

Article

SIMBA: A Bidirectional Retrieval–Forward Simulation Framework for Modeling FY-4A/GIIRS Hyperspectral Infrared Radiances Toward NWP Applications

Jingdong Shen ^{1,†}, Fu Wang ^{2,†} , Qifeng Lu ², Hao Huang ¹, Chunqiang Wu ² , Chi Yang ¹ and Xiaofang Liu ^{1,*}

¹ School of Computer Science and Engineering, Sichuan University of Science & Engineering, Yibin 644000, China; 324085406126@stu.suse.edu.cn (J.S.); 324085406108@stu.suse.edu.cn (H.H.); 323085406228@stu.suse.edu.cn (C.Y.)

² State Key Laboratory of Severe Weather Meteorological Science and Technology, CMA Earth System Modeling and Prediction Centre and Key Laboratory of Earth System Modeling and Prediction, China Meteorological Administration, Beijing 100081, China; wangfu@cma.cn (F.W.); luqf@cma.gov.cn (Q.L.); wucq@cma.gov.cn (C.W.)

* Correspondence: lxf1969@suse.edu.cn

† These authors contributed equally to this work.

Highlights

What are the main findings?

- SIMBA, a cycle-consistent bidirectional framework inspired by the data assimilation of the numerical weather prediction (NWP) system, is developed to jointly perform atmospheric profile retrieval and forward brightness temperature (BT) simulation for hyperspectral infrared observations.
- SIMBA achieves the best overall performance among the compared baseline models in temperature and specific humidity retrieval, and BT simulation, while improving vertical dependency modeling, observation-space consistency, and channel-wise stability across pressure levels and spectral regions.

What are the implications of the main findings?

- The results indicate that joint retrieval and forward simulation is an effective way to build a more physically consistent relationship between hyperspectral observations and atmospheric states, making the framework better aligned with the needs of data assimilation of NWP.
- SIMBA includes a differentiable profile-to-radiance reconstruction branch that enables observation-space consistency assessment of retrieved atmospheric profiles. This validate-by-reconstruction strategy may provide a methodological reference for other retrieval problems in which estimated geophysical states need to remain consistent with their corresponding observations.



Academic Editor: Jiaojiao Li

Received: 2 April 2026

Revised: 3 June 2026

Accepted: 16 June 2026

Published: 1 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 \(CC BY\) license](#).

Abstract

Hyperspectral infrared observations are an important data source for numerical weather prediction (NWP) because they provide rich information on the vertical structure of atmospheric temperature and humidity. However, most existing deep learning methods mainly focus on one-way retrieval from radiances to atmospheric profiles, while the reverse radiance simulation process and the consistency between atmospheric state space and radiance observation space are insufficiently considered. In this study, we propose SIMBA, a unified bidirectional retrieval–forward simulation framework for FY-4A/GIIRS hyperspectral infrared radiance modeling toward NWP applications. The framework

jointly performs atmospheric profile retrieval and radiance reconstruction, introduces a cycle-consistency constraint to strengthen the coupling between the two processes, and employs a bidirectional Mamba state-space module to capture long-range dependencies along pressure levels. Using collocated FY-4A/GIIRS observations and ERA5 reanalysis data, the proposed method is evaluated for temperature retrieval, specific humidity retrieval, long-wave radiance reconstruction, and medium wave radiance reconstruction. Experimental results show that SIMBA outperforms several representative deep learning baselines across both retrieval and reconstruction tasks, while ablation experiments confirm the contribution of the bidirectional design and cycle-consistency mechanism. These results demonstrate that the proposed framework is effective for joint atmospheric profile retrieval and hyperspectral infrared radiance modeling, and suggests potential for future Jacobian-related analysis and NWP-oriented extensions.

Keywords: bidirectional Mamba; atmospheric profiles; FY-4A/GIIRS; hyperspectral infrared observation

1. Introduction

The quality of numerical weather prediction (NWP) [1] strongly depends on the accuracy of the initial atmospheric state, making the effective use of satellite observations [2] essential for improving forecast skill. Among available observation types, hyperspectral infrared measurements [3] provide abundant information on the vertical structure of atmospheric temperature and water vapor. Compared with conventional radiosonde and surface observations, they offer broader spatial coverage [4], higher temporal continuity, and richer vertically resolved thermodynamic information, and are therefore particularly valuable [5] over regions with sparse conventional observations.

The Geostationary Interferometric Infrared Sounder (GIIRS) onboard FY-4A [6,7] is the world's first geostationary hyperspectral infrared sounder. Unlike polar-orbiting hyperspectral infrared instruments [8], GIIRS provides high-frequency regional observations over fixed domains [9], creating new opportunities for continuous monitoring of atmospheric thermodynamic conditions. However, GIIRS does not directly observe atmospheric temperature and humidity profiles, but rather top-of-atmosphere infrared radiances. Therefore, a key task is to establish an effective and stable mapping between atmospheric state variables and observed radiances [10].

In conventional practice, this mapping is established through radiative transfer models (RTMs) [11,12], which connect atmospheric state space and observation space. RTMs play a central role in radiance simulation, profile retrieval, and hyperspectral infrared applications in NWP [13]. In particular, variational data assimilation [14] requires accurate forward operators and reliable sensitivity information to describe the relationship between atmospheric states and radiance observations [15]. However, under high-dimensional channel settings and large-scale data conditions, conventional RTMs can be computationally expensive, which motivates the development of more efficient surrogate modeling strategies [16,17].

Recent advances in deep learning have provided a new pathway for learning the nonlinear relationship between hyperspectral infrared radiances and atmospheric profiles [18–20]. Existing studies have employed convolutional neural networks (CNNs) [21], long short-term memory networks (LSTMs) [22], Transformers [23], and multilayer perceptrons (MLPs) [24] to estimate temperature or humidity profiles directly from radiance observations. In the broader hyperspectral remote-sensing literature, generative adversarial networks (GANs) have also been used for spectral–spatial representation learning, data augmentation, and hyperspectral image classification [25]. However, most of these studies

mainly focus on one-way retrieval [26], namely, learning the mapping from radiances to atmospheric profiles, while the forward simulation process from atmospheric states to radiances is rarely modeled explicitly within the same framework [27]. As a result, model optimization is dominated by state-space errors, whereas consistency in observation space [28] is insufficiently constrained. This limitation weakens the physical coherence of the learned mapping and reduces its ability to maintain consistency between atmospheric state space and radiance observation space [29,30].

Another challenge arises from the intrinsic vertical structure of atmospheric thermodynamic profiles [31]. Temperature and humidity profiles are organized along pressure levels and exhibit pronounced cross-level coupling and long-range dependencies [32]. Thus, hyperspectral infrared radiance–profile modeling is not only a nonlinear regression problem [33], but also a vertical sequence modeling problem [34]. Traditional recurrent neural networks may suffer from gradient attenuation when handling long sequences, whereas Transformer-based models often incur high computational cost due to global attention operations. Recently developed state-space models, especially the Mamba architecture, provide strong long-sequence modeling capability with linear computational complexity [35], making them well suited for representing multilayer vertical dependencies [36].

To address these issues, we propose SIMBA, a bidirectional retrieval–forward simulation framework for FY-4A/GIIRS hyperspectral infrared radiance modeling toward NWP applications [37]. Instead of focusing only on atmospheric profile retrieval, we jointly model the radiance-to-profile retrieval process and the profile-to-radiance forward simulation process within a unified architecture [38]. Specifically, we design a retrieval branch to estimate atmospheric thermodynamic profiles from hyperspectral infrared radiances, and a forward branch [39] to reconstruct radiances from retrieved or reference profiles. We further couple the two branches through a cycle-consistency constraint [40], so that the framework can be optimized in both state space and observation space simultaneously [41]. In this way, we strengthen the linkage between atmospheric states and radiative observations and construct a data-driven differentiable forward-modeling framework that improves radiance–profile consistency under the evaluated conditions [42,43]. For this purpose, a bidirectional Mamba state-space module is incorporated into the framework to effectively capture the long-range vertical dependencies inherent in atmospheric thermodynamic profiles, enabling more accurate modeling of the structured vertical coupling in temperature and humidity.

The main contributions of this study are as follows: (1) We propose a unified bidirectional retrieval–forward simulation framework for FY-4A/GIIRS hyperspectral infrared radiance modeling, which jointly links atmospheric state space and radiance observation space. (2) We introduce a cycle-consistency constraint to couple the retrieval and forward branches, thereby improving observation-space consistency and the stability of bidirectional radiance–profile mapping. (3) We incorporate a bidirectional Mamba state-space module to model long-range vertical dependencies of atmospheric thermodynamic profiles, and demonstrate its effectiveness in both profile retrieval and radiance reconstruction tasks. (4) We provide a data-driven differentiable forward-modeling basis for future extensions toward sensitivity analysis and assimilation-oriented applications, while recognizing that rigorous RTM-based validation remains necessary.

The remainder of this paper is organized as follows. Section 2 describes the data, sample construction strategy, and methodology. Section 3 presents the experimental results and discussion. Section 4 concludes the paper and outlines future work.

2. Data and Methodology

2.1. Satellite and Reanalysis Data

This study uses observations from the Geostationary Interferometric Infrared Sounder (GIIRS) and AGRI onboard China's FY-4A satellite [7]. GIIRS measures infrared radiances in two spectral bands: a long-wave (LW) band ($700\text{--}1130\text{ cm}^{-1}$, 689 channels) and a medium-wave (MW) band ($1650\text{--}2250\text{ cm}^{-1}$, 961 channels), both with a spectral resolution of 0.625 cm^{-1} . The instrument operates with a 32×4 field-of-view array scanning mode, providing a spatial resolution of approximately 16 km at the sub-satellite point and a temporal resolution of 67 min for the China area and 35 min for the mesoscale area. The main technical specifications of GIIRS are listed in Table 1. In this study, Level 1 radiance products from the LW and MW bands are used together with AGRI Level 2 cloud-mask products and auxiliary information, including latitude, longitude, and observation time. All GIIRS and AGRI data are obtained from the National Satellite Meteorological Centre (NSMC) of China.

Table 1. FY-4A GIIRS specifications.

Parameter	Performance
Spectral bandwidth	Long-wave: $700\text{--}1130\text{ cm}^{-1}$ Medium-wave: $1650\text{--}2250\text{ cm}^{-1}$
Spectral channels	Long-wave: 689 Medium-wave: 961
Spectral resolution	Long-wave: 0.625 cm^{-1} Medium-wave: 0.625 cm^{-1}
Sensitivity	Long-wave: $0.5\text{--}1.1\text{ mW m}^{-2}\text{ sr}^{-1}\text{ cm}^{-2}$ Medium-wave: $0.1\text{--}0.14\text{ mW m}^{-2}\text{ sr}^{-1}\text{ cm}^{-2}$
Operational model	China area: $5000 \times 5000\text{ km}^2$ Mesoscale area: $2000 \times 2000\text{ km}^2$
Spatial resolution	16 km
Temporal resolution	China area: 67 min Mesoscale area: 35 min
Calibration accuracy	1.5 K (3σ) radiation 10 ppm (3σ) spectrum

ERA5 reanalysis data [44] (ECMWF) are used as reference atmospheric states for model training and evaluation. Temperature and specific humidity profiles at 101 vertical levels, from the near surface to roughly 0.005 hPa, are used. ERA5 has a horizontal resolution of roughly 31 km and a temporal resolution of 60 min, enabling precise spatiotemporal collocation with GIIRS observations. ERA5 is used as a supervisory reference rather than as in situ ground truth, because GIIRS provides footprint-scale radiances while ERA5 represents grid-scale fields. This scale mismatch may introduce representativeness errors into the training labels and force the model to map detailed observational features to smoothed targets, potentially degrading the representation of fine-scale or extreme localized events. The ERA5-based evaluation should therefore be interpreted as a consistency assessment with reanalysis profiles, not as independent in situ validation.

2.2. Matching Methods and Dataset

GIIRS provides 1650 spectral channels with rich but highly correlated information. Direct use of all channels would introduce substantial redundancy and increase computational cost [45]. Therefore, the input channels used in this study were selected with

reference to the previous GIIRS channel-selection study by Zhang et al. [46], rather than being newly optimized in the present study. Based on Jacobian characteristics and cumulative influence coefficients, channels with high sensitivity to temperature and humidity were retained. In the LW band ($700\text{--}1130\text{ cm}^{-1}$), channels in the CO_2 absorption region ($700\text{--}790\text{ cm}^{-1}$) and H_2O absorption region ($779\text{--}951\text{ cm}^{-1}$) were selected; in the MW band ($1650\text{--}2250\text{ cm}^{-1}$), channels in the H_2O absorption region ($1650\text{--}2000\text{ cm}^{-1}$) were retained, along with several window channels for near-surface thermodynamic information. This yields 106 LW and 99 MW channels, used in all experiments.

