JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2024JH000449

A Machine Learning-Based Observation Operator for FY-4B GIIRS Brightness Temperatures Considering the Uncertainty of Label Data

Yonghui Li[1,2], Wei Han[3,4], Wansuo Duan[1,2], Zeting Li[4,5], and Hao Li[6,7]

[1]State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]State Key Laboratory of Severe Weather Meteorological Science and Technology (LASW), Beijing, China, [4]CMA Earth System Modeling and Prediction Centre (CEMC), China Meteorological Administration, Beijing, China, [5]Key Laboratory of Meteorological Disaster, Ministry of Education (KLME)/Joint International Research Laboratory of Climate and Environment Change (ILCEC)/Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters (CIC-FEMD), Nanjing University of Information Science & Technology, Nanjing, China, [6]Artificial Intelligence Innovation and Incubation Institute, Fudan University, Shanghai, China, [7]Shanghai Academy of Artificial Intelligence for Science (SAIS), Shanghai, China

Abstract The increasing volume of satellite data, particularly hyperspectral infrared data, combined with the real-time monitoring requirements of numerical weather prediction (NWP) systems, has heightened the demand for computational efficiency and accuracy in radiative transfer models (RTM). Machine learning (ML) offers a promising approach, and numerous studies on ML-based RTM have emerged recently. However, existing ML-based RTMs for hyperspectral infrared were not end-to-end. Moreover, since the label data do not represent truth, models trained with loss functions like mean squared error (MSE) or mean absolute error (MAE) fail to account for its uncertainty. This limitation can lead to suboptimal model parameters, as training may assign higher weights to labels with larger errors. This study construct an end-to-end ML-based RTM focused on clear sky conditions over the ocean for the FengYun-4B satellite (FY-4B) Geostationary Interferometric Infrared Sounder (GIIRS), using maximum likelihood estimation (MLE) and MSE for training, respectively. MLE accounts for the uncertainty of labels. The results indicate both models achieve high accuracy, with mean errors within 0.1 K (K) and standard deviation (STD) of errors within 0.04 K compared to the labels. The model trained with MLE exhibits a mean error closer to 0 and a STD similar to the error STD of labels, suggesting better parameter configurations to reflect the actual error distribution of the labels. Additionally, the temperature and water vapor Jacobian computed by both models are comparable to those obtained from RTTOV, highlighting their potential for application as observational operator in satellite data assimilation for hyperspectral infrared sounder.

Plain Language Summary Satellite hyperspectral infrared data accounts for a significant portion of satellite data and is increasing annually. Moreover, numerical weather prediction requires real-time monitoring of satellites, leading to increasingly higher demands for computational efficiency and accuracy in radiative transfer models. This article develops a machine learning-based radiative transfer model specifically for hyperspectral infrared sounder. The model is trained using both maximum likelihood estimation and mean squared error. The former considers the uncertainty of the labels and both achieve high accuracy, with MLE exhibiting even higher accuracy and better parameter configurations. Additionally, by calculating the gradients of brightness temperature with respect to atmospheric profiles, that is, the inputs, it is found that the accuracy is similar to that of the traditional model RTTOV, demonstrating its potential for application as observational operator in satellite data assimilation for hyperspectral infrared sounder.

Key Points:

- An end-to-end machine learning-based radiative transfer model for clear sky has been developed for FY-4B GIIRS
- The model, trained using a likelihood-based loss function that weights each channel differently, captures the uncertainty of labels
- The Jacobian matrix calculated by the constructed model resembles the results from the radiative transfer model RTTOV

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

W. Han and W. Duan,
hanwei@cma.gov.cn;
duanws@lasg.iap.ac.cn

Citation:

Li, Y., Han, W., Duan, W., Li, Z., & Li, H. (2025). A machine learning-based observation operator for FY-4B GIIRS brightness temperatures considering the uncertainty of label data. *Journal of Geophysical Research: Machine Learning and Computation*, *2*, e2024JH000449. https://doi.org/10.1029/2024JH000449

