements group-wise structural information on top of variable-importance gating, allowing the model to consider not only whether each variable is important but also how different variable groups interact with each other.

Finally, the affinity-modulated representation is fused with the original encoded features through a residual connection, followed by normalization:

$$\mathbf{M}_{\text{out}} = \text{LN}(\mathbf{M} + \mathbf{M}_\alpha), \quad (25)$$

where $\mathbf{M}_{\text{out}}$ is the final output of the bidirectional state-space recalibration encoder, and $\text{LN}(\cdot)$ denotes layer normalization. Since $\mathbf{M}_\alpha$ is reconstructed from $\mathbf{M}_g$ through group-level affinity mixing, it contains the group-wise dependency information after variable contribution recalibration. Therefore, the final residual fusion uses the original encoded representation $\mathbf{M}$ as the main branch and injects $\mathbf{M}_\alpha$ as a structure-enhanced branch, thereby preserving the basic temporal representation while introducing variable-aware group-level interaction information.

3.5. Prediction Head and Denormalization

After the bidirectional state-space recalibration encoder, a linear prediction head maps the high-level temporal representation to the required forecasting horizon and output dimension:

$$\hat{\mathbf{Y}}_t = \text{Head}(\mathbf{M}_{\text{out}}), \quad (26)$$

where $\hat{\mathbf{Y}}_t \in \mathbb{R}^{P \times N_y}$. For the Hisense-AC dataset, the prediction head outputs the future indoor temperature trajectory, whereas for the public benchmark datasets it jointly predicts all target variables.

Compared with complex decoder structures, the linear prediction head keeps the model architecture simple while effectively utilizing the multi-scale dynamics, long-range dependencies, and contribution-recalibrated features extracted by the preceding modules. This design improves forecasting efficiency and reduces the risk of overfitting caused by additional parameters.

In addition, variables in HVAC systems usually have different physical units and numerical ranges. Directly feeding raw sequences into the model causes unstable training and weakens the comparability among variables. Therefore, normalization is applied

to the historical multivariate sequence before model training to reduce the influence of scale differences and distribution fluctuations. After prediction, inverse normalization is performed to restore the predicted values to their original physical units. This process not only improves training stability but also ensures the practical interpretability of the forecasting results.

4. Experimental Design

4.1. Experimental Environment

All experiments are conducted under the same hardware and software environment. The detailed configuration is shown in Table 1.

Table 1. Experimental environment configuration.

Environment	Configuration
Operating system	Ubuntu 20.04.2 LTS (64-bit)
Processor	Intel(R) Xeon(R) Gold 5118 CPU @ 2.30 GHz
GPU	NVIDIA Tesla P100 PCIe 16 GB
Programming language	Python 3.9
Development tool	Visual Studio Code

4.2. Datasets

4.2.1. Private Dataset: Hisense-AC

The private Hisense-AC dataset is collected from real-world operation logs of Hisense air-conditioning systems in actual user scenarios. It covers operation records from 15 cities over approximately half a year and is constructed as a real-world measured dataset for indoor temperature forecasting in practical HVAC systems. The dataset contains external weather disturbances, user setting differences, and various unit operating conditions, with indoor temperature selected as the forecasting target. In the recorded samples, the target indoor air temperature t_i ranges from 5.00 to 49.00 °C, while the user-defined `set_temperature` ranges from 16.00 to 32.00 °C. Outdoor-side conditions are described by weather-related variables, such as humidity 12 with a range of 5–100%, and the outdoor discomfort index `outDI` with a range of 36.99–78.86. Equipment-side information includes capacity-related variables, efficiency, running-status indicators, and airflow-related features where available, reflecting observable equipment operating conditions in practical HVAC logs. According to the type of operation event, the dataset is divided into four operating categories: `Heating_on`, `Cooling_on`, `Heating_off`, and `Cooling_off`.

Different from public benchmark datasets that are usually regularly sampled and relatively stable in temporal evolution, Hisense-AC is organized by air-conditioning operation events and is jointly affected by control commands, external disturbances, and historical thermal states. Therefore, it exhibits stronger operating-condition heterogeneity and non-stationary characteristics. The power-on stages usually present coupled fast control responses and slow thermal-inertia effects, while the power-off stages mainly reflect slow variations caused by heat storage and release in the building envelope. Moreover, heating and cooling conditions differ significantly in variable contribution intensity, temperature response direction, and dynamic evolution path.

4.2.2. Public Time-Series Datasets

To evaluate the generalization ability of the proposed model on standard long-term time-series forecasting tasks, experiments are conducted on three public benchmark collections comprising six datasets: ETTh1, ETTh2, ETTm1, ETTm2, Exchange, and Weather.

The ETT dataset contains four subsets, covering both minute-level and hourly sampling granularities. The Exchange dataset consists of continuous exchange-rate observations from multiple countries and is characterized by strong non-stationarity and trend fluctuations. The Weather dataset contains long-term observations of multiple meteorological variables, such as temperature, humidity, air pressure, and wind speed. It exhibits evident multivariate correlations and complex external disturbance patterns, which are partially consistent with the environmental factors in HVAC temperature forecasting. Therefore, it is used to further validate the model's ability to capture coupled meteorological dependencies.

4.3. Experimental Settings

For the Hisense-AC dataset, indoor-related variables sampled at 1 min intervals and outdoor variables sampled every 30 min are aligned to a unified 1 min timeline, with the latter treated as piecewise constants within each operation event. Continuous operation logs are segmented by event, and input–output samples are generated using a sliding window with a step size of 4. The multivariate-to-univariate (MS) setting is adopted to predict the future indoor temperature sequence from historical multivariate observations.