Based on the selected channels, a collocated dataset is constructed from FY-4A/GIIRS radiance observations and ERA5 reanalysis profiles, with GIIRS as the reference. ERA5 profiles within ± 30 min are matched to the nearest spatial grid point and interpolated to the 101 pressure levels of the satellite product. Only samples with valid LW and MW radiances and corresponding ERA5 profiles are retained. Cloudy-sky samples are selected because they represent a more challenging radiative regime with stronger nonlinear interactions among temperature, humidity, clouds, and surface effects. After spatiotemporal matching and quality control, 158,649 cloudy-sky samples are obtained. Each sample contains the selected radiance channels, pressure-level coordinates, and the corresponding ERA5 temperature and humidity profiles.

To examine cross-condition generalization, a clear-sky dataset is constructed using the same procedures, yielding 112,151 samples. The cloudy-sky-trained models are evaluated on this dataset without retraining; results are reported in Section 3.1.

2.3. Proposed Network

In this study, we propose SIMBA, a bidirectional retrieval–forward simulation framework for atmospheric profile retrieval and radiance reconstruction from FY-4A/GIIRS hyperspectral infrared observations. As shown in Figure 1, SIMBA consists of a retrieval branch and a forward simulation branch coupled through a cycle-consistency mechanism.

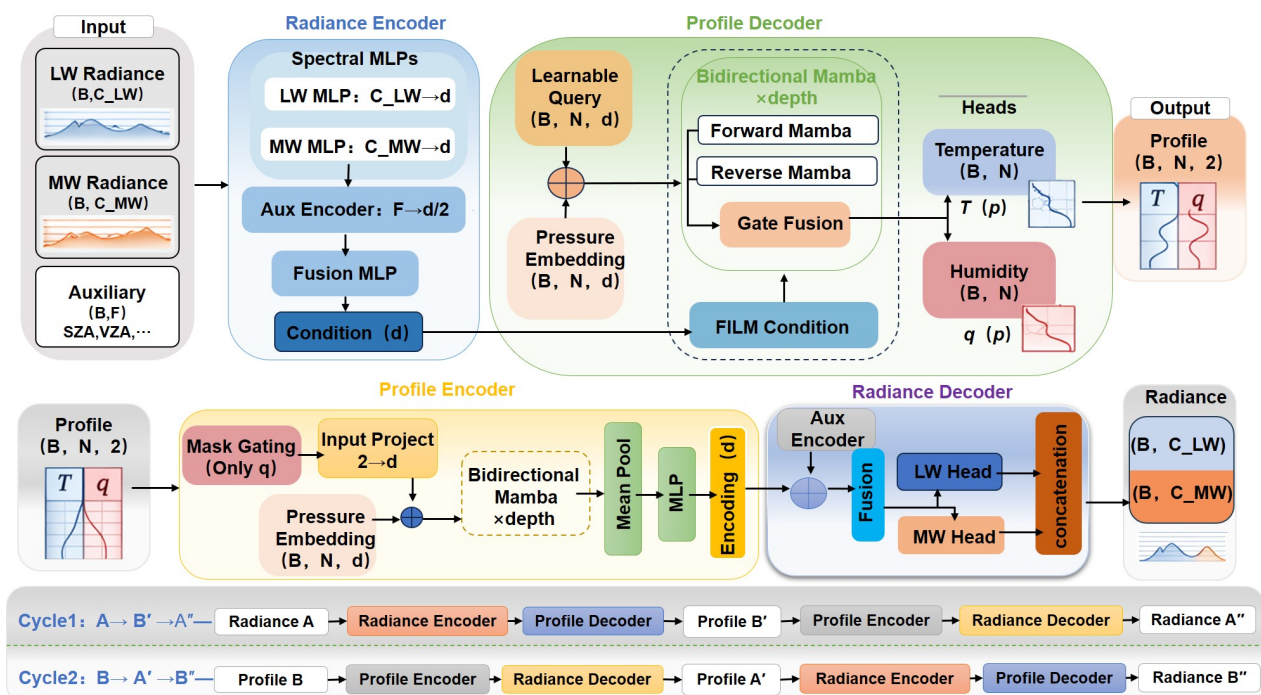


Figure 1. Overall architecture of SIMBA. The framework consists of a retrieval branch and a forward simulation branch for bidirectional radiance–profile modeling. A bidirectional Mamba module is used for vertical sequence modeling, while FiLM conditioning injects radiative information into the profile generation process. The two branches are coupled through a cycle-consistency constraint.

SIMBA uses GIIRS LW and MW radiance observations together with auxiliary variables for radiance-to-profile retrieval, while atmospheric profiles are used in the forward branch for radiance reconstruction. ERA5 reanalysis profiles serve as supervisory reference states during training.

2.3.1. Retrieval and Forward Simulation Branches

In the retrieval branch, LW and MW radiances are first encoded by two spectral multilayer perceptron (MLP) encoders and then fused with auxiliary variables to form a unified conditional vector. The fused features are passed to a profile decoder composed of stacked bidirectional Mamba blocks, which model the vertical dependencies of temperature and specific humidity along pressure levels. Pressure-level positional encoding is introduced to preserve vertical structure. To inject radiative information into the profile decoder, the unified conditional vector is further used as the FiLM condition. Specifically, this vector summarizes the radiative and geometric information of each GIIRS sample and is passed through a fully connected layer to generate two sample-dependent modulation vectors, namely the scaling vector $\gamma(c)$ and the offset vector $\beta(c)$. These two vectors are then applied to the hidden states of the profile decoder, where $\gamma(c)$ adaptively adjusts the amplitude of the hidden features and $\beta(c)$ shifts their feature responses. In this way, the vertical profile generation process is dynamically conditioned on the radiative characteristics of each observation through a Feature-wise Linear Modulation (FiLM) mechanism [47]:

$$\text{FiLM}(h, c) = \gamma(c) \odot h + \beta(c) \quad (1)$$

where h denotes the hidden state, c denotes the conditional vector, and $\gamma(\cdot)$ and $\beta(\cdot)$ are scaling and offset parameters generated by a fully connected layer.

In the forward simulation branch, temperature and specific humidity profiles are encoded into latent atmospheric-state representations through a profile encoder with bidirectional Mamba blocks and pressure-level positional encoding. The encoded features are then mapped to the LW and MW radiance spaces through a radiance decoder with band-specific output heads, enabling profile-to-radiance reconstruction.

The two branches are coupled through a cycle-consistency constraint: profiles retrieved from radiances are required to reconstruct the original observations through the forward branch, while radiances reconstructed from reference profiles are required to recover the corresponding atmospheric profiles through the retrieval branch. This bidirectional closed-loop design improves the consistency and stability of radiance–profile modeling.

2.3.2. Loss Function

To jointly optimize the retrieval and forward simulation branches, we define a multi-objective loss function consisting of profile retrieval losses, radiance reconstruction losses, and a cycle-consistency loss.

The profile retrieval losses constrain the predicted temperature and specific humidity profiles using the corresponding ERA5 reference profiles:

$$\mathcal{L}_T = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{T}}_i - \mathbf{T}_i\|^2, \quad \mathcal{L}_q = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{q}}_i - \mathbf{q}_i\|^2 \quad (2)$$

where $\hat{\mathbf{T}}_i$ and $\hat{\mathbf{q}}_i$ denote the predicted temperature and specific humidity profiles, and $\mathbf{T}_i$ and $\mathbf{q}_i$ denote the corresponding ERA5 reference profiles.

The radiance reconstruction losses are defined for the LW and MW bands as

$$\mathcal{L}_{\text{LW}} = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{R}}_i^{\text{LW}} - \mathbf{R}_i^{\text{LW}}\|^2, \quad \mathcal{L}_{\text{MW}} = \frac{1}{N} \sum_{i=1}^N \|\hat{\mathbf{R}}_i^{\text{MW}} - \mathbf{R}_i^{\text{MW}}\|^2 \quad (3)$$

where $\hat{\mathbf{R}}_i^{\text{LW}}$ and $\hat{\mathbf{R}}_i^{\text{MW}}$ denote the reconstructed LW and MW radiances, and $\mathbf{R}_i^{\text{LW}}$ and $\mathbf{R}_i^{\text{MW}}$ denote the corresponding GIIRS observations.

The cycle-consistency loss is defined as

$$\mathcal{L}_{\text{cycle}} = \frac{1}{N} \sum_{i=1}^N \left(\|G(F(\mathbf{R}_i)) - \mathbf{R}_i\|^2 + \|F(G(\mathbf{P}_i)) - \mathbf{P}_i\|^2 \right), \quad (4)$$

where F denotes the retrieval mapping (radiance $\rightarrow$ profile), G denotes the forward simulation mapping (profile $\rightarrow$ radiance), $\mathbf{R}_i$ denotes the input GIIRS radiance observation, and $\mathbf{P}_i$ denotes the corresponding ERA5 reference profile.

The total loss is formulated as

$$\mathcal{L}_{\text{total}} = \alpha_T \mathcal{L}_T + \alpha_q \mathcal{L}_q + \alpha_{\text{LW}} \mathcal{L}_{\text{LW}} + \alpha_{\text{MW}} \mathcal{L}_{\text{MW}} + \beta \mathcal{L}_{\text{cycle}}, \quad (5)$$

where α_T , α_q , α_{LW} , α_{MW} , and β are hyperparameters controlling the relative contributions of the loss terms.

We implemented SIMBA and all baseline models using PyTorch 2.7.0. The experimental dataset comprises FY-4A GIIRS radiation observations and ERA5 reanalysis profiles. Each sample contains preprocessed LW and MW radiation vectors, auxiliary information, pressure coordinates, and corresponding reference temperature and specific humidity profiles. The dataset was partitioned into training, validation, and test sets at an 8:1:1 ratio, maintaining temporal sequence independence during splitting to objectively evaluate model generalization. Training employed the AdamW optimiser with an initial learning rate of 1×10^{-3} and weight decay of 1×10^{-4} , combined with ReduceLROnPlateau learning rate scheduling to dynamically adjust the rate based on validation loss. Training was conducted on an NVIDIA RTX 4090 GPU with 24 GB of VRAM, employing a batch size of 128 and running for 100 epochs. All input data underwent standardization prior to training.

2.4. Evaluation Metrics

To quantitatively evaluate the performance of the proposed framework and baseline models, we used three standard regression metrics, namely root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). These metrics measure the magnitude of prediction errors and the overall goodness of fit. The definitions are given as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (6)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

where n is the number of samples, y_i is the reference value, $\hat{y}_i$ is the model prediction, and $\bar{y}$ is the sample mean of the reference values. In the subsequent analysis, lower RMSE and MAE indicate better predictive accuracy, while an R^2 value closer to 1 indicates a better overall fit.

3. Results and Discussion

3.1. Overall Evaluation

Bidirectional modeling has been applied in satellite data retrieval, but not yet in hyperspectral retrieval of temperature and humidity, where CNN [48], MLP [49], Transformer [50], and LSTM [51] remain widely used. We therefore augment these architectures with bidirectional processing, yielding Bidirectional CNN (BiCNN), Bidirectional MLP (BiMLP), Bidirectional Transformer (BiTransformer), and Bidirectional LSTM (BiLSTM) as comparison methods.

Table 2 reports the mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) for temperature and specific humidity profile retrieval. The LW and MW radiance reconstruction errors are also included to evaluate the consistency between the retrieved atmospheric states and the observed radiances. To examine cross-condition generalization, the results are divided into cloudy-sky testing and clear-sky generalization, as shown in Table 2(a) and Table 2(b), respectively.

Table 2. Comparative performance of SIMBA and the baseline models under cloudy-sky and clear-sky conditions. (a) Performance on the cloudy cases. (b) Performance on cases under clear-sky conditions.