Received 26 SEP 2024
Accepted 27 FEB 2025

Author Contributions:

Conceptualization: Yonghui Li, Wei Han, Wansuo Duan, Zeting Li, Hao Li
Data curation: Yonghui Li, Zeting Li, Hao Li
Formal analysis: Yonghui Li, Wei Han
Funding acquisition: Wei Han, Hao Li
Investigation: Yonghui Li, Wei Han, Zeting Li
Methodology: Yonghui Li, Wei Han, Wansuo Duan, Zeting Li
Project administration: Wei Han, Hao Li

1. Introduction

Satellite data assimilation plays a crucial role in enhancing forecasting skills of NWP systems, especially assimilating hyperspectral infrared observations which can greatly improve the initial conditions of atmospheric and enhance the accuracy of numerical weather predictions (Cardinali & Prates, 2009; Eresmaa et al., 2017; A. Geer et al., 2019; A. J. Geer et al., 2019; Li, Han, & Duan, 2024; McNally et al., 2014; Yin et al., 2021). On 10 December 2016, China successfully launched the FengYun-4A satellite (FY-4A) into a geostationary orbit. The

10.1029/2024JH000449

Resources: Wei Han, Wansuo Duan, Hao Li
Software: Yonghui Li, Zeting Li, Hao Li
Supervision: Wei Han, Wansuo Duan, Zeting Li, Hao Li
Validation: Yonghui Li, Wei Han, Wansuo Duan, Zeting Li, Hao Li
Visualization: Yonghui Li, Wansuo Duan, Zeting Li
Writing – original draft: Yonghui Li, Wei Han, Wansuo Duan, Zeting Li
Writing – review & editing: Yonghui Li, Wei Han, Wansuo Duan