For the public datasets, the multivariate-to-multivariate (M) setting is adopted following the standard long-term forecasting protocol. The input sequence length is set to 96, and the prediction horizons are set to 96, 192, 336, and 720. For all datasets, the patch sizes in the multi-scale temporal interaction embedding module are set to 4, 8, 16, with a unified stride of 8.

All models are trained using Adam with an initial learning rate of 1×10^{-4} , a weight decay of 1×10^{-5} , a batch size of 32, and a dropout rate of 0.2. Early stopping is applied with a patience of 10 epochs. For the Mamba-related modules, the local convolution kernel size and expansion factor are set to 4 and 2, respectively. All datasets are chronologically divided into training, validation, and testing sets with a ratio of 70:10:20.

4.4. Evaluation Metrics

To remain consistent with long-term forecasting benchmarks, mean square error (MSE) and mean absolute error (MAE) are adopted as the main evaluation metrics on the public datasets. For the private Hisense-AC dataset, root mean square error (RMSE) is additionally reported to measure the prediction deviation in the original temperature scale. Lower values of these metrics indicate better forecasting performance.

The three metrics are defined as follows:

$$\text{MSE} = \frac{1}{M} \sum_{m=1}^M (y_m - \hat{y}_m)^2, \quad (27)$$

$$\text{MAE} = \frac{1}{M} \sum_{m=1}^M |y_m - \hat{y}_m|, \quad (28)$$

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{m=1}^M (y_m - \hat{y}_m)^2}. \quad (29)$$

where y_m and $\hat{y}_m$ denote the ground-truth value and the predicted value of the m -th prediction point, respectively, and M denotes the total number of prediction points involved in evaluation. Compared with MSE, MAE is less sensitive to large deviations and better reflects the average robust prediction error.

5. Results and Analysis

5.1. Comparative Experiments on the Hisense-AC Dataset

The comparison results on the Hisense-AC dataset under four operating modes are shown in Table 2. MIRA-Mamba achieves the lowest error values across all modes, indicating its stable forecasting performance under different operating conditions. Specifically, in the Heating_on mode, MIRA-Mamba reduces MSE, MAE, and RMSE by 4.0%, 2.1%, and 1.7%, respectively, compared with the best-performing baseline. In the Heating_off mode, the corresponding reductions are 5.4%, 3.4%, and 2.5%. These results suggest that the proposed model can more effectively capture the dynamic response during the heating process and the long-range decay pattern after shutdown.

Table 2. Experimental comparison results on the Hisense-AC dataset.

Mode	Informer [11]			Autoformer [12]			PatchTST [15]			CIBG-Mamba [38]			DLinear [14]			iTransformer [16]			MIRA-Mamba		
	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE	MSE	MAE	RMSE
Heating_on	0.977	0.739	0.988	0.735	0.644	0.857	0.490	0.495	0.700	<u>0.472</u>	<u>0.471</u>	<u>0.687</u>	0.487	0.496	0.698	0.520	0.496	0.723	0.453	0.461	0.675
Heating_off	0.996	0.832	0.998	0.656	0.626	0.810	0.317	0.393	0.564	<u>0.312</u>	<u>0.377</u>	<u>0.559</u>	0.330	0.407	0.574	0.336	0.397	0.580	0.295	0.364	0.545
Cooling_on	0.516	0.509	0.718	0.592	0.565	0.770	<u>0.406</u>	<u>0.401</u>	<u>0.638</u>	0.411	0.402	0.641	0.421	0.436	0.649	0.422	0.421	0.650	0.393	0.390	0.629
Cooling_off	0.179	0.266	0.424	0.207	0.301	0.455	<u>0.162</u>	<u>0.238</u>	<u>0.402</u>	0.165	0.240	0.406	0.181	0.265	0.425	0.175	0.250	0.418	0.159	0.233	0.399

The best results are highlighted in bold, and the second-best results are underlined. Additional paired Wilcoxon signed-rank tests are conducted to assess statistical significance.

For the cooling modes, MIRA-Mamba also shows consistent improvements. Compared with PatchTST [15], it reduces MSE, MAE, and RMSE by 3.2%, 2.7%, and 1.4% in the Cooling_on mode, and by 1.9%, 2.1%, and 0.7% in the Cooling_off mode, respectively. Overall, the relative gains are more pronounced under heating conditions. The explanation is that temperature evolution during heating is more strongly affected by delayed responses and slowly accumulated thermal effects, making multi-scale temporal representation and adaptive variable contribution modeling more beneficial. These findings demonstrate that MIRA-Mamba can improve multi-step temperature forecasting under complex HVAC operating conditions.

To further assess the statistical reliability of the results in Table 2, paired Wilcoxon signed-rank tests were conducted between MIRA-Mamba and the strongest baseline under each operating mode. CIBG-Mamba [38] was selected as the reference model for Heating_on and Heating_off, while PatchTST [15] was selected for Cooling_on and Cooling_off. The tests indicate statistically significant reductions in sample-wise squared errors and absolute errors across all four modes, with all p -values below 0.001.

Figure 3 further presents the average MSE, MAE, and RMSE of different models on the Hisense-AC dataset. MIRA-Mamba achieves the lowest average errors across all three metrics, indicating its overall advantage over the compared baselines. The bootstrap-based 95% confidence intervals provide additional uncertainty information for the average error estimates. In addition, the relatively small performance variation across different operating modes suggests that the proposed model does not rely on a specific operating condition, but maintains robust forecasting performance under diverse real-world HVAC scenarios.