(a) Performance on the cloudy cases.												
Model	Temperature			Humidity			Long-wave			Medium-wave		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
BiCNN	0.7399	1.0489	0.9990	0.2601	0.5815	0.9891	4.0954	6.3157	0.8901	0.2712	0.3923	0.8511
BiMLP	0.8265	1.1911	0.9987	0.2753	0.6171	0.9877	4.0434	6.1790	0.8948	0.2897	0.4099	0.8375
BiTransformer	0.7744	1.0890	0.9989	0.2650	0.5949	0.9886	4.2335	6.4741	0.8846	0.2800	0.4009	0.8445
BiLSTM	0.7338	1.0346	0.9990	0.2578	0.5772	0.9892	4.1409	6.4350	0.8859	0.2774	0.4004	0.8450
SIMBA	0.7131	1.0032	0.9991	0.2530	0.5684	0.9896	3.9674	6.1228	0.8967	0.2711	0.3887	0.8538

(b) Performance on cases under clear-sky conditions.												
Model	Temperature			Humidity			Long-wave			Medium-wave		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
BiCNN	0.7955	1.3117	0.9984	0.4281	1.0801	0.9565	2.0381	2.9410	0.9851	0.2687	0.3637	0.9289
BiMLP	0.9137	1.5644	0.9977	0.4522	1.1222	0.9530	1.9585	2.8969	0.9856	0.2885	0.3944	0.9164
BiTransformer	0.8152	1.3035	0.9984	0.4172	1.0473	0.9591	2.7880	4.0576	0.9717	0.3262	0.4465	0.8929
BiLSTM	0.7266	1.1860	0.9987	0.4051	1.0430	0.9594	1.7530	2.5279	0.9890	0.2659	0.3626	0.9293
SIMBA	0.6958	1.1006	0.9989	0.4016	1.0320	0.9603	1.6894	2.4016	0.9901	0.2614	0.3550	0.9323

Table 2(a) shows that SIMBA performs best under cloudy-sky conditions. For temperature retrieval, it achieves an RMSE of 1.0032 K, an MAE of 0.7131 K, and an R^2 of 0.9991, reducing the RMSE by approximately 3.0% relative to BiLSTM. For specific humidity retrieval, SIMBA yields an RMSE of 0.5684, which is approximately 1.5% lower than that of BiLSTM. Radiance reconstruction errors are also the lowest for both the LW and MW bands. When evaluated on clear-sky samples (Table 2(b)), SIMBA continues to outperform all baselines, with RMSE reductions of 7.20% for temperature, 1.05% for specific humidity, 5.00% for LW radiance, and 2.10% for MW radiance relative to BiLSTM. This result indicates that cloudy-sky training does not restrict SIMBA to cloudy pixels; instead, the bidirectional radiance–profile mapping generalizes to clear-sky conditions.

Overall, SIMBA enhances retrieval and reconstruction under cloudy skies and generalizes to clear-sky conditions without retraining. These findings support the bidirectional framework and indicate that Mamba is well-suited for modeling vertical dependencies in atmospheric profiles.

We further evaluated the computational efficiency of SIMBA against all baseline models. All measurements were conducted on an NVIDIA RTX 4090 GPU (24 GB VRAM) with PyTorch 2.7.0, using identical preprocessing and inference settings. SIMBA has 1.13 M trainable parameters—fewer than BiTransformer (1.76 M) and comparable to BiLSTM (1.04 M). Its inference time is 2.159 s for 13,785 test samples (0.157 ms per sample), and its peak GPU memory usage is 146.50 MB. That is lower than BiLSTM (234.27 MB) but slightly higher than BiTransformer (130.60 MB). Inference speeds under cloudy-sky conditions are detailed in Table S3.

3.2. Evaluation of Atmospheric Profile Retrieval

3.2.1. Temperature Profiles

To further evaluate temperature profile retrieval performance, the overall scatter distributions, layered scatter characteristics, and vertical error profiles of different models are analyzed in this section.

Figure 2 presents the overall density scatter distributions of temperature retrievals on the test set for different models. In general, the predictions of all models follow the 1:1 reference line, indicating that they can reproduce the overall temperature distribution reasonably well. To quantitatively characterize the scatter dispersion, an expected error (EE) envelope of ± 1 K was introduced around the 1:1 line, and the percentage of temperature points within this envelope was calculated as the within-EE percentage. SIMBA achieves the highest within-EE value of 76.74%, compared with 75.54% for BiCNN, 71.69% for BiMLP, 73.56% for BiTransformer, and 75.53% for BiLSTM. In addition, SIMBA obtains the lowest RMSE of 1.0032 K and MAE of 0.7131 K, as well as the highest R^2 of 0.9991. These quantitative results indicate that SIMBA has a more compact error distribution and better overall agreement with the reference temperature than the comparison models.

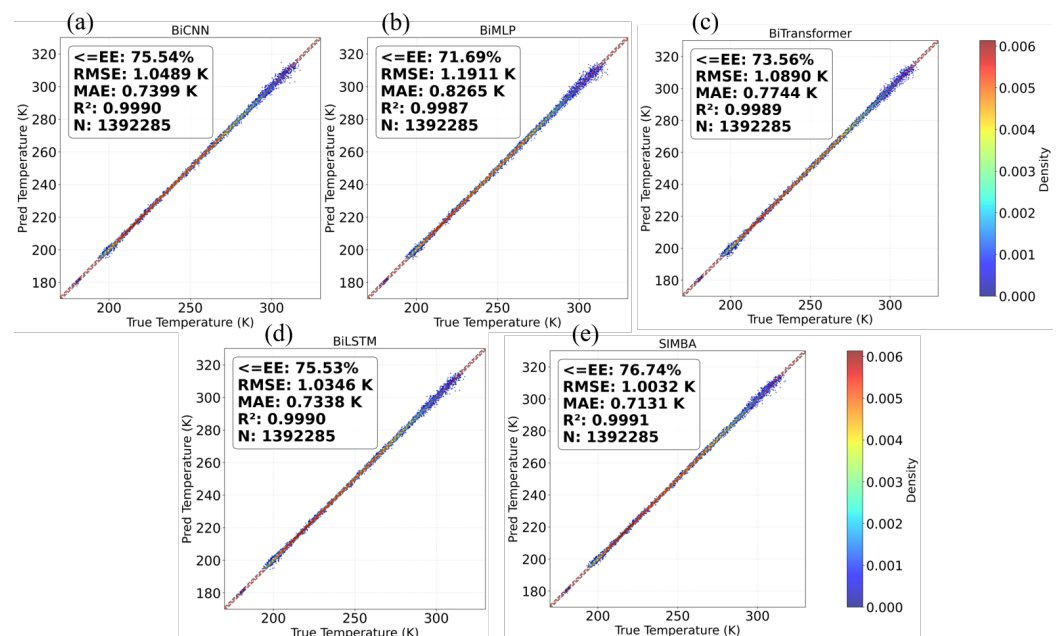


Figure 2. Overall density scatter comparison of temperature retrievals on the test set for different models: (a) BiCNN; (b) BiMLP; (c) BiTransformer; (d) BiLSTM; and (e) SIMBA. The red dashed line represents the 1:1 reference line, while the gray dashed lines represent the expected error (EE) envelope of ± 1 K. The color bar indicates the kernel density of the samples. The metric $\leq \text{EE}$ indicates the proportion of temperature points within the EE envelope, i.e., $|T_{\text{pred}} - T_{\text{true}}| \leq 1$ K. N denotes the total number of temperature points obtained by flattening all vertical levels of the test profiles.

The layered scatter plots in Figure 3 further illustrate temperature retrieval performance at six representative pressure levels (100, 300, 500, 700, 850, and 1000 hPa). Overall, the predicted values of all models generally follow the 1:1 reference line, but SIMBA shows a more concentrated distribution and reduced scatter at most pressure levels. The advantage is particularly evident in the near-surface layer (850–1000 hPa) and the upper-level region (100–300 hPa), where the retrieval task is more challenging.

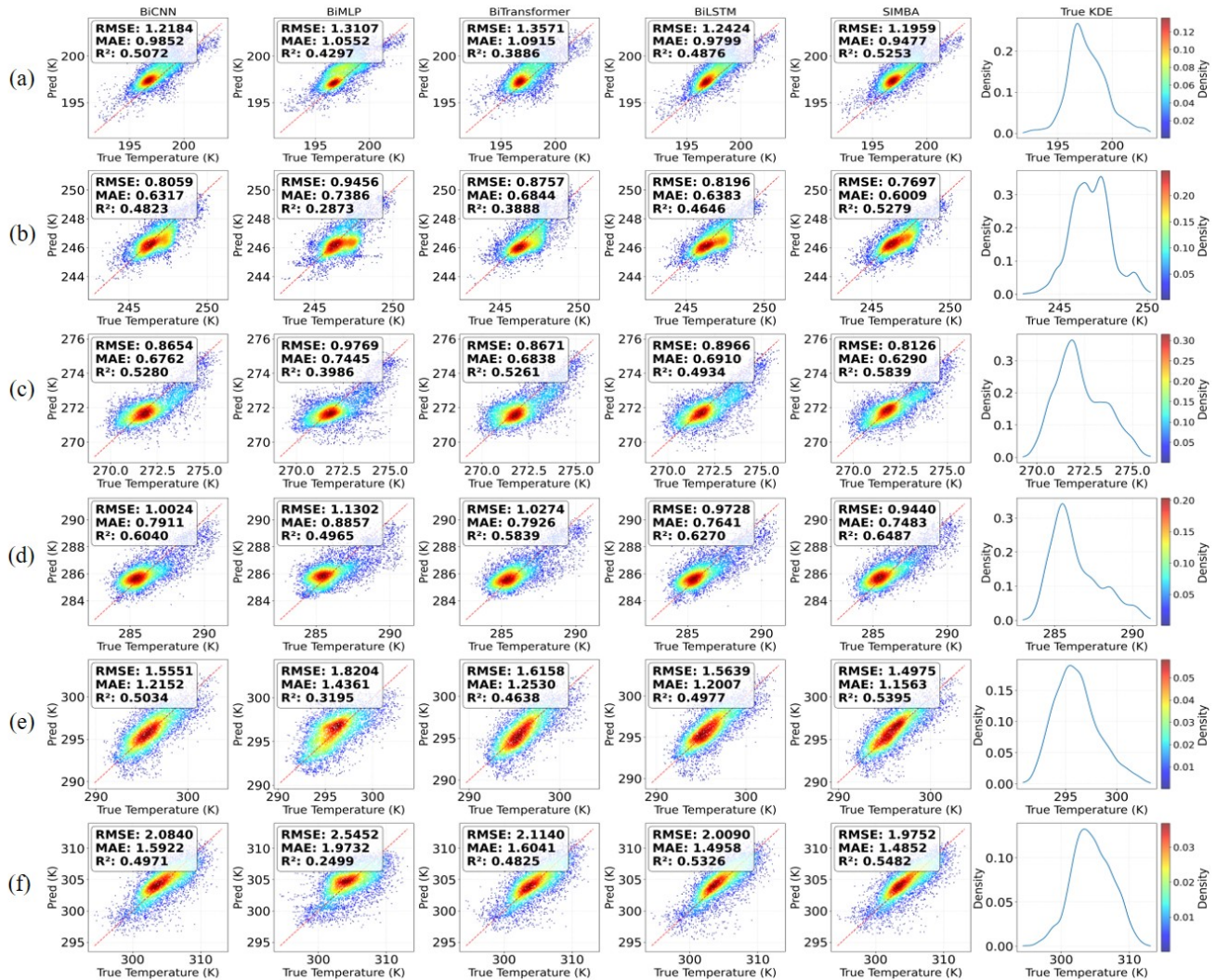


Figure 3. Scatter plot comparison of temperature retrieval at six representative pressure levels: (a) 100 hPa; (b) 300 hPa; (c) 500 hPa; (d) 700 hPa; (e) 850 hPa; (f) 1000 hPa. Columns correspond to BiCNN, BiMLP, BiTransformer, BiLSTM, and SIMBA, while the rightmost column shows the kernel density estimate (True KDE) of ERA5 temperature at each pressure level.

Figures 4 and 5 show the vertical distributions of RMSE and relative bias for temperature retrieval. Overall, SIMBA maintains comparatively low RMSE over most pressure levels, with more noticeable improvements below 700 hPa and above 300 hPa. The relative bias profiles further show that SIMBA exhibits smaller oscillation amplitudes throughout the vertical column, indicating a more stable error distribution across pressure levels. In addition, we conducted an independent validation using 83 collocated radiosonde profiles. Between 200 hPa and 800 hPa, the mean temperature bias is merely -0.1637 K with a standard deviation of 1.6833 K, while the mean specific-humidity bias is 0.1812 g kg $^{-1}$ with a standard deviation of 1.3435 g kg $^{-1}$, as detailed in Figure S2.

Overall, the temperature retrieval results demonstrate that SIMBA provides more accurate and vertically stable predictions than the comparison models. These results further support the effectiveness of the proposed bidirectional Mamba framework for modeling the vertical structure of atmospheric temperature profiles.