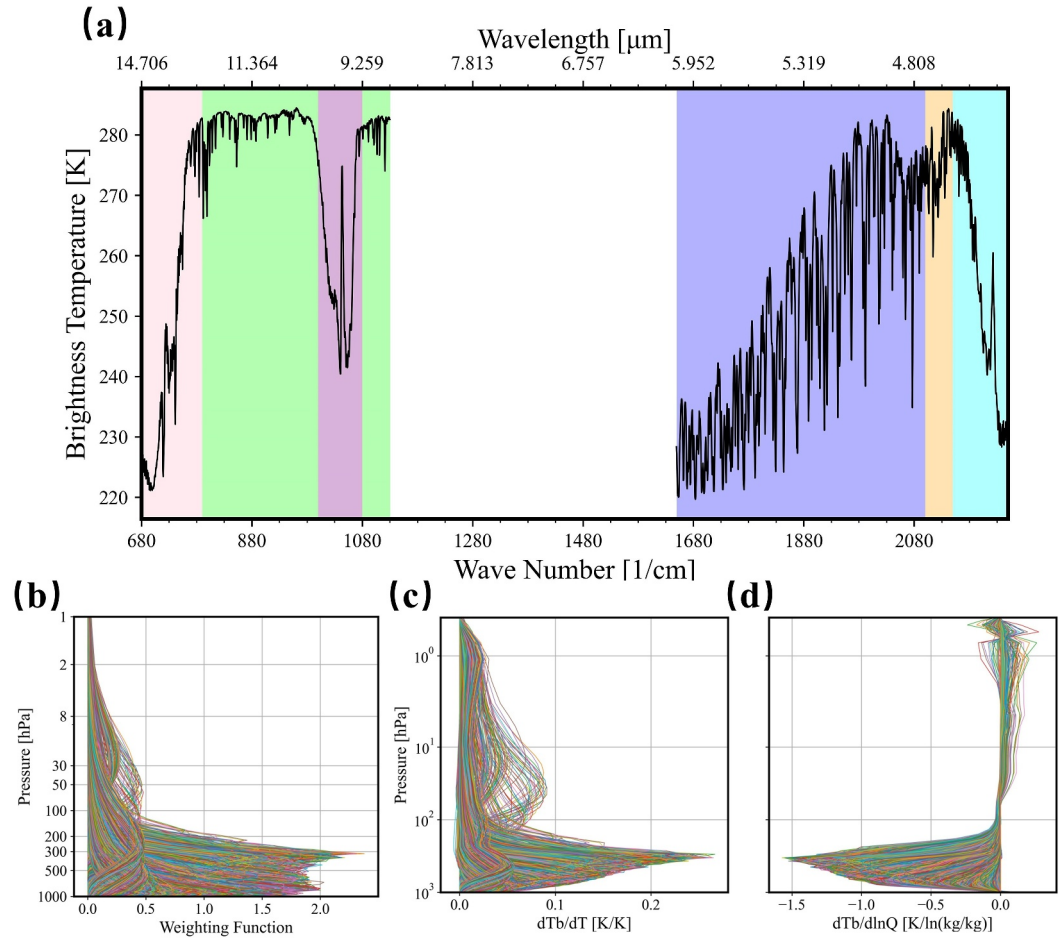

Figure 1. Characteristics of FY-4B GIIRS data. Panel (a) shows the brightness temperature of various channels of FY-4B GIIRS. The pink shaded area represents temperature channels, the green represents window channels, the purple represents ozone channels, the blue represents water vapor channels, the orange represents carbon monoxide channels, and the cyan represents nitrogen dioxide channels. Panel (b) displays the weight functions (WF) for each channel, panel (c) shows the temperature Jacobian for each channel, and panel (d) presents the water vapor Jacobian.

satellite carries the GIIRS, which represents a significant advancement in Earth observation systems (Yang et al., 2017). Following this success, the FY-4B was launched on 3 June 2021, also equipped with GIIRS, making it the second Geostationary Hyperspectral Infrared Sounder in orbit internationally, further contributing to the enhancement of Earth observation capabilities, and the spectral range of FY-4B GIIRS spans from 680 to 1,130 cm^{-1} (8.85 to 14.71 μm) and 1,650 to 2,250 cm^{-1} (4.44 to 6.06 μm) with the same spectral interval of 0.625 cm^{-1} (Figure 1a). This instrument have thousands of channels with and extremely high vertical spatial resolution, allowing precise detection of the vertical structure of atmospheric temperature and humidity (Figures 1a–1d) and provides large-scale, continuous, fast, and precise data, which could capture the temporal and spatial variability for high-impact weather or rapidly changing weather events, significantly enhancing weather forecasting skills (Feng et al., 2022; Han et al., 2023; Li, Han, & Duan, 2024; Niu et al., 2023; Yin et al., 2021, 2022). To efficiently and accurately assimilate the FY-4B GIIRS data, an efficient and precise RTM, serving as an observational operator within satellite data assimilation, is indispensable. It acts as a bridge between the observation space and the model space. The hyperspectral infrared sounder like the FY-4B GIIRS, with its high spatial, spectral, and temporal resolution, demands significant computational resources for data assimilation.

To improve computational efficiency, extensive research has been conducted by scholars, which can be roughly divided into three categories. The first category focuses on relaxing the accuracy constraints for each individual simulation, such as two streams (Meador & Weaver, 1980), low orders of scatterings (Natraj & Spurr, 2007), or precomputed look-up tables (Wang et al., 2011). The widely used radiative transfer models, such as the Community