The rolling visualization in Figure 4 is constructed by extracting the first predicted step from consecutive test windows, providing a continuous view of the rolling forecasting behavior. MIRA-Mamba generally follows the ground-truth temperature evolution with competitive stability under different operating modes, offering an intuitive complement to the quantitative results in Table 2.

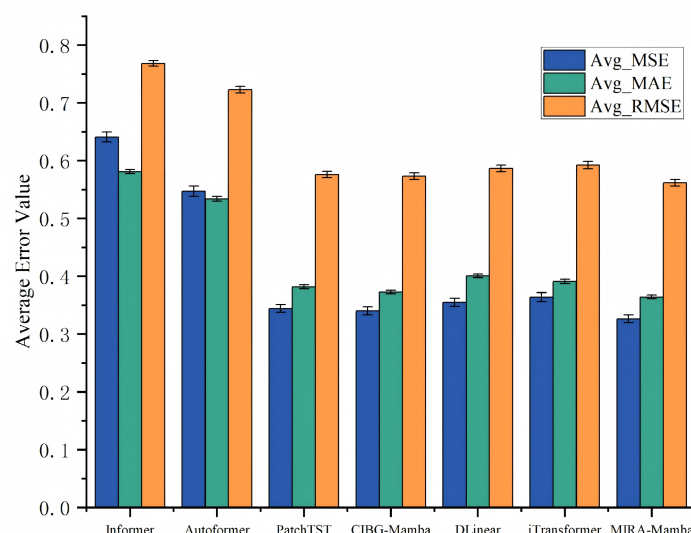


Figure 3. Average error metrics on the Hisense-AC dataset. Error bars denote bootstrap-based 95% confidence intervals estimated over test samples within each operating mode.

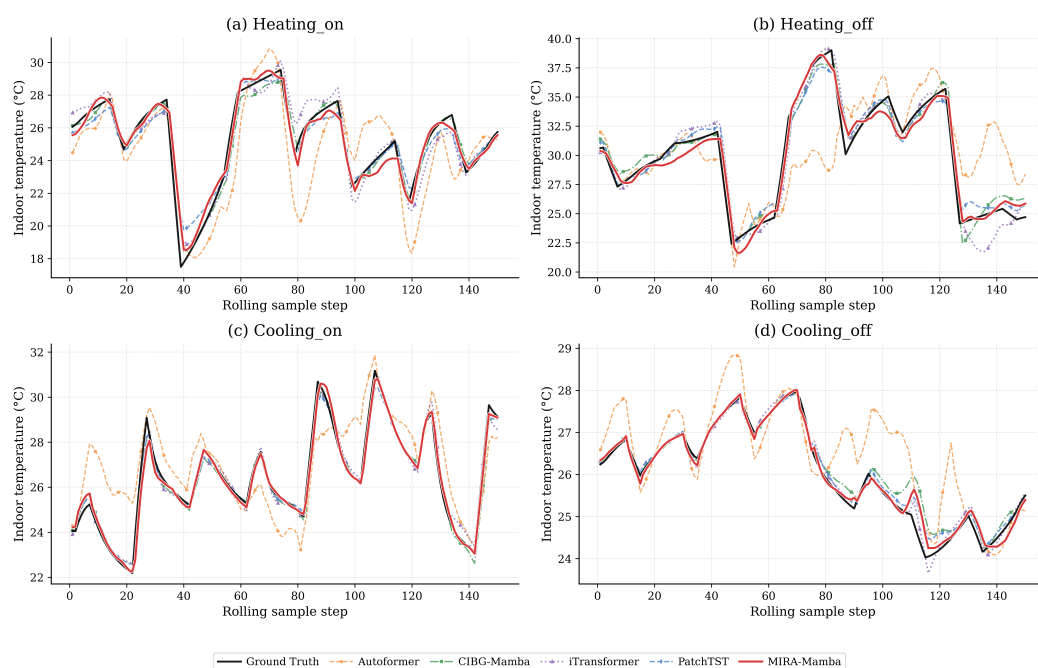


Figure 4. Rolling one-step-ahead forecasting visualizations on the Hisense-AC dataset under four operating modes.

5.2. Comparative Experiments on Public Datasets

To further evaluate the general forecasting ability of MIRA-Mamba beyond the HVAC scenario, we conduct comparative experiments on six public benchmark datasets, including ETTh1, ETTh2, ETTm1, ETTm2, Exchange, and Weather. The results are reported in Table 3. Overall, MIRA-Mamba achieves the best or competitive MSE and MAE under most prediction horizons, indicating that the proposed model is not limited to indoor temperature forecasting but also shows good applicability to general long-term time series forecasting tasks.

Table 3. Experimental comparison results on public datasets.