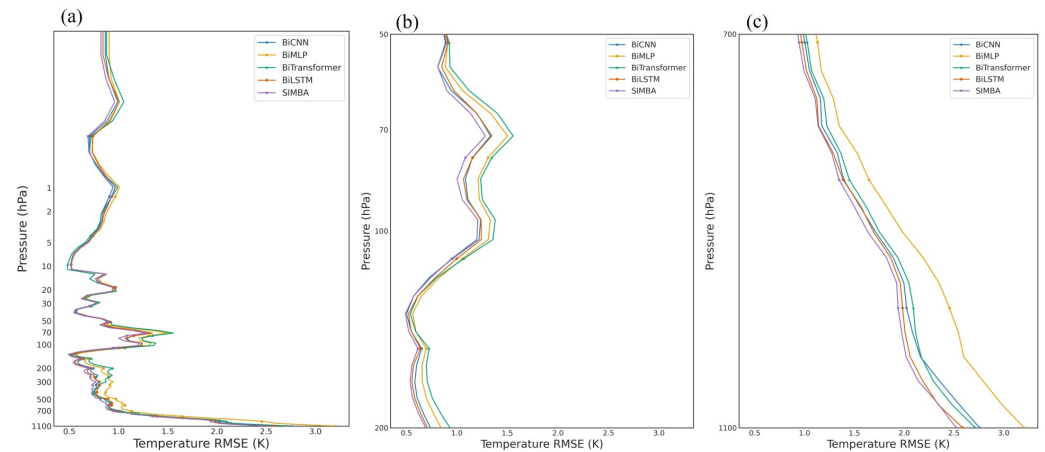


Figure 4. Vertical distribution of RMSE for temperature retrieval across different models: (a) full pressure range; (b) enlarged view of 50–200 hPa; (c) enlarged view of 700–1100 hPa.

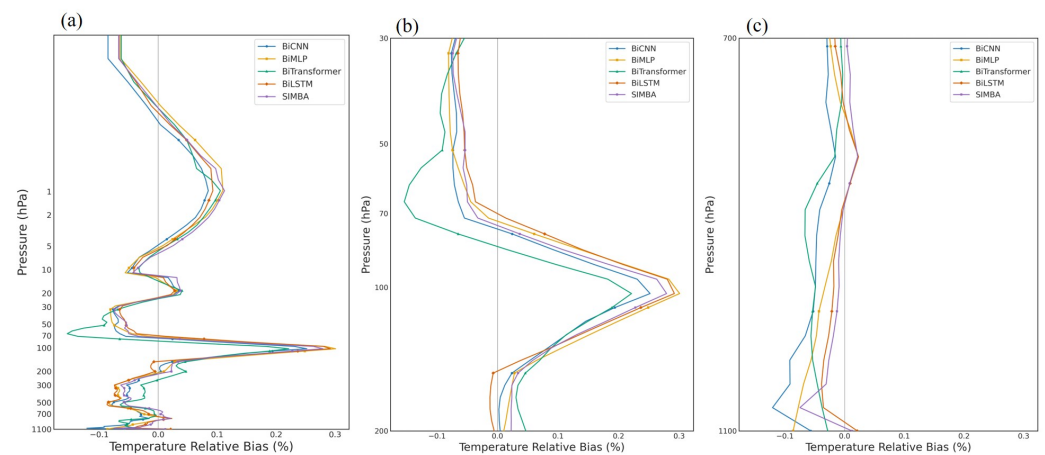


Figure 5. Vertical distribution of relative bias for temperature retrieval across different models: (a) full pressure range; (b) enlarged view of 30–200 hPa; (c) enlarged view of 700–1100 hPa.

3.2.2. Humidity Profiles

To further evaluate specific humidity retrieval performance, the overall scatter distributions, layered scatter characteristics, and vertical error profiles of different models are analyzed in this section. Compared with temperature retrieval, specific humidity retrieval is generally more challenging because of the stronger nonlinearity and vertical variability of water vapor.

Figure 6 presents the overall scatter distributions of specific humidity retrieval on the test set for different models. In general, the predictions of all models follow the 1:1 reference line, indicating that they can reproduce the overall humidity distribution reasonably well. However, SIMBA exhibits a narrower scatter dispersion and achieves the lowest overall RMSE and MAE, indicating superior fitting performance for specific humidity retrieval. Compared with temperature retrieval, the scatter becomes more dispersed in the high-humidity range (greater than 15 g/kg), reflecting the increased difficulty of humidity inversion. This visual pattern is consistent with the quantitative results in Table 2, where

SIMBA achieves the lowest specific humidity RMSE (0.5684) and MAE (0.2530), corresponding to an RMSE reduction of approximately 1.5% relative to the best baseline, BiLSTM.

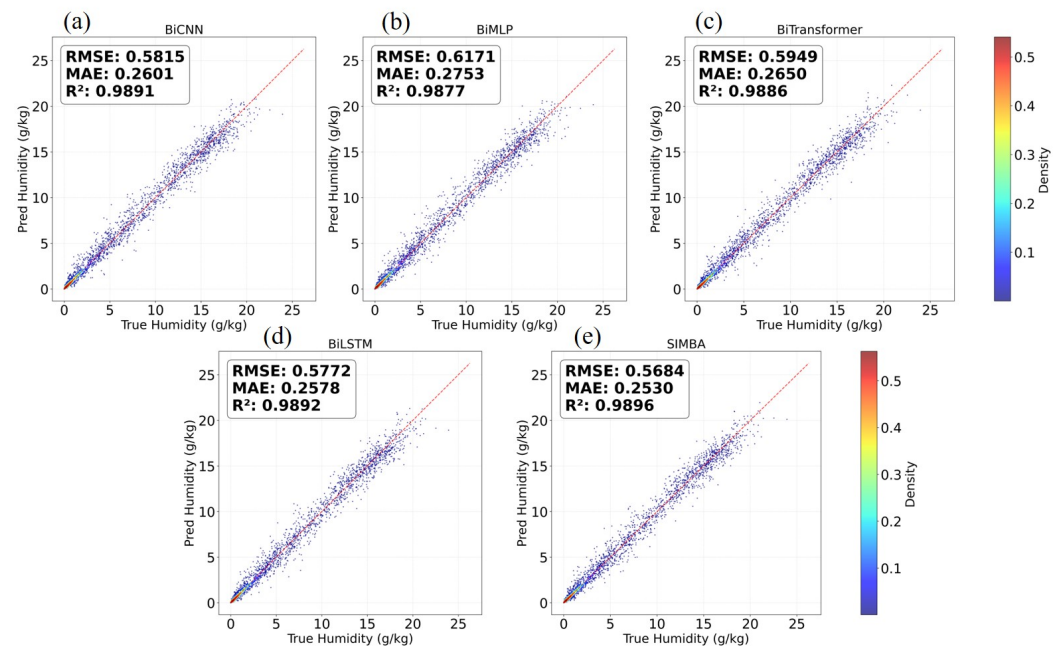


Figure 6. Overall scatter plot comparison of specific humidity retrieval on the test set for different models: (a) BiCNN; (b) BiMLP; (c) BiTransformer; (d) BiLSTM; (e) SIMBA.

The layered scatter plots in Figure 7 further demonstrate retrieval performance at six representative pressure levels (200, 500, 700, 850, 925, and 1000 hPa). Overall, SIMBA shows a more concentrated scatter distribution and lower overall error than most baseline models in the mid-to-lower troposphere, especially around 700 hPa and in the near-surface layers (850–1000 hPa), where humidity gradients are relatively strong and retrieval is more challenging. In the upper layer (around 200 hPa), the dispersion of all models increases, reflecting the greater uncertainty associated with the very low water vapor content at these altitudes. Specifically, at (a) 200 hPa, SIMBA achieves an MAE of 0.0213 g/kg and an RMSE of 0.0278 g/kg; at (b) 500 hPa, the MAE and RMSE increase to 0.6099 and 0.7978 g/kg, respectively. At (e) 925 hPa and (f) 1000 hPa, SIMBA yields MAE values of 0.7771 and 0.8270 g/kg with RMSE values of 1.0080 and 1.0761 g/kg, consistently outperforming all baseline models in these high-humidity regimes.

Figures 8 and 9 present the vertical distributions of RMSE and relative bias for specific humidity retrieval. Overall, SIMBA maintains comparatively low RMSE over most pressure levels, with particularly notable improvements in the lower troposphere between 700 and 1000 hPa. The relative bias profiles further show that SIMBA exhibits smaller oscillation amplitudes over the vertical column, indicating a more stable error distribution across pressure levels.

Overall, the humidity retrieval results demonstrate that SIMBA provides more accurate and vertically stable predictions than the comparison models, particularly in the moisture-rich lower troposphere. These results further support the effectiveness of the proposed bidirectional Mamba framework for modeling the vertically coupled structure of atmospheric moisture profiles.

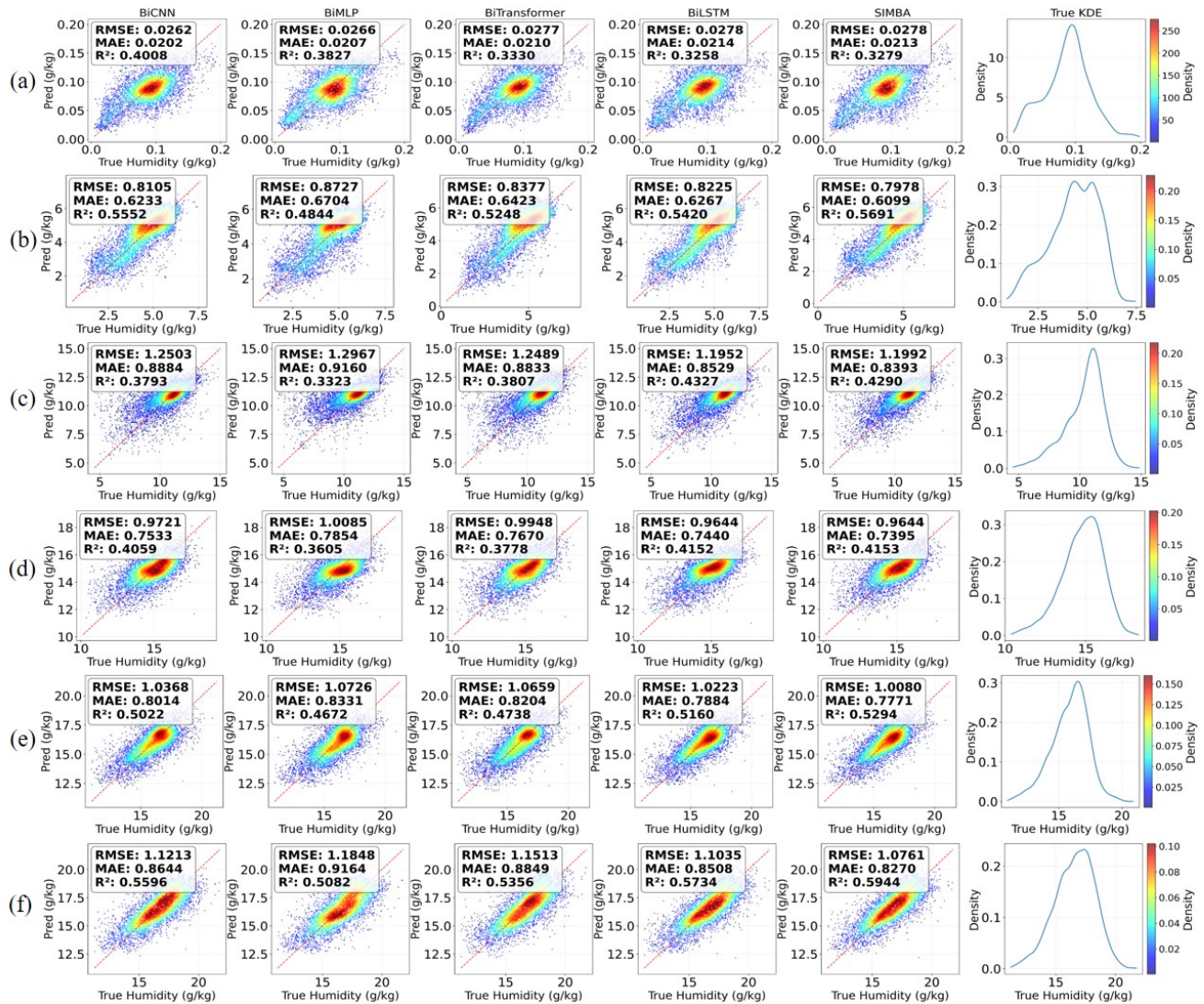


Figure 7. Scatter plot comparison of specific humidity retrieval at six representative pressure levels: (a) 200 hPa; (b) 500 hPa; (c) 700 hPa; (d) 850 hPa; (e) 925 hPa; (f) 1000 hPa. Columns correspond to BiCNN, BiMLP, BiTransformer, BiLSTM, and SIMBA, while the rightmost column shows the kernel density estimate (True KDE) of ERA5 specific humidity at each pressure level.