Radiative Transfer Model (CRTM) (X.-M. Liang et al., 2009), the Radiative Transfer for TOVS (RTTOV) (Saunders et al., 1999), and the recently developed Advanced Radiative Transfer Modeling System (ARMS) (Weng et al., 2020), all rely on the aforementioned techniques. The second category aims to minimize computational costs by reducing the number of individual radiative transfer simulations within a spectral band while maintaining calculation accuracy, such as the correlated k distribution (CKD) method (Arking & Grossman, 1972; Fu & Liou, 1992; Kratz, 1995; Lacis & Oinas, 1991; C. Liu et al., 2015; Shi et al., 2009), the principal component method (X. Liu et al., 2006, 2009; C. Liu et al., 2020), and the optimal spectral sampling method (Moncet et al., 2008). The third category leverages the rapid development of machine learning (ML) within the context of big data. ML offers a promising avenue for training an RTM to capture complex relationships between atmospheric parameters and radiative variables, thereby reducing computational resources. For instance, Chevallier et al. (1998) used an multi layer perceptron (MLP) to simulate longwave radiation. Le et al. (2020) introduced a model based on the MLP to extend a small fraction of monochromatic radiances to the entire spectral range. However, its inputs differ from traditional RTM models. Instead of atmospheric variables, it uses radiation from specific wavelengths. Su et al. (2023) introduced a fast and accurate RTM for the FY-4B GIIRS based on principal component analysis or MLP, but the model is not end-to-end. Instead of directly predicting from atmospheric variables to radiation, it uses the standard Schwarzschild Equation to simulate brightness temperature (BT) in intermediate steps. There are also studies based on ML that target microwaves instead of infrared (X. Liang et al., 2022), or utilize ML to replace components within the RTM (Stegmann et al., 2022; Ukkonen, 2022).

Although ML models for infrared instruments have been developed as mentioned above, they are either not specifically designed for hyperspectral data from geostationary satellites or are not end-to-end ML models. Additionally, the loss functions utilized in the aforementioned ML-based models is typically either mean squared error (MSE) or mean absolute error (MAE). This implies that their learning objective is focused solely on the model reference labels. However, the labels (either the actual observations or the BT simulated by RTM) are inaccurate, with errors existing between the labels and the true value. Using MSE or MAE as the loss function does not account for the magnitude of errors of labels itself between different channels, assigning equal weights to each channel which is unreasonable. It is conceivable that if there is a large error of label in a certain channel, we should give that channel a smaller weight during model training in order to ensure the reliability of the model.

This paper introduces an end-to-end hyperspectral RTM based on ML for the FY-4B GIIRS, specifically for clear-sky conditions over the ocean. The model features an encoder-decoder architecture where a Bidirectional Long Short-Term Memory (BiLSTM) (Zhang et al., 2015) network functions as the encoder and the MLP serves as the decoder. The MLP takes the encoded information from the BiLSTM and transforms it to produce the final output, which predicts BT. The model proposed in this paper is named BH-RTM. To overcome the issue of MSE treating all channels with equal weights without considering uncertainty in different channels, this paper incorporates a strategy akin to data assimilation (Bouttier & Courtier, 2002). We estimate the error distribution between the labels and the true data, and then assign different weights to different channels through the error covariance matrix, that is, the uncertainty. This loss function is derived from the likelihood function of Maximum Likelihood Estimation (MLE), which involves finding the optimal parameters that maximize the probability of the labels, providing more robust and accurate parameter estimation. This paper used MSE and likelihood function of MLE as the loss functions to train the BH-RTM, respectively, and both had high accuracy. Furthermore, the Jacobian computed from both models were validated and found to be comparable to those from RTTOV, indicating their potential as alternative models for traditional RTM in data assimilation.

The organization of this paper is structured as follows: Section 2 introduces the methodology employed. Section 3 provides an overview of the data used. Section 4 presents the key findings and results, and Section 5 encompasses a comprehensive discussion, conclusions drawn from the study, as well as future prospects and directions.

2. Method

2.1. Model

The satellite radiative transfer equation typically consists of three components (Weng & Liu, 2003). The first component describes the radiation intensity transmitted through the atmosphere after attenuation due to absorption by the medium, following reflection from the Earth's surface. The second component involves integrating the thermal emission and single scattering by the medium along the upward path. The third component considers multiple scattering. This equation indicates that radiation is accumulated layer by layer before reaching