Model		Informer [11]		Autoformer [12]		TimesNet [26]		SMamba [32]		DLinear [14]		iTransformer [16]		MIRA-Mamba	
Dataset	Horizon	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	96	0.945	0.769	0.446	0.456	<u>0.389</u>	0.412	0.399	0.413	0.396	0.410	0.394	<u>0.409</u>	0.384	0.405
	192	1.018	0.791	0.499	0.483	0.438	0.441	0.455	0.447	0.445	<u>0.440</u>	0.448	<u>0.441</u>	<u>0.441</u>	0.437
	336	1.029	0.780	0.508	0.492	0.493	0.470	0.501	0.472	<u>0.487</u>	<u>0.465</u>	0.491	<u>0.465</u>	0.485	0.460
	720	1.190	0.853	0.568	0.531	0.518	<u>0.495</u>	0.544	0.520	<u>0.513</u>	0.510	0.517	0.501	0.493	0.483
ETTh2	96	2.747	1.314	0.381	0.414	0.337	0.371	<u>0.295</u>	<u>0.347</u>	0.341	0.395	0.301	0.350	0.284	0.340
	192	6.304	2.103	0.452	0.451	0.408	0.416	<u>0.377</u>	<u>0.398</u>	0.482	0.479	0.382	0.400	0.355	0.387
	336	5.343	1.946	0.480	0.487	0.453	0.451	<u>0.420</u>	<u>0.433</u>	0.593	0.542	0.425	<u>0.433</u>	0.407	0.425
	720	4.221	1.713	0.500	0.501	0.439	0.451	<u>0.428</u>	<u>0.447</u>	0.840	0.661	0.427	0.445	0.431	<u>0.447</u>
ETTm1	96	0.624	0.559	0.543	0.495	<u>0.334</u>	0.375	0.338	<u>0.372</u>	0.346	<u>0.372</u>	0.341	0.376	0.321	0.360
	192	0.716	0.608	0.543	0.498	0.408	0.414	<u>0.379</u>	<u>0.393</u>	0.382	0.361	0.253	0.312	0.395	0.396
	336	1.261	0.897	0.621	0.532	0.414	0.422	<u>0.410</u>	<u>0.414</u>	0.415	0.415	0.419	0.419	0.394	0.406
	720	0.973	0.745	0.625	0.540	0.496	0.465	0.476	<u>0.452</u>	<u>0.473</u>	0.451	0.487	0.456	0.464	<u>0.452</u>
ETTm2	96	0.369	0.455	0.260	0.328	0.188	0.268	0.184	0.269	0.193	0.293	<u>0.183</u>	<u>0.266</u>	0.177	0.261
	192	0.816	0.708	0.283	0.342	0.252	0.307	<u>0.249</u>	0.309	0.285	0.361	0.253	0.312	0.246	0.308
	336	1.394	0.906	0.332	0.367	0.321	0.349	<u>0.311</u>	<u>0.348</u>	0.385	0.430	0.315	0.351	0.306	0.346
	720	4.222	1.532	0.423	0.419	0.419	<u>0.406</u>	<u>0.413</u>	0.405	0.556	0.524	0.412	<u>0.406</u>	<u>0.413</u>	0.409
Exchange	96	0.870	0.766	0.145	0.275	0.103	0.232	0.095	<u>0.218</u>	<u>0.094</u>	0.227	0.095	<u>0.218</u>	0.093	0.215
	192	1.180	0.874	0.281	0.386	0.225	0.345	0.189	<u>0.311</u>	<u>0.186</u>	0.325	<u>0.186</u>	<u>0.311</u>	0.182	0.307
	336	1.659	1.040	0.669	0.620	0.364	0.439	0.337	0.422	0.327	0.435	0.334	0.420	<u>0.331</u>	<u>0.421</u>
	720	2.830	1.391	1.114	0.819	1.010	0.765	<u>0.882</u>	<u>0.712</u>	0.749	0.664	0.922	0.731	0.912	0.721
Weather	96	0.442	0.466	0.313	0.368	0.169	0.220	0.180	0.223	0.196	0.255	0.178	<u>0.218</u>	<u>0.170</u>	0.213
	192	0.474	0.472	0.331	0.373	0.225	0.267	0.228	0.262	0.240	0.299	0.224	0.257	0.221	<u>0.260</u>
	336	0.461	0.459	0.370	0.395	0.282	0.305	<u>0.281</u>	<u>0.301</u>	<u>0.281</u>	0.331	0.282	0.299	0.279	<u>0.301</u>
	720	1.056	0.747	0.590	0.548	0.356	0.354	0.357	<u>0.352</u>	0.345	0.382	0.360	0.353	<u>0.352</u>	0.350

The best results are highlighted in bold, and the second-best results are underlined.

Specifically, MIRA-Mamba achieves the lowest MSE in 17 of 24 forecasting settings, with an average MSE of 0.355, which is 2.39% lower than SMamba [32]. For MAE, it achieves the lowest value in 15 settings, with an average MAE of 0.380, 1.32% lower than iTransformer. The improvement is relatively clear on ETTh2, where the series shows stronger fluctuations than ETTh1. At the 192-step horizon, MIRA-Mamba obtains an MSE of 0.355, compared with 0.377 for SMamba, corresponding to a 5.84% reduction. This suggests that the model can better handle fluctuating patterns in long-horizon forecasting. For the higher-frequency ETTm datasets, local variations are more fine-grained. On ETTm1, MIRA-Mamba reduces the MSE by 3.90% compared with SMamba at the 336-step horizon, indicating its effectiveness for minute-level temporal changes. On non-ETT datasets, Exchange contains stronger non-stationary trends, while Weather involves more complex multivariate dependencies. The results on short- and medium-term Exchange forecasting and most Weather settings suggest that MIRA-Mamba also remains competitive beyond the ETT datasets.