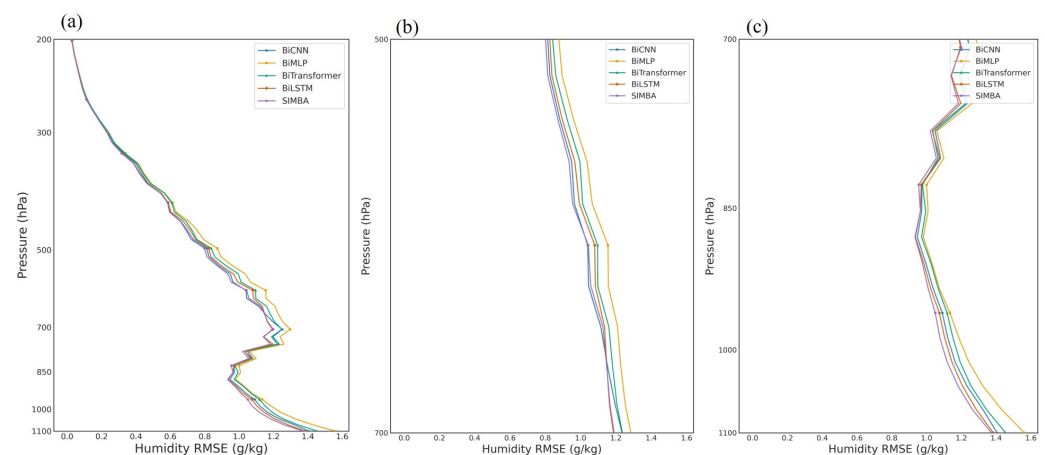


Figure 8. Vertical distribution of RMSE for specific humidity retrieval across different models: (a) full pressure range; (b) enlarged view of 500–700 hPa; (c) enlarged view of 700–1100 hPa.

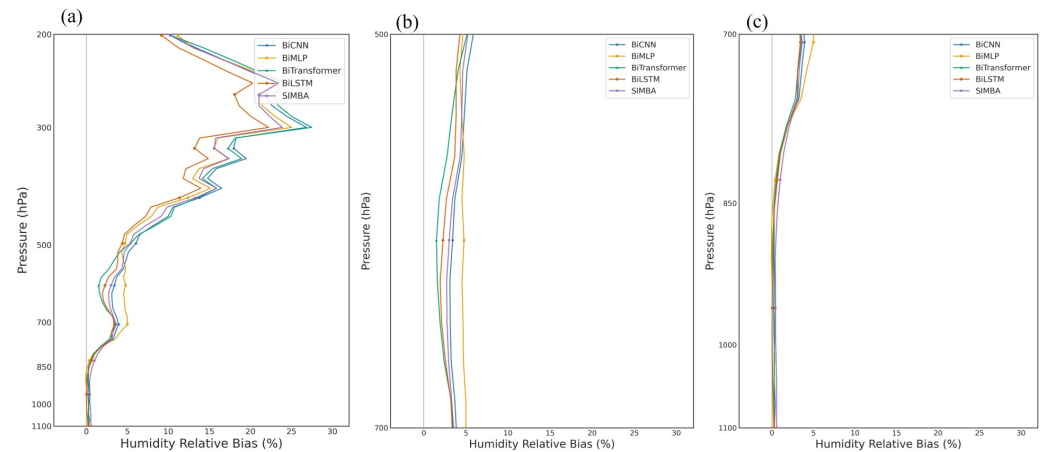


Figure 9. Vertical distribution of relative bias for specific humidity retrieval across different models: (a) full pressure range; (b) enlarged view of 200–500 hPa; (c) enlarged view of 700–1100 hPa.

3.3. Evaluation for Forward Models

3.3.1. Radiation of Long-Wave Bands

From the perspective of NWP data assimilation, forward consistency in observation space is important for assessing whether the retrieved atmospheric states remain compatible with the observed radiance field. Therefore, the LW radiance reconstruction results are analyzed in this section.

Figure 10 presents the overall scatter distributions of reconstructed LW radiances on the test set for different models. In general, the reconstructed values of all models follow the 1:1 reference line, indicating that the retrieved atmospheric profiles can reproduce the main characteristics of the observed LW radiances. However, SIMBA exhibits a more concentrated scatter distribution and a narrower dispersion range than the comparison models, and achieves the best overall performance in terms of RMSE, MAE, and R^2 . Although the numerical improvements in LW reconstruction are smaller than those in profile retrieval, the more compact scatter pattern indicates better consistency between the retrieved profiles and the observed radiances.

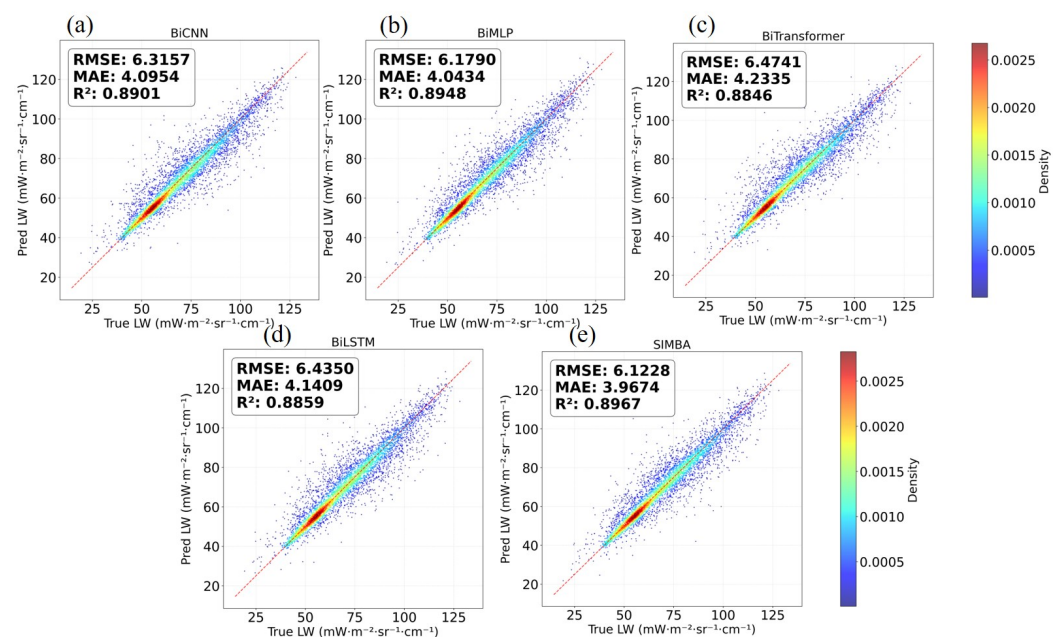


Figure 10. Overall scatter plot comparison of long-wave radiance reconstruction on the test set for different models: (a) BiCNN; (b) BiMLP; (c) BiTransformer; (d) BiLSTM; (e) SIMBA.

To further examine reconstruction differences across distinct spectral positions, six representative LW channels (CH1, CH20, CH35, CH60, CH80, and CH100) are selected for detailed analysis in Figure 11. Channels in the window region (e.g., CH1 and CH20) are more sensitive to near-surface thermodynamic structure, and all models show relatively good linear relationships in these channels. By contrast, channels within or near the CO₂ absorption region (e.g., CH60, CH80, and CH100) are more sensitive to the middle and upper troposphere, and the scatter distributions become more dispersed across all models. Nevertheless, SIMBA maintains comparatively lower dispersion and smaller systematic deviation in these absorption-sensitive channels.

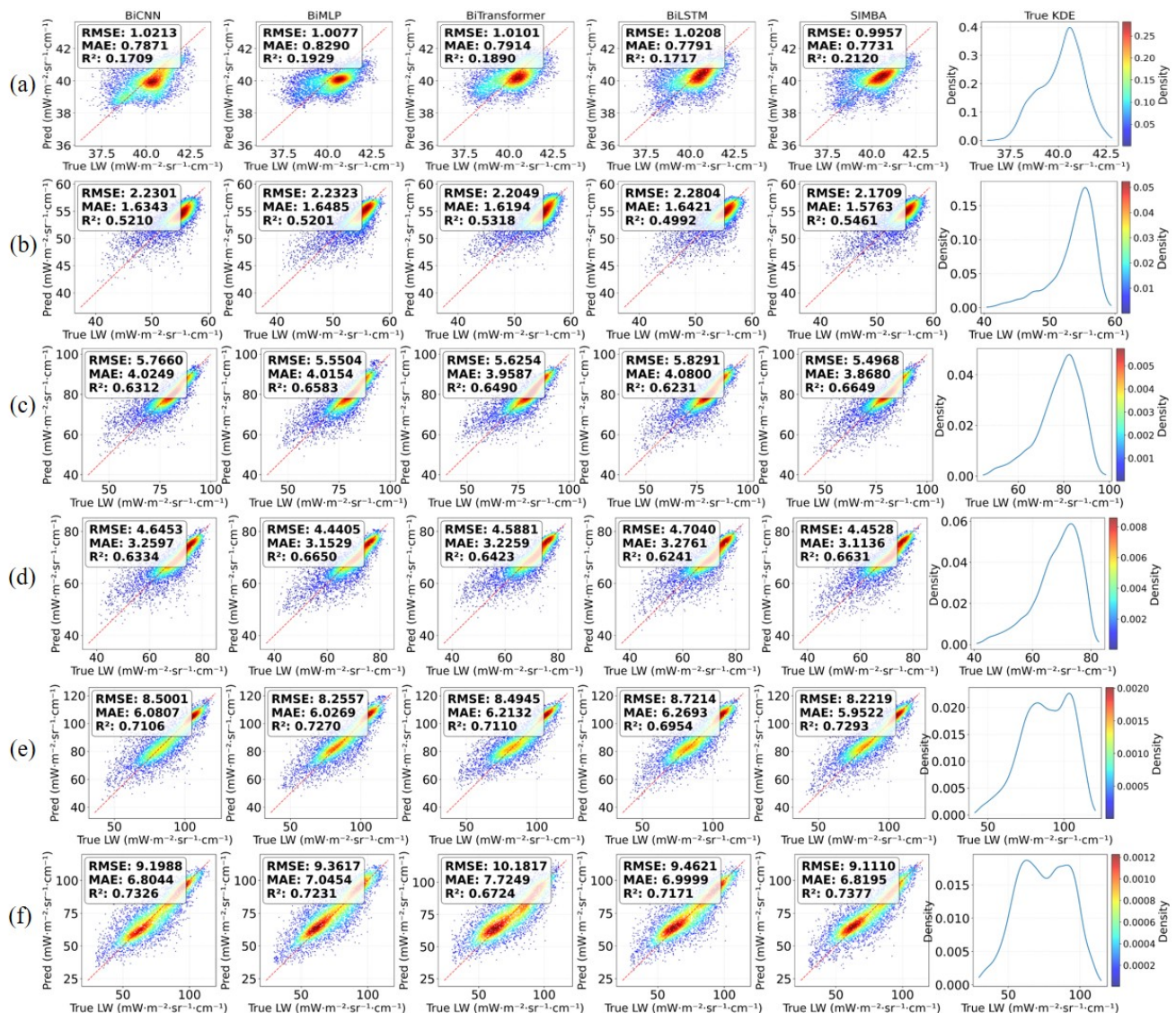


Figure 11. Scatter plot comparison of long-wave radiance reconstruction for six representative channels: (a) CH1; (b) CH20; (c) CH35; (d) CH60; (e) CH80; (f) CH100. Columns correspond to BiCNN, BiMLP, BiTransformer, BiLSTM, and SIMBA, while the rightmost column shows the kernel density estimate (True KDE) of the observed radiance at each channel.

Figures 12 and 13 present the RMSE and relative bias distributions across the 106 LW channels, respectively. The RMSE curves show that all models generally maintain relatively low errors in the window-region channels, whereas larger errors appear in channels associated with stronger absorption. SIMBA maintains lower or comparable RMSE values across most channels, with a more gradual increase in error and smaller fluctuations, especially in channels that are more sensitive to middle- and upper-level atmospheric structures. The rel-

active bias curves further show that some baseline models exhibit larger oscillations or localized peaks in certain absorption-sensitive channels, whereas SIMBA remains generally more stable without obvious local amplification.