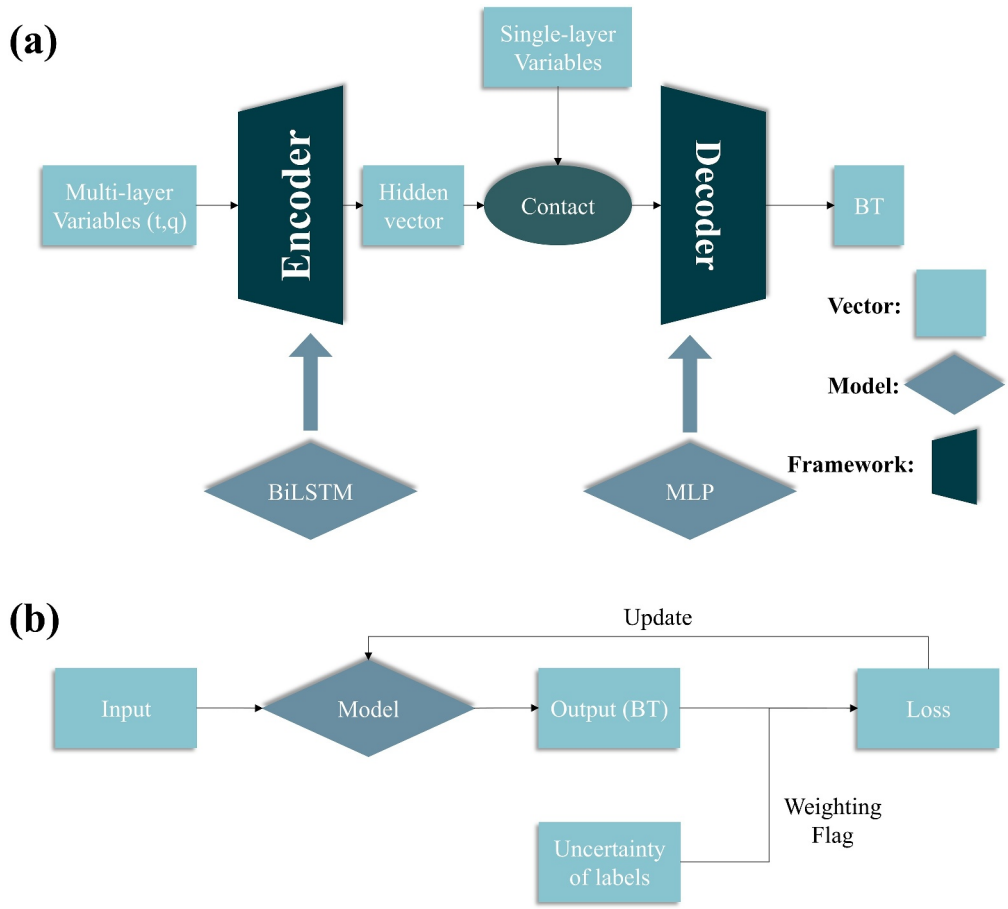

Figure 2. Panel (a) shows the model framework, where the multi-layer variables represent the atmospheric temperature and humidity profiles, and the single-layer variables are listed in Table 1. Panel (b) illustrates the training process. When the weighting flag is set to "True," the loss function incorporates label uncertainty, treating it as a weight. When the weighting flag is set to "False," the loss function does not account for label uncertainty.

the satellite probe, being attenuated or enhanced through interactions with objects within the atmospheric medium as it propagates.

As mentioned above, most ML-based on RTM utilizes MLP. MLP takes multi-layer variables and single-layer variables as inputs and learns relationships between different atmospheric layers in the hidden layers to predict BT. However, it does not explicitly connect the layers together, which means it may can not capture the flow of information between different layers. Even though MLP shows promising results, we may not fully understand this phenomenon. BiLSTM is a type of recurrent neural network particularly adept at handling time-series data. Its layer-by-layer processing mechanism bears similarity to radiative transfer models. The input data, representing atmospheric profiles at different altitude layers, are fed as time steps into BiLSTM.