Figure 5 further presents the normalized performance comparison based on MSE and MAE. The radar charts in panels (a) and (b) average the relative scores over four prediction horizons for each dataset, providing a view of the overall performance across datasets. Some baselines show noticeable inward shifts on specific dataset axes, indicating that their performance is more affected by dataset types. In contrast, MIRA-Mamba shows radar profiles generally closer to the outer boundary on both metrics, suggesting better average relative performance. Its relatively balanced shape across different axes further indicates more stable forecasting results across datasets. Panels (c) and (d) further show the horizon-level 95% confidence intervals for the normalized MSE-based and MAE-based scores, respectively. The results indicate that MIRA-Mamba generally maintains high normalized scores with relatively stable performance across different forecasting horizons, further demonstrating its competitive forecasting ability on public benchmarks.

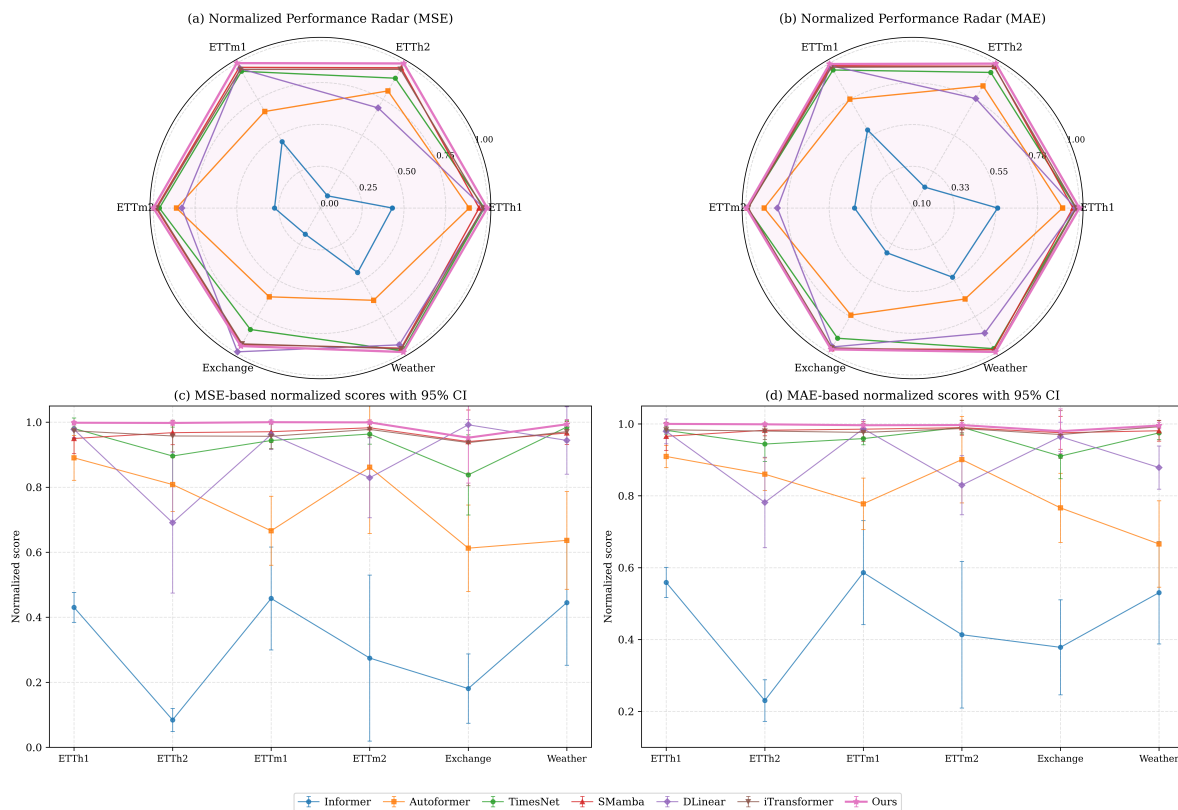


Figure 5. Normalized performance comparison on public benchmarks. (a,b) MSE-based and MAE-based normalized radar charts, respectively. (c,d) Corresponding horizon-level 95% confidence intervals of the normalized scores across the four forecasting horizons.

5.3. Ablation Study

To evaluate the contribution of the main designs in MIRA-Mamba, we organize the ablation study in three steps. First, module-level ablations are conducted by separately removing the Multi-scale Temporal Interaction Embedding (MTE) module and the Feature Contribution Recalibration (FCR) module on the four operating modes of Hisense-AC, as well as the Exchange and Weather datasets. Second, to further analyze the internal design of FCR, we remove the variable-importance gate and the channel affinity modulation module under the Heating_on and Heating_off modes. Third, to further assess the design of the embedding module, we remove its positional encoding and grouped channel attention on the Heating_on mode of Hisense-AC.

5.3.1. Ablation of Key Components

As shown in Table 4, both FCR-only and MTE-only consistently reduce prediction errors across all datasets and operating modes, while the full MIRA-Mamba achieves the lowest errors in all settings. The two single-module variants exhibit different strengths. On Weather, MTE-only achieves an MSE of 0.172 and an MAE of 0.215, compared with 0.180 and 0.219 for FCR-only, suggesting the benefit of multi-scale temporal modeling for complex temporal variations. In contrast, FCR-only achieves an MSE of 0.102 and an MAE of 0.234 on Exchange, compared with 0.106 and 0.241 for MTE-only, indicating the effectiveness of adaptive variable contribution recalibration when variable relationships are more unstable.

With the joint use of MTE and FCR, the full model further improves upon the single-module variants across all operating modes. On Heating_off, the MSE decreases from 0.335 with MTE-only to 0.295 after incorporating FCR, indicating that contribution recalibration provides additional benefits on top of multi-scale temporal modeling. For the remaining

HVAC modes, MIRA-Mamba also maintains consistent advantages over both singlemodule variants, achieving the lowest MSE and MAE on Heating_on, Cooling_on, and Cooling_off. On Exchange, the MAE further decreases from 0.234 with FCR-only to 0.215 in the full model. These results indicate that MTE and FCR contribute complementary effects rather than redundant improvements.