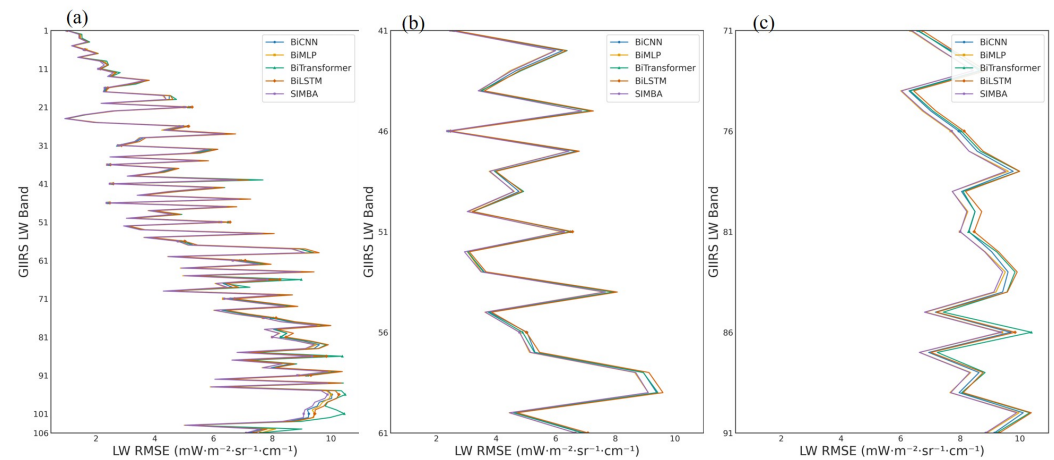


Figure 12. Distribution of RMSE across 106 long-wave channels for different models: (a) full channel range; (b) enlarged view of CH41–CH61; (c) enlarged view of CH71–CH91.

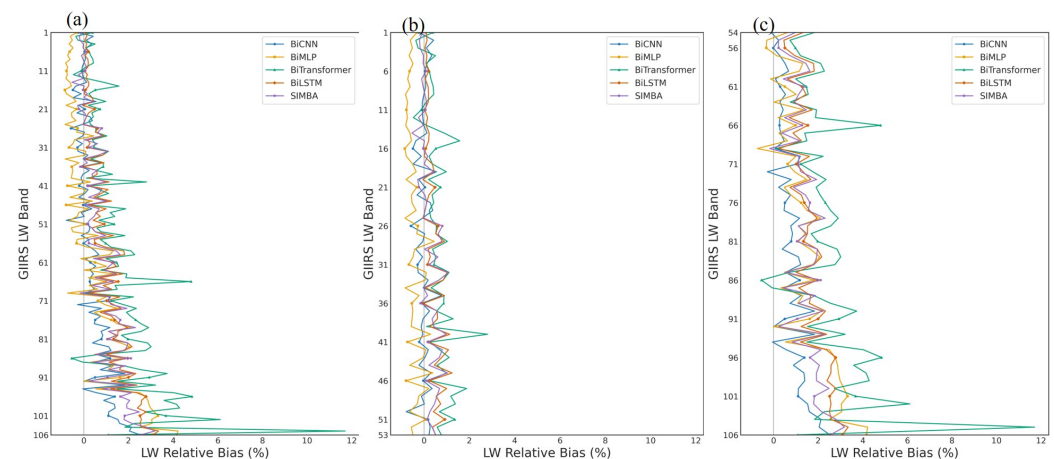


Figure 13. Distribution of relative bias across 106 long-wave channels for different models: (a) full channel range; (b) enlarged view of CH1–CH53; (c) enlarged view of CH54–CH106.

Overall, the LW radiance reconstruction results demonstrate that SIMBA provides better observation-space consistency and more stable channel-wise performance than the comparison models, indicating a more reliable radiance–profile mapping for NWP-oriented applications.

3.3.2. Radiation of Medium-Wave Bands

From the perspective of NWP data assimilation, forward consistency in observation space is important for assessing whether the retrieved atmospheric states remain compatible with the observed radiance field in different spectral regimes. Therefore, the MW radiance reconstruction results are analyzed in this section.

Figure 14 presents the overall scatter distributions of reconstructed MW radiances on the test set for different models. Overall, the reconstructed values of all models follow the 1:1 reference line, indicating that the retrieved atmospheric profiles can reproduce the main characteristics of the observed MW radiances. Compared with the LW band, the scatter distribution is generally more concentrated because of the smaller dynamic range of MW radiances, although some dispersion remains in the higher-value region.

SIMBA achieves the best overall performance in terms of RMSE and R^2 , indicating slightly better consistency between the retrieved profiles and the observed radiances. Although the numerical advantage is relatively moderate, the more compact scatter pattern suggests that the proposed framework remains effective in preserving radiative consistency in the MW band.

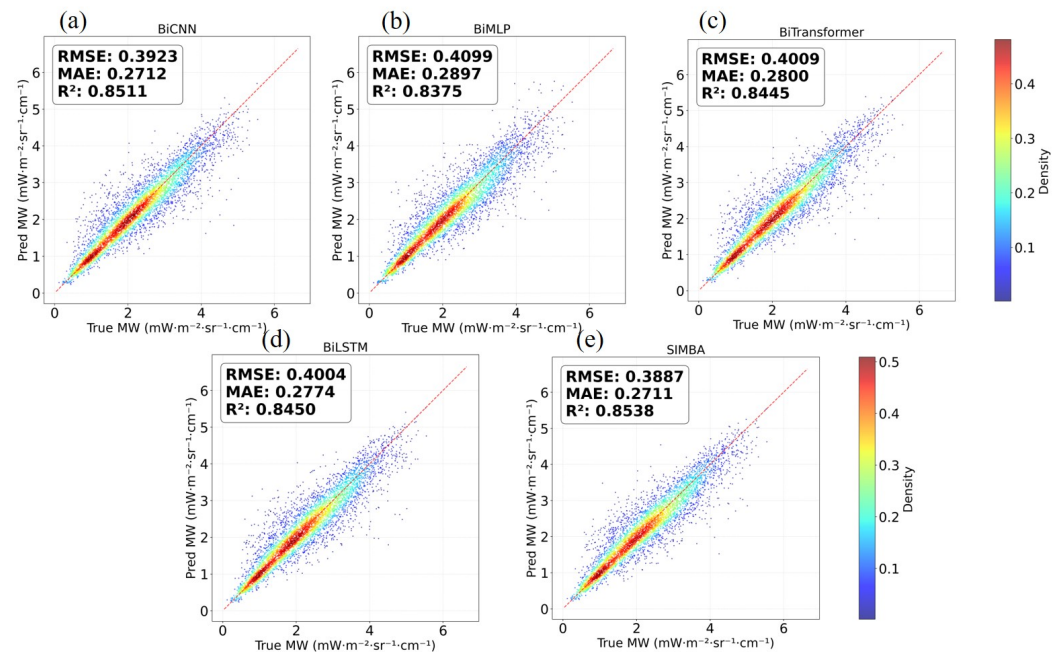


Figure 14. Overall scatter plot comparison of medium-wave radiance reconstruction on the test set for different models: (a) BiCNN; (b) BiMLP; (c) BiTransformer; (d) BiLSTM; (e) SIMBA.

To further analyze reconstruction differences across spectral positions, six representative channels (CH2, CH20, CH37, CH42, CH70, and CH88) are selected from the 99 MW channels for detailed analysis in Figure 15. In the window-region channels (e.g., CH2 and CH20), all models exhibit relatively good linear relationships. In the weak absorption-band channels (e.g., CH37 and CH42), the scatter becomes more dispersed, reflecting the increased uncertainty associated with channels sensitive to the upper troposphere. In relatively higher-level sensitive channels (e.g., CH70 and CH88), the dispersion further increases for all models. Nevertheless, SIMBA maintains a narrower scatter range and a more gradual error growth trend across most representative channels, indicating stronger reconstruction stability in spectrally sensitive MW regions.

Figures 16 and 17 present the RMSE and relative bias distributions across the 99 MW channels, respectively. Overall, all models exhibit relatively low errors in the window-region channels, whereas larger errors appear in weak absorption bands and in channels that are more sensitive to the middle and upper troposphere. The channel-wise errors also show a certain oscillatory structure with increasing channel number. In comparison, SIMBA maintains lower or comparable RMSE levels across most channels, with relatively smaller fluctuations and without obvious localized amplification. The relative bias curves further show that, although fluctuations remain in certain channels, SIMBA is generally more stable than several baseline models and does not exhibit obvious local amplification in the medium- to high-numbered channels.

Overall, the MW radiance reconstruction results demonstrate that SIMBA provides better observation-space consistency and more stable channel-wise performance than the comparison models. This behavior is consistent with the trends observed in the LW band and further indicates a more reliable radiance–profile mapping for NWP-oriented applications.

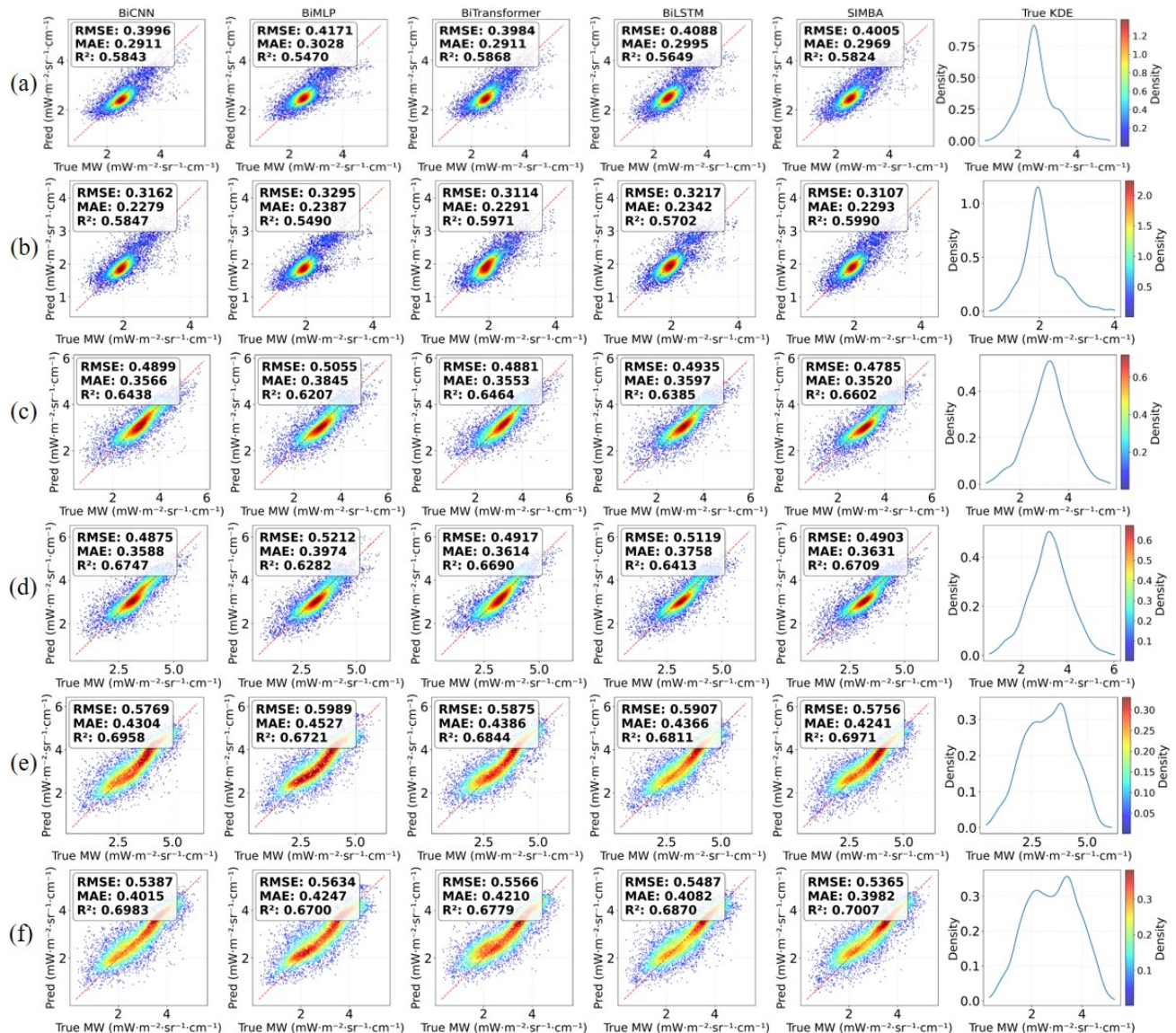


Figure 15. Scatter plot comparison of medium-wave radiance reconstruction for six representative channels: (a) CH2; (b) CH20; (c) CH37; (d) CH42; (e) CH70; (f) CH88. Columns correspond to BiCNN, BiMLP, BiTransformer, BiLSTM, and SIMBA, while the rightmost column shows the kernel density estimate (True KDE) of the observed radiance at each channel.