The BH-RTM designed in this paper aims to simulate the atmospheric radiative transfer process and predict BT of FY-4B GIIRS. The specific framework of BH-RTM is illustrated in Figure 2a. BiLSTM, serving as encoder, processes the multi-layer variables through multiple hidden layers bidirectionally, considering information flow both from the surface to the top layer (forward LSTM) and from the top layer to the surface (backward LSTM). The final hidden state of both the forward and backward LSTMs are concatenated to form a hidden vector. The hidden vector is then combined with the single-layer variables to create a input, integrating atmospheric profiles and single-layer information to provide a comprehensive set of features for the decoder. The decoder is composed of MLP with ReLU activation function, which helps prevent overfitting and introduces non-linearity into BH-RTM. This allows BH-RTM to capture intricate patterns and interactions within the data. This layer-by-layer processing of BiLSTM and information integration design not only captures the complex relationships among

the various altitude layers of the atmospheric profiles but also incorporates the single-layer variable information. Such a design has the potential to improve the accuracy of BH-RTM predictions and provide physical interpretability for BH-RTM. In the BiLSTM, the hyperparameter hidden size is set to 128, and num layers is set to 5. For detailed information on the input data of the model, please refer to Section 3.

2.2. Loss Function

This paper uses two types of loss functions to train BH-RTM. The first is the widely used MSE:

$$J_{MSE}(\theta) = \frac{1}{N}(y - \hat{y}(\theta,x))^T (y - \hat{y}(\theta,x)) \tag{1}$$

where y represents the label (either the actual FY-4B GIIRS observations or the BT of FY-4B GIIRS simulated by RTTOV or LBL) and $\hat{y}(\theta,x)$ denotes the prediction from BH-RTM, with θ being model parameters. The labels used in this paper are detailed in Section 3. The BH-RTM that utilizes MSE as the loss function is subsequently referred to as BH-RTM-S.

However, neither actual FY-4B GIIRS observations nor FY-4B GIIRS BT simulated by RTTOV or LBL represent the true BT. Therefore, the predicted results of trained BH-RTM are only close to the BT with errors. Assuming we know the error distribution between labels and the true BT.

$$y_t = y + \epsilon \tag{2}$$

where y is label, y_t is true BT, and ϵ represents the error. We assume that error ϵ follows a Gaussian distribution with mean μ and covariance matrix Σ, where:

$$\epsilon \sim \mathcal{N}(\mu, \Sigma) \tag{3}$$

Here, $\Sigma \in \mathbb{R}^{N*N}$ with $N = 1682$ representing the total number of FY-4B GIIRS channels and we assume $\mu = 0$ where $\mu \in \mathbb{R}^N$. In Section 3, we will discuss the methodology for estimating the standard deviation (STD) of this error distribution to quantify the uncertainty of labels.

Compared to MSE, using likelihood function of MLE as loss function aims to bring the prediction $\hat{y}(\theta,x)$ to estimate true BT y_t. The likelihood function can be written as:

$$L(\theta;y,x) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left(-\frac{(y - \hat{y}(\theta,x))^T \Sigma^{-1} (y - \hat{y}(\theta,x))}{2}\right) \tag{4}$$

By maximizing the likelihood function, we obtain the loss function

$$J_{MLE}(\theta) = (y - \hat{y}(\theta,x))^T \Sigma^{-1} (y - \hat{y}(\theta,x)) \tag{5}$$

While both Equations 1 and 5 share similarities, MLE enables the model to not only focus on the direct difference between predicted and labels, but also to comprehensively assess the uncertainty of the labels through assigning different weights. This, in turn, enhances the overall reliability of the BH-RTM. Conversely, MSE primarily strives to align the prediction of model with the labels. The BH-RTM that uses likelihood function as the loss function is subsequently referred to as BH-RTM-L. Figure 3 shows the uncertainty Σ. Analysis found that the STD is higher in the absorption bands of water vapor and ozone, indicating significant forecast uncertainty of RTTOV in these regions. In contrast, the uncertainty is lower in the atmospheric window regions (green shade). Through this weighting scheme, the BH-RTM-L will place greater emphasis on channels with smaller uncertainty during training, as these channels will contribute more to the loss function. Conversely, the BH-RTM-L will pay relatively less attention to channels with larger uncertainty. Considering the subsequent comparison with BH-RTM-S, the Σ is normalized such that the diagonal elements sum to one during training. As for how Σ is computed, please refer to the appendix Appendix A.