Table 4. Ablation study on key components. The symbols ✓ and × indicate whether a component is used or not, respectively; ↓ indicates that lower values are better; bold values denote the best results.

Variant	MTE	FCR	Heating_on		Heating_off		Cooling_on		Cooling_off		Exchange		Weather	
			MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓	MSE ↓	MAE ↓
Baseline	×	×	0.470	0.470	0.347	0.380	0.411	0.402	0.165	0.240	0.110	0.265	0.186	0.223
+FCR only	×	✓	0.461	0.468	0.346	0.377	0.403	0.400	0.164	0.237	0.102	0.234	0.180	0.219
+MTE only	✓	×	0.464	0.468	0.335	0.370	0.399	0.394	0.162	0.235	0.106	0.241	0.172	0.215
Ours	✓	✓	0.453	0.461	0.295	0.364	0.393	0.390	0.159	0.233	0.093	0.215	0.170	0.213

5.3.2. Internal Ablation of FCR

To analyze the individual roles of the two sequential adaptation mechanisms in the Feature Contribution Recalibration (FCR) module, we conduct an internal ablation study under the Heating_on and Heating_off modes. As shown in Table 5, removing either channel affinity modulation or the variable-importance gate increases the forecasting errors, indicating that both components contribute to feature recalibration. Under Heating_on, the MSE increases from 0.453 to 0.461 without affinity modulation and to 0.458 without variable gating. Under Heating_off, removing the variable-importance gate leads to a larger degradation than removing affinity modulation, suggesting that adaptive variable weighting is particularly important for the shutdown process, where the temperature trajectory is mainly driven by residual thermal effects and the relative influence of input variables changes over time. Overall, combining variable-level recalibration with channel-group interaction modeling provides a more effective encoded representation.

Table 5. Ablation results of internal components in the FCR module. The symbol ↓ indicates that lower values are better; bold values denote the best results.

Variant	Heating_on		Heating_off	
	MSE ↓	MAE ↓	MSE ↓	MAE ↓
Ours (Full)	0.453	0.461	0.295	0.364
w/o aff	0.461	0.465	0.310	0.372
w/o gate	0.458	0.463	0.335	0.368

5.3.3. Internal Ablation of MTE

To examine the effect of the internal components in MTE, positional encoding and grouped channel attention are removed on the Heating_on mode of the Hisense-AC dataset, as shown in Table 6. The complete model achieves an MSE of 0.453 and an MAE of 0.461. After removing positional encoding, the MSE and MAE increase to 0.464 and 0.468, corresponding to relative increases of 2.43% and 1.52%, respectively. This indicates that explicit positional information helps preserve temporal order in multi-scale patch representations. After removing grouped channel attention, the MSE and MAE increase to 0.457 and 0.463, corresponding to relative increases of 0.88% and 0.43%. This suggests that channel interaction further enhances the modeling of coupling relationships among multivariate inputs. Overall, both positional encoding and grouped channel attention contribute to the effectiveness of MTE, with positional encoding showing a more pronounced impact.

Table 6. Ablation results of internal components in the embedding layer. The symbol ↓ indicates that lower values are better; bold values denote the best results.

Variant	MSE ↓	MAE ↓
Ours (Full)	0.453	0.461
w/o positional encoding	0.464	0.468
w/o grouped channel attention	0.457	0.463

6. Conclusions

This study investigates multi-step indoor temperature forecasting for HVAC systems, where accurate prediction of future temperature trajectories is essential for adapting to varying operating conditions. To address this task, we propose the MIRA-Mamba forecasting framework. The proposed method captures temperature dynamics across multiple temporal scales, preserves informative historical dependencies, and adaptively adjusts the influence of input variables under different operating conditions, thereby improving the representation of complex temperature evolution patterns. Experimental results on the Hisense-AC dataset demonstrate that MIRA-Mamba achieves stable forecasting performance across heating and cooling modes. Its competitive results on the ETT, Exchange, and Weather datasets further indicate that the proposed framework is not restricted to HVAC temperature forecasting and possesses good applicability to general long-term time series forecasting tasks. Quantitatively, compared with the strongest competing baseline under each operating mode, MIRA-Mamba achieves average relative reductions of approximately 3.63%, 2.60%, and 1.60% in MSE, MAE, and RMSE across the four Hisense-AC modes, respectively; on the six public benchmarks, it achieves the best MSE in 17 out of 24 settings and the best MAE in 15 out of 24 settings. The ablation study also confirms the effectiveness of the principal design components. Future work will extend the evaluation to longer observation periods, a wider range of building types, and more diverse climate conditions, while further investigating model robustness under missing sensor data, measurement noise, and unreliable observations. In addition, the integration of the proposed forecasting model with model predictive control (MPC)-based HVAC control frameworks will be explored in future work to evaluate its practical effects on energy consumption, thermal comfort, and closed-loop control performance.