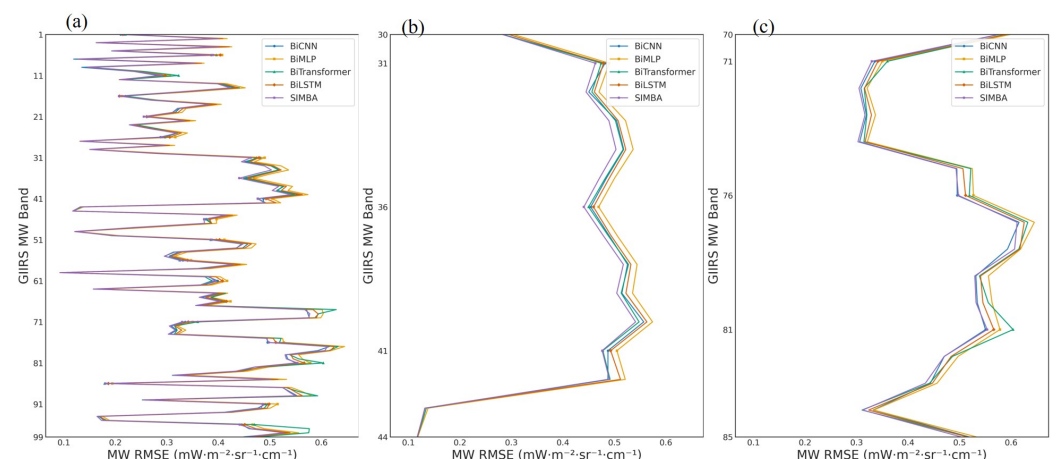


Figure 16. Distribution of RMSE across 99 medium-wave channels for different models: (a) full channel range; (b) enlarged view of CH30-CH44; (c) enlarged view of CH70-CH85.

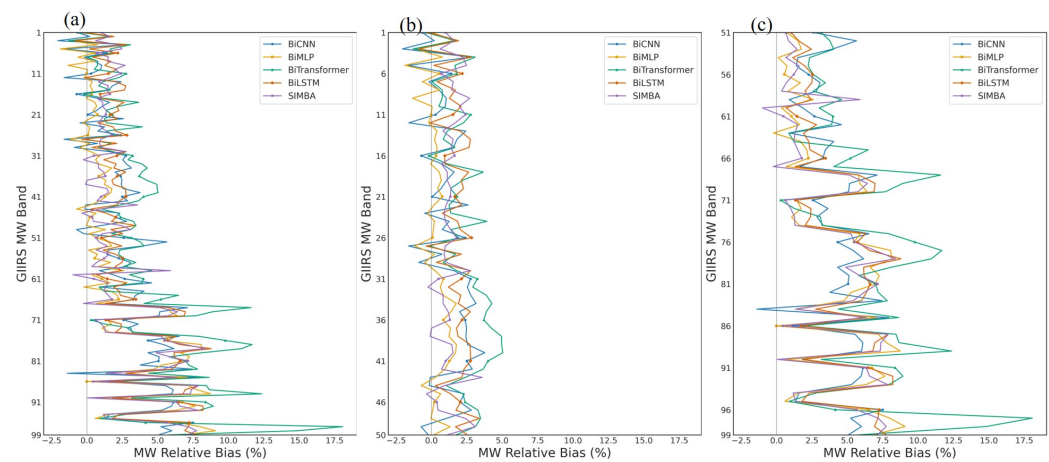


Figure 17. Distribution of relative bias across 99 medium-wave channels for different models: (a) full channel range; (b) enlarged view of CH1–CH50; (c) enlarged view of CH51–CH99.

3.4. Ablation Study

To isolate the contribution of the bidirectional closed-loop design, ablation experiments compare three variants under the same Mamba backbone, data split, preprocessing, training protocol, and evaluation metrics: Retrieval-Mamba (BT $\rightarrow$ profile, retrieval-only), Forward-Mamba (profile $\rightarrow$ BT, forward-only), and SIMBA (bidirectional closed-loop with cycle consistency). Results are averaged over three repeated runs per setting.

As shown in Table 3, SIMBA outperforms both baselines across all four variables. For temperature retrieval, SIMBA achieves an MAE of 0.7141 (vs. 0.7372 for Retrieval-Mamba, roughly 3.1% lower) and an RMSE of 1.0057 (vs. 1.0382, roughly 3.1% lower). For humidity, SIMBA reduces MAE from 0.2619 to 0.2527 (roughly 3.5% lower) and RMSE from 0.5875 to 0.5677 (roughly 3.4% lower) relative to Retrieval-Mamba. For radiance reconstruction, SIMBA yields lower MAE and RMSE than Forward-Mamba on both LW (MAE: 3.9527 vs. 4.0335, roughly 2.0% lower; RMSE: 6.0977 vs. 6.2830, roughly 2.9% lower) and MW (MAE: 0.2712 vs. 0.2781, roughly 2.5% lower; RMSE: 0.3879 vs. 0.3976, roughly 2.4% lower), with higher R^2 values in all cases. These results confirm that SIMBA's gains are not solely attributable to the Mamba backbone—coupling retrieval and forward simulation through the bidirectional closed-loop mechanism improves both retrieval accuracy and radiance reconstruction consistency.

Table 3. Ablation results of SIMBA and baseline models.

Model	Temperature			Humidity			Long-Wave			Medium-Wave		
	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
Retrieval-Mamba	0.7372	1.0382	0.9990	0.2619	0.5875	0.9888	—	—	—	—	—	—
Forward-Mamba	—	—	—	—	—	—	4.0335	6.2830	0.8912	0.2781	0.3976	0.8470
SIMBA	0.7141	1.0057	0.9991	0.2527	0.5677	0.9896	3.9527	6.0977	0.8975	0.2712	0.3879	0.8544

3.5. Discussion: Differentiable Forward Modeling and Remaining Validation Requirements

The above results show that SIMBA improves both atmospheric profile retrieval and radiance reconstruction through explicit coupling between atmospheric state space and radiance observation space. From an NWP-oriented perspective, the forward simulation branch can be regarded as a data-driven differentiable forward operator with RTM-like functionality under the evaluated conditions, rather than as a complete replacement for a physical radiative transfer model, as illustrated in Figure S3. It learns the mapping

from atmospheric thermodynamic states to top-of-atmosphere radiances within a unified bidirectional framework:

$$\hat{\mathbf{P}}_i = F_\theta(\mathbf{R}_i, \mathbf{a}_i), \quad \hat{\mathbf{R}}_i = G_\phi(\mathbf{P}_i, \mathbf{a}_i). \quad (9)$$

Because the forward branch is fully differentiable, its local sensitivity to atmospheric profile perturbations can be characterized by the Jacobian

$$\mathbf{J}_G(\mathbf{P}_i) = \frac{\partial G_\phi(\mathbf{P}_i, \mathbf{a}_i)}{\partial \mathbf{P}_i} = \left[\frac{\partial \hat{\mathbf{R}}_i}{\partial \mathbf{T}_i}, \frac{\partial \hat{\mathbf{R}}_i}{\partial \mathbf{q}_i} \right], \quad (10)$$

and the corresponding radiance response to a small perturbation $\delta \mathbf{P}_i$ can be approximated as

$$G_\phi(\mathbf{P}_i + \delta \mathbf{P}_i, \mathbf{a}_i) \approx G_\phi(\mathbf{P}_i, \mathbf{a}_i) + \mathbf{J}_G(\mathbf{P}_i) \delta \mathbf{P}_i. \quad (11)$$

This formulation provides a possible basis for future Jacobian-related interpretation and channel-sensitivity analysis. However, differentiability alone does not demonstrate equivalence to a physical RTM or guarantee physically consistent sensitivity estimates.

In this study, we focus on validating the bidirectional radiance–profile modeling capability of the proposed framework and its consistency in observation space. Nevertheless, rigorous validation using independent in situ observations and quantitative Jacobian comparisons with physical RTMs, such as RTTOV, is still required before the framework can be considered for operational assimilation-oriented applications.

4. Conclusions

This study proposed SIMBA, a simulation-constrained bidirectional Mamba framework for retrieving atmospheric temperature and specific humidity profiles from FY-4A/GIIRS observations. By coupling retrieval (radiance $\rightarrow$ profile) and forward simulation (profile $\rightarrow$ radiance) within a unified architecture, SIMBA enables joint optimization across atmospheric state space and radiance observation space via a cycle-consistency mechanism, while the Mamba backbone captures long-range vertical dependencies in thermodynamic profiles.

We evaluated SIMBA against several baseline models under both cloudy-sky and clear-sky conditions. Experimental results demonstrate that SIMBA consistently achieves the best RMSE, MAE, and R^2 across temperature and specific humidity retrieval, and yields the best LW and MW radiance reconstruction. Further analyses of vertical error distributions, channel-wise errors, and ablation experiments confirm that the bidirectional closed-loop design improves both retrieval accuracy and radiance reconstruction stability. Overall, jointly constraining state space and observation space benefits hyperspectral infrared retrieval under complex radiative transfer conditions, and provides a useful basis for NWP-oriented radiance–profile modeling.

Several limitations should be acknowledged. First, SIMBA should be regarded as a data-driven differentiable forward-modeling framework rather than a complete replacement for a physical RTM—residual retrieval and reconstruction errors remain non-negligible. Second, training and evaluation rely on ERA5 reanalysis as supervisory references. The reported accuracy therefore reflects consistency with ERA5 rather than independent validation against radiosonde observations, and the spatial-scale mismatch between GIIRS footprint and ERA5 grid may further introduce representativeness errors. Third, clear-sky and cloudy-sky conditions are not modeled separately. Both temperature and specific humidity retrieval accuracy under clear-sky conditions is lower than under cloudy skies (Table 2(b)), indicating that separate modeling per sky condition could yield improvements.; evaluation across broader seasons and regions is also needed. Fourth,

the model achieved good fitting performance, but the results lack physical interpretability. To assess what the model has actually learned, we conducted a perturbation sensitivity analysis following the RTM approach (Figure S3). The resulting channel sensitivities contain substantial high-frequency noise, which limits their physical interpretability and remains an open problem for future work.

Future work will address these limitations along several directions. Independent validation using radiosonde and other in situ observations will be conducted to assess model performance beyond ERA5 consistency. Separate modeling for clear-sky and cloudy-sky conditions will be explored, together with evaluation across broader seasons and regions. Longer time series will be incorporated to better capture seasonal and interannual variability. Finally, quantitative SIMBA–RTTOV Jacobian comparisons and noise-suppression techniques will be pursued to improve the physical interpretability of the learned forward sensitivities.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/doi/s1>.

Author Contributions: Conceptualization, J.S. and F.W.; methodology, J.S.; software, J.S.; validation, J.S., H.H. and C.Y.; formal analysis, J.S.; investigation, J.S.; resources, X.L. and F.W.; data curation, J.S. and C.W.; writing—original draft preparation, J.S.; writing—review and editing, J.S., H.H., C.Y., X.L. and F.W.; visualization, J.S.; supervision, Q.L., C.W., X.L. and F.W.; project administration, F.W. and Q.L.; funding acquisition, X.L. and F.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 No. 42471437).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Acknowledgments: We are grateful to the anonymous reviewers for their precious opinions and suggestions, whose professional reviews conspicuously enhanced the quality of this paper. Furthermore, our appreciation goes to the editorial team for their painstaking efforts and guidance throughout the publication process.