10.1029/2024JH000449

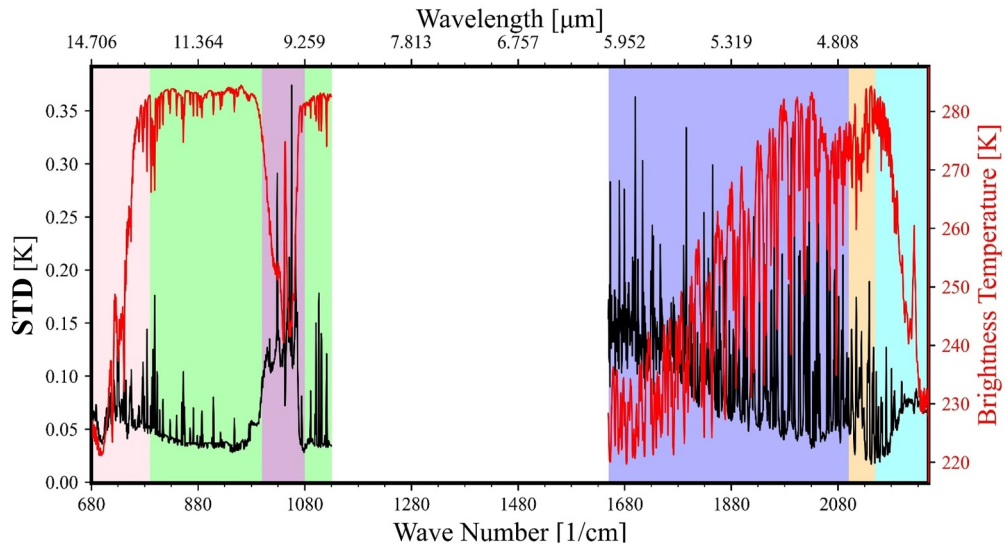

Figure 3. The STD, used to quantify uncertainty, of the statistics. The black line represents the square root of the diagonal elements of the Σ matrix, the red line represents the BT, and the shaded areas are consistent with those in Figure 1a.

3. Data

The data set used to train the BH-RTM includes input features and labels. The input features are divided into two categories: single-layer variables and multi-layer variables (details in Table 1). The single-layer variables include 2 m dew point temperature, surface temperature, 10 m wind speed, surface atmospheric pressure, satellite azimuth angle, satellite zenith angle, solar azimuth angle, and solar zenith angle. The first four variables are from ECMWF Reanalysis v5 (ERA5) (Hersbach et al., 2020), while the last four variables are from FY-4B GIIRS L1 data (https://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx). The multi-layer variables are the temperature and specific humidity atmospheric profiles data, which have 37 layers from ERA5. Some studies (Graham et al., 2019; Hersbach et al., 2020; Tetzner et al., 2019) have evaluated the ERA5 data set and reported that it shows good performance in variables such as the wind field, temperature and humidity, and 2-m temperature. It is important to note that both single-layer and multi-layer variables are interpolated to match the time, latitude and longitude of the actual FY-4B GIIRS L1 data. This interpolated data set is used for training, validating, testing.

Table 1
The Details of Data Set

Input features			Labels		
Names	Number	Data source	Names	Number	Data source
Air temperature	37[a]	ERA5	BT of FY-4B GIIRS (Ch1-Ch1682)	1,682	Simulation of FY-4B GIIRS by RTTOV
Air specific humidity	37	ERA5			
Surface temperature	1	ERA5			
Two meter dew point temperature	1	ERA5			
10 m wind speed	1	ERA5			
Surface air pressure	1	ERA5			
Earth azimuth	1	L1 data of FY-4B GIIRS observation			
Earth zenith	1	L1 data of FY-4B GIIRS observation			
Solar azimuth	1	L1 data of FY-4B GIIRS observation			
Solar zenith	1	L1 data of FY-4B GIIRS observation			

[a]1, 2, 3, 5, 7, 10, 20, 30, 50, 70, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 775, 800,825, 850, 875, 900, 925, 950, 975, 1,000 hPa.