Author Contributions: Conceptualization, M.J., C.W. and L.W.; methodology, M.J.; software, M.J.; validation, M.J., D.R. and T.S.; formal analysis, M.J.; investigation, M.J., D.R. and T.S.; resources, D.R., T.S., C.W. and L.W.; data curation, M.J., D.R. and T.S.; writing—original draft preparation, M.J.; writing—review and editing, M.J., C.W. and L.W.; visualization, M.J.; supervision, C.W. and L.W.; project administration, C.W. and L.W.; funding acquisition, C.W. and L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 62072469; the Natural Science Foundation of Shandong Province, grant number ZR2024MF145; the Qingdao West Coast New Area “Unveiling the List” Science and Technology Breakthrough Project, grant number 2024-1; and the Key Technology Research and Development and Industrialization Demonstration Project of Qingdao City, grant number 24-1-2-gljh-8-gx.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: Authors Daiwei Ruan and Tiejun Sun were employed by the company Qingdao Hisense Hitachi Air-conditioning Systems Co., Ltd., R&D Center, Qingdao, China. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations and Notation

The following abbreviations and notation are used in this manuscript:

HVAC	Heating, ventilation, and air conditioning
MPC	Model Predictive Control
MIRA-Mamba	Multi-scale Interactive Recalibrated Mamba
MTE	Multi-scale Temporal Interaction Embedding
BSRE	Bidirectional State-space Recalibration Encoder
FCR	Feature Contribution Recalibration
GCA	Grouped Channel Attention
GSA	Global Self-Attention
GRU	Gated Recurrent Unit
MSE	Mean Squared Error
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
B	Batch size
L	Historical input window length
P	Forecasting horizon
N	Number of input variables
d	Hidden feature dimension
K	Number of temporal scales
p_k	Patch length at the k -th temporal scale
s	Sliding-window stride
$\mathbf{X}$	Historical input sequence
$\hat{\mathbf{Y}}$	Predicted future sequence
$\odot$	Element-wise multiplication

References

1. International Energy Agency. Buildings. In *Energy Efficiency 2025*; International Energy Agency: Paris, France, 2025.
2. Zhuang, D.; Gan, V.J.; Tekler, Z.D.; Chong, A.; Tian, S.; Shi, X. Data-driven predictive control for smart HVAC system in IoT-integrated buildings with time-series forecasting and reinforcement learning. *Appl. Energy* **2023**, *338*, 120936. [\[CrossRef\]](#)
3. Banjac, A.; Novak, H.; Vašak, M. Implementation of model predictive indoor climate control for hierarchical building energy management. *Control Eng. Pract.* **2023**, *136*, 105536. [\[CrossRef\]](#)
4. Sun, Y.; Haghighat, F.; Fung, B.C. A review of the-state-of-the-art in data-driven approaches for building energy prediction. *Energy Build.* **2020**, *221*, 110022. [\[CrossRef\]](#)
5. Liu, H.; Liang, J.; Liu, Y.; Wu, H. A review of data-driven building energy prediction. *Buildings* **2023**, *13*, 532. [\[CrossRef\]](#)
6. Xu, C.; Chen, H.; Wang, J.; Guo, Y.; Yuan, Y. Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method. *Build. Environ.* **2019**, *148*, 128–135. [\[CrossRef\]](#)
7. Mtibaa, F.; Nguyen, K.K.; Azam, M.; Papachristou, A.; Venne, J.S.; Cheriet, M. LSTM-based indoor air temperature prediction framework for HVAC systems in smart buildings. *Neural Comput. Appl.* **2020**, *32*, 17569–17585. [\[CrossRef\]](#)
8. Fang, Z.; Crimier, N.; Scanu, L.; Midelet, A.; Alyafi, A.; Delinchant, B. Multi-zone indoor temperature prediction with LSTM-based sequence to sequence model. *Energy Build.* **2021**, *245*, 111053. [\[CrossRef\]](#)
9. Elmaz, F.; Eyckerman, R.; Casteels, W.; Latré, S.; Hellinckx, P. CNN-LSTM architecture for predictive indoor temperature modeling. *Build. Environ.* **2021**, *206*, 108327. [\[CrossRef\]](#)
10. Zheng, G.; Jia, R.; Yi, W.; Yue, X. Multi-step fusion model for predicting indoor temperature in residential buildings based on attention mechanism and neural network. *J. Build. Eng.* **2025**, *102*, 112057. [\[CrossRef\]](#)
11. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Palo Alto, CA, USA, 2021; Volume 35, pp. 11106–11115. [\[CrossRef\]](#)
12. Wu, H.; Xu, J.; Wang, J.; Long, M. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 22419–22430. [\[CrossRef\]](#)

13. Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; Jin, R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the International Conference on Machine Learning*; PMLR: Cambridge, MA, USA, 2022; pp. 27268–27286. [\[CrossRef\]](#)
14. Zeng, A.; Chen, M.; Zhang, L.; Xu, Q. Are transformers effective for time series forecasting? In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2023; Volume 37, pp. 11121–11128. [\[CrossRef\]](#)
15. Nie, Y.; Nguyen, N.H.; Sinthong, P.; Kalagnanam, J. A time series is worth 64 words: Long-term forecasting with transformers. *arXiv* **2022**, arXiv:2211.14730. [\[CrossRef\]](#)
16. Liu, Y.; Hu, T.; Zhang, H.; Wu, H.; Wang, S.; Ma, L.; Long, M. itransformer: Inverted transformers are effective for time series forecasting. In *Proceedings of the International Conference on Learning Representations, ICLR*; OpenReview.net: Vienna, Austria, 2024; Volume 2024, pp. 11116–11140. [\[CrossRef\]](#)
17. Wang, S.; Wu, H.; Shi, X.; Hu, T.; Luo, H.; Ma, L.; Zhang, J.; Zhou, J. Timemixer: Decomposable multiscale mixing for time series forecasting. In *Proceedings of the International Conference on Learning Representations, ICLR*; OpenReview.net: Vienna, Austria, 2024; Volume 2024, pp. 38626–38652. [\[CrossRef\]](#)
18. Gu, A.; Dao, T. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv* **2023**, arXiv:2312.00752. [\[CrossRef\]](#)
19. Loffa, M.A.; Macii, E.; Patti, E.; Bottaccioli, L. Physics-Informed vs. Deep Learning: Indoor Temperature Prediction with Different Data Availability. In *Proceedings of the 16th ACM International Conference on Future and Sustainable Energy Systems*, Rotterdam, The Netherlands, 17–20 June 2025; pp. 742–750. [\[CrossRef\]](#)
20. Hannula, E.; Häkkinen, A.; Solonen, A.; Uribe, F.; de Wiljes, J.; Roininen, L. *Bayesian LSTM for Indoor Temperature Modeling*; SSRN: Rochester, NY, USA, 2025. [\[CrossRef\]](#)
21. Machacuay, J.; Manrique-Silupu, J.; Ipanaque, W. LSTM with shallow NNs for indoor temperature long-term predictions in refrigeration systems. *Neural Comput. Appl.* **2025**, *37*, 1067–1088. [\[CrossRef\]](#)
22. Drgoňa, J.; Tuor, A.R.; Chandan, V.; Vrabie, D.L. Physics-constrained deep learning of multi-zone building thermal dynamics. *Energy Build.* **2021**, *243*, 110992. [\[CrossRef\]](#)
23. Bellagarda, A.; Cesari, S.; Aliberti, A.; Ugliotti, F.; Bottaccioli, L.; Macii, E.; Patti, E. Effectiveness of neural networks and transfer learning for indoor air-temperature forecasting. *Autom. Constr.* **2022**, *140*, 104314. [\[CrossRef\]](#)
24. Li, H.; Pinto, G.; Piscitelli, M.S.; Capozzoli, A.; Hong, T. Building thermal dynamics modeling with deep transfer learning using a large residential smart thermostat dataset. *Eng. Appl. Artif. Intell.* **2024**, *130*, 107701. [\[CrossRef\]](#)
25. Wang, X.; Wang, X.; Yin, X.; Li, K.; Wang, L.; Wang, R.; Song, R. Distributed LSTM-GCN-based spatial-temporal indoor temperature prediction in multizone buildings. *IEEE Trans. Ind. Inform.* **2023**, *20*, 482–491. [\[CrossRef\]](#)
26. Wu, H.; Hu, T.; Liu, Y.; Zhou, H.; Wang, J.; Long, M. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv* **2022**, arXiv:2210.02186. [\[CrossRef\]](#)
27. Zhang, Y.; Yan, J. Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting. In *Proceedings of the International Conference on Learning Representations*, Kigali, Rwanda, 1–5 May 2023.
28. Wang, H.; Peng, J.; Huang, F.; Wang, J.; Chen, J.; Xiao, Y. MICN: Multi-Scale Local and Global Context Modeling for Long-Term Series Forecasting. In *Proceedings of the International Conference on Learning Representations*, Kigali, Rwanda, 1–5 May 2023.
29. Liu, M.; Zeng, A.; Chen, M.; Xu, Z.; Lai, Q.; Ma, L.; Xu, Q. Scinet: Time series modeling and forecasting with sample convolution and interaction. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 5816–5828. [\[CrossRef\]](#)
30. Gu, A.; Goel, K.; Ré, C. Efficiently modeling long sequences with structured state spaces. *arXiv* **2021**, arXiv:2111.00396. [\[CrossRef\]](#)
31. Smith, J.T.; Warrington, A.; Linderman, S.W. Simplified state space layers for sequence modeling. *arXiv* **2022**, arXiv:2208.04933. [\[CrossRef\]](#)
32. Wang, Z.; Kong, F.; Feng, S.; Wang, M.; Yang, X.; Zhao, H.; Wang, D.; Zhang, Y. Is mamba effective for time series forecasting? *Neurocomputing* **2025**, *619*, 129178. [\[CrossRef\]](#)
33. Cai, X.; Zhu, Y.; Wang, X.; Yao, Y. MambaTS: Improved selective state space models for long-term time series forecasting. *arXiv* **2024**, arXiv:2405.16440. [\[CrossRef\]](#)
34. Liang, A.; Jiang, X.; Sun, Y.; Lu, C. Bi-mamba4ts: Bidirectional Mamba for time series forecasting. *arXiv* **2024**, arXiv:2404.15772. [\[CrossRef\]](#)
35. Hu, J.; Lan, D.; Zhou, Z.; Luo, G.; Wen, Q.; Liang, Y. How to Train Your Mamba for Time Series Forecasting. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 1*, Jeju, Republic of Korea, 9–13 August 2026; pp. 476–485. [\[CrossRef\]](#)
36. Karadag, Y.M.; Talaz, I.; Dino, I.G.; Kalkan, S. ms-mamba: Multi-scale mamba for time-series forecasting. *Neurocomputing* **2026**, *680*, 133226. [\[CrossRef\]](#)

37. Cho, K.; Van Merriënboer, B.; Gulçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [[CrossRef](#)]
38. Li, P.; Zheng, X.; Xiang, S.; Hou, J.; Qin, Y.; Kurboniyon, M.S.; Ren, W. Channel Independence Bidirectional Gated Mamba With Interactive Recurrent Mechanism for Time Series Forecasting. *IEEE Trans. Ind. Electron.* **2025**, *72*, 14593–14602. [[CrossRef](#)]

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