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

References

1. Brotzge, J.A.; Berchoff, D.; Carlis, D.L.; Carr, F.H.; Carr, R.H.; Gerth, J.J.; Gross, B.D.; Hamill, T.M.; Haupt, S.E.; Jacobs, N.; et al. Challenges and opportunities in numerical weather prediction. *Bull. Am. Meteorol. Soc.* **2023**, *104*, E698–E705. [\[CrossRef\]](#)
2. Lei, L.; Weng, F.; Duan, W.; Chen, Y.; Zhang, L.; Wang, R.; Yang, J.; Qin, X.; Han, W.; Li, J.; et al. Overview and prospect of data assimilation in numerical weather prediction. *J. Meteorol. Res.* **2025**, *39*, 559–592. [\[CrossRef\]](#)
3. Carminati, F.; Xiao, X.; Lu, Q.; Atkinson, N.; Hocking, J. Assessment of the hyperspectral infrared atmospheric sounder (HIRAS). *Remote Sens.* **2019**, *11*, 2950. [\[CrossRef\]](#)
4. Zhao, Q.; Yu, L.; Du, Z.; Peng, D.; Hao, P.; Zhang, Y.; Gong, P. An overview of the applications of earth observation satellite data: Impacts and future trends. *Remote Sens.* **2022**, *14*, 1863. [\[CrossRef\]](#)
5. Okamoto, K.; Owada, H.; Fujita, T.; Kazumori, M.; Otsuka, M.; Seko, H.; Ota, Y.; Uekiyo, N.; Ishimoto, H.; Hayashi, M.; et al. Assessment of the potential impact of a hyperspectral infrared sounder on the Himawari follow-on geostationary satellite. *SOLA* **2020**, *16*, 162–168. [\[CrossRef\]](#)
6. Yin, R.; Han, W.; Gao, Z.; Li, J. Impact of high temporal resolution FY-4A Geostationary Interferometric Infrared Sounder (GIIRS) radiance measurements on Typhoon forecasts: Maria (2018) case with GRAPES global 4D-Var assimilation system. *Geophys. Res. Lett.* **2021**, *48*, e2021GL093672.

7. Yang, J.; Zhang, Z.; Wei, C.; Lu, F.; Guo, Q. Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1637–1658. [\[CrossRef\]](#)
8. Chahine, M.T.; Pagano, T.S.; Aumann, H.H.; Atlas, R.; Barnett, C.; Blaisdell, J.; Chen, L.; Divakarla, M.; Fetzer, E.J.; Goldberg, M.; et al. AIRS: Improving weather forecasting and providing new data on greenhouse gases. *Bull. Am. Meteorol. Soc.* **2006**, *87*, 911–926.
9. Schmit, T.J.; Griffith, P.; Gunshor, M.M.; Daniels, J.M.; Goodman, S.J.; Lebair, W.J. A closer look at the ABI on the GOES-R series. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 681–698. [\[CrossRef\]](#)
10. Rodgers, C.D. *Inverse Methods for Atmospheric Sounding: Theory and Practice*; World Scientific: Singapore, 2000; Volume 2.
11. Yang, J.; Ding, S.; Dong, P.; Bi, L.; Yi, B. Advanced radiative transfer modeling system developed for satellite data assimilation and remote sensing applications. *J. Quant. Spectrosc. Radiat. Transf.* **2020**, *251*, 107043. [\[CrossRef\]](#)
12. Lu, Q.; Hu, J.; Wu, C.; Qi, C.; Wu, S.; Xu, N.; Sun, L.; Li, X.; Liu, H.; Guo, Y.; et al. Monitoring the performance of the Fengyun satellite instruments using radiative transfer models and NWP fields. *J. Quant. Spectrosc. Radiat. Transf.* **2020**, *255*, 107239. [\[CrossRef\]](#)
13. Eyre, J.; Bell, W.; Cotton, J.; English, S.; Forsythe, M.; Healy, S.; Pavelin, E. Assimilation of satellite data in numerical weather prediction. Part II: Recent years. *Q. J. R. Meteorol. Soc.* **2022**, *148*, 521–556.
14. Rabier, F.; Järvinen, H.; Klinker, E.; Mahfouf, J.F.; Simmons, A. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.* **2000**, *126*, 1143–1170. [\[CrossRef\]](#)
15. de Rosnay, P.; Browne, P.; de Boissésón, E.; Fairbairn, D.; Hirahara, Y.; Ochi, K.; Schepers, D.; Weston, P.; Zuo, H.; Alonso-Balmaseda, M.; et al. Coupled data assimilation at ECMWF: Current status, challenges and future developments. *Q. J. R. Meteorol. Soc.* **2022**, *148*, 2672–2702. [\[CrossRef\]](#)
16. Krasnopolsky, V.M. The application of neural networks in the earth system sciences. *Neural Netw. Emul. Complex Multidimens. Mapp.* **2013**, *46*.
17. Krishnan, P.; Srinivasa Ramanujam, K.; Balaji, C. An artificial neural network based fast radiative transfer model for simulating infrared sounder radiances. *J. Earth Syst. Sci.* **2012**, *121*, 891–901. [\[CrossRef\]](#)
18. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat, F. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Chen, L.; Han, B.; Wang, X.; Zhao, J.; Yang, W.; Yang, Z. Machine learning methods in weather and climate applications: A survey. *Appl. Sci.* **2023**, *13*, 12019. [\[CrossRef\]](#)
20. Kashinath, K.; Mustafa, M.; Albert, A.; Wu, J.L.; Jiang, C.; Esmailzadeh, S.; Azizzadenesheli, K.; Wang, R.; Chattopadhyay, A.; Singh, A.; et al. Physics-informed machine learning: Case studies for weather and climate modelling. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2021**, *379*. [\[CrossRef\]](#)
21. Li, H.; Gu, M.; Zhang, C.; Xie, M.; Yang, T.; Hu, Y. Retrieving atmospheric gas profiles using FY-3E/HIRAS-II infrared hyperspectral data by neural network approach. *Remote Sens.* **2023**, *15*, 2931.
22. Ye, X.; Ren, H.; Nie, J.; Hui, J.; Jiang, C.; Zhu, J.; Fan, W.; Qian, Y.; Liang, Y. Simultaneous estimation of land surface and atmospheric parameters from thermal hyperspectral data using a LSTM-CNN combined deep neural network. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 5508705. [\[CrossRef\]](#)
23. Kong, S.; Bi, L.; Han, W.; Yin, R.; Zhang, H. Hybrid Bayesian-machine learning framework for multi-profile atmospheric retrieval from hyperspectral infrared observations. *Adv. Atmos. Sci.* **2026**, *43*, 373–389.
24. Zhao, Y.; Zhou, D.; Yan, H. An improved retrieval method of atmospheric parameter profiles based on the BP neural network. *Atmos. Res.* **2018**, *213*, 389–397. [\[CrossRef\]](#)
25. Ranjan, P.; Nandal, A.; Agarwal, S.; Kumar, R. A Dive into Generative Adversarial Networks in the World of Hyperspectral Imaging: A Survey of the State of the Art. *Remote Sens.* **2026**, *18*, 196. [\[CrossRef\]](#)
26. Malmgren-Hansen, D.; Laparra, V.; Nielsen, A.A.; Camps-Valls, G. Statistical retrieval of atmospheric profiles with deep convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 231–240. [\[CrossRef\]](#)
27. Lahoz, W.A.; Schneider, P. Data assimilation: Making sense of Earth Observation. *Front. Environ. Sci.* **2014**, *2*, 16. [\[CrossRef\]](#)
28. Gettelman, A.; Geer, A.J.; Forbes, R.M.; Carmichael, G.R.; Feingold, G.; Posselt, D.J.; Stephens, G.L.; Van den Heever, S.C.; Varble, A.C.; Zuidema, P. The future of Earth system prediction: Advances in model-data fusion. *Sci. Adv.* **2022**, *8*, eabn3488. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Wang, X.; Jiang, H. Physics-guided deep learning for skillful wind-wave modeling. *Sci. Adv.* **2024**, *10*, eadr3559. [\[PubMed\]](#)
30. Bocquet, M. Surrogate modeling for the climate sciences dynamics with machine learning and data assimilation. *Front. Appl. Math. Stat.* **2023**, *9*, 1133226. [\[CrossRef\]](#)
31. Kotthaus, S.; Bravo-Aranda, J.A.; Collaud Coen, M.; Guerrero-Rascado, J.L.; Costa, M.J.; Cimini, D.; O'Connor, E.J.; Hervo, M.; Alados-Arboledas, L.; Jiménez-Portaz, M.; et al. Atmospheric boundary layer height from ground-based remote sensing: A review of capabilities and limitations. *Atmos. Meas. Tech.* **2023**, *16*, 433–479. [\[CrossRef\]](#)

32. Xue, Q.; Guan, L.; Shi, X. One-dimensional variational retrieval of temperature and humidity profiles from the FY4A GIIRS. *Adv. Atmos. Sci.* **2022**, *39*, 471–486.
33. Koner, P.K.; Harris, A.R.; Dash, P. A deterministic method for profile retrievals from hyperspectral satellite measurements. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5657–5670.
34. Xiao, J.; Liu, Y.; Wei, X. Region-aware sequence-to-sequence learning for hyperspectral denoising. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 218–235.
35. Gu, A.; Goel, K.; Ré, C. Efficiently modeling long sequences with structured state spaces. *arXiv* **2021**, arXiv:2111.00396.
36. Chen, W.; Zhang, Y.; Liu, R.; Sun, S.; Feng, Q. A Mamba-Based Hierarchical Partitioning Framework for Upper-Level Wind Field Reconstruction. *Aerospace* **2025**, *12*, 842.
37. Feng, J.; Qin, X.; Wu, C.; Zhang, P.; Yang, L.; Shen, X.; Han, W.; Liu, Y. Improving typhoon predictions by assimilating the retrieval of atmospheric temperature profiles from the FengYun-4A's Geostationary Interferometric Infrared Sounder (GIIRS). *Atmos. Res.* **2022**, *280*, 106391.
38. Ye, X.; Wang, P.; Zhu, J.; Duan, Y.; Yang, B. Land Surface Temperature end-to-end retrieval considering the topographic effect using radiative transfer model-driven convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5001010.
39. Gonzalez, J.; Dipu, S.; Sourdeval, O.; Siméon, A.; Camps-Valls, G.; Quaas, J. Emulation of Forward Modeled Top-of-Atmosphere MODIS-Based Spectral Channels Using Machine Learning. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *18*, 1896–1911.
40. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
41. Cheng, S.; Argaud, J.P.; Iooss, B.; Ponçot, A.; Lucor, D. A graph clustering approach to localization for adaptive covariance tuning in data assimilation based on state-observation mapping. *Math. Geosci.* **2021**, *53*, 1751–1780. [[CrossRef](#)] [[PubMed](#)]
42. Brence, J.; Tanevski, J.; Adams, J.; Malina, E.; Džeroski, S. Surrogate models of radiative transfer codes for atmospheric trace gas retrievals from satellite observations. *Mach. Learn.* **2023**, *112*, 1337–1363.
43. Tahseen, T.P.; Mendonça, J.M.; Yip, K.H.; Waldmann, I.P. Enhancing 3D planetary atmosphere simulations with a surrogate radiative transfer model. *Mon. Not. R. Astron. Soc.* **2024**, *535*, 2210–2227. [[CrossRef](#)]
44. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
45. Chang, S.; Sheng, Z.; Du, H.; Ge, W.; Zhang, W. A channel selection method for hyperspectral atmospheric infrared sounders based on layering. *Atmos. Meas. Tech.* **2020**, *13*, 629–644. [[CrossRef](#)]
46. Zhang, L.; Bao, Y.; Liu, H.; Lu, Q.; Wang, Y.; Huang, Y.; Wu, Y. Retrieval and fusion of atmospheric temperature and humidity profiles in the East China based on FY-4B/GIIRS. *J. Meteorol. Sci.* **2026**, *46*, 80–91.
47. Perez, E.; Strub, F.; De Vries, H.; Dumoulin, V.; Courville, A. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
48. Tan, X.; Ma, K.; Dou, F. A convolutional neural network and attention-based retrieval of temperature profile for a satellite hyperspectral microwave sensor. *Atmosphere* **2024**, *15*, 235. [[CrossRef](#)]
49. Xu, X.; Han, W.; Gao, Z.; Li, J.; Yin, R. Retrieval of atmospheric temperature profiles from FY-4A/GIIRS hyperspectral data based on TPE-MLP: Analysis of retrieval accuracy and influencing factors. *Remote Sens.* **2024**, *16*, 1976.
50. Xiao, C.; Dong, J.; Dou, H.; Li, Y.; Wang, W.; Ren, F. A transformer network air temperature and humidity inversion method based on ATMS brightness temperature data. *IEEE Geosci. Remote Sens. Lett.* **2024**, *22*, 4500205. [[CrossRef](#)]
51. Jiang, S.; Ma, Y.; Deng, F.; Lei, L. A deep learning framework for enhanced retrieval of atmospheric temperature and humidity profiles across China: Unifying inversion algorithms across multiple stations. *Atmos. Res.* **2025**, *315*, 107793.

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.