The labels are FY-4B GIIRS' BT simulated by RTTOV version 12.3, based on the input features along with the FY-4B GIIRS scan geometry. Using RTTOV-simulated labels helps estimate the calculation accuracy of the Jacobian in BH-RTM. Considering that RTTOV simulations are inaccurate under cloud-sky conditions, we focus only on BT of FY-4B GIIRS for clear-sky conditions. Additionally, we only focused on the ocean domain to avoid the complexity of surface emissivity.

Considering the existing FY-4B GIIRS data, the time range for the all data set is from 31 May 2023, to 31 January 2024. The corresponding ERA5 data is hourly, but only the 1st, 7th, 14th, 21st, and 28th of each month are considered to capture the variations within a month for FY-4B GIIRS. The data set is adequately shuffled and roughly separated into training, testing, and validation data sets in the ratio of 8:1:1, wherein the number of training data samples is approximately three million. Since the training, testing, and validation data set are shuffled and redistributed from the original data set, it is not possible to extract actual prediction results for a specific day from the testing data set, making it difficult to effectively demonstrate the generalization results. Therefore, 2 May 2024, a specific day outside the original data set, is selected to test the generalization ability of BH-RTM. Since this classification results in minimal differences between the profiles in the training and test sets, we further selected data from 1 March to 5 March 2024, as an additional data set to validate the model's generalization capability.

Before training, the data set undergoes preprocessing. By calculating the maximum and minimum values for each feature and label in the training data set and applying min-max normalization to all features and labels, consistency and stability of data during the training process are ensured. This preprocessing step accelerates the training process, improves prediction performance of model, avoids numerical computation issues, and enhances the stability and robustness of model, providing a solid foundation for subsequent modeling and analysis. During the training process, the batch size was set to 1,024 with model trained for 200 epochs, and the optimizer used was stochastic gradient descent (SGD). The training process is shown in Figure 2b. When the model is trained using the likelihood function, the weighting flag is set to 'True'; when the model is trained using MSE, the weighting flag is set to 'False'.

4. Results

4.1. Model Evaluation

With consistent model configuration, training is conducted using MSE and likelihood function of MLE as the loss functions, respectively. The convergence process is shown in Figure 4. During training, the STD (Figure 4a) of the model predictions and the labels, as well as the loss (Figure 4b), gradually decreased, indicating that the fitting ability of model progressively improved. This trend appeared in both the training data set and the validation data set, suggesting no overfitting occurred during the training process and both BH-RTM-L and BH-RTM-S have generalization ability. After 200 epochs of training, it can be observed that the loss has converged. It is noteworthy that loss value of BH-RTM-L is much higher than loss value of BH-RTM-S (Figure 4b). This is understandable, as for BH-RTM-L, when the STD is small (i.e., the weight is large) and the difference between the model prediction and the label is large, the loss value increases significantly. On the other hand, for BH-RTM-S, each error of channel is equally weighted, so even if the difference between the model prediction and the label is large for some channels, it does not significantly affect the overall loss. This difference in weighting results in the BH-RTM-L having a higher overall loss value, especially when dealing with channels that have large weights and where the prediction of model significantly deviates from the labels.

When evaluating model, it is crucial to not only the immediate fit between predicted BT and labels but also the generalization ability of model. The evaluation results on the testing data set are as follows. Both BH-RTM-L and BH-RTM-S demonstrate good simulation capabilities (Figures 5a and 5b), with the maximum difference between the model prediction and the label being about −0.09 K. The bias of BH-RTM-L in most channels is closer to 0 compared to the BH-RTM-S, which is especially notable in the water vapor channels (the blue shade area of Figure 5a). Figure 5b depicts the error STD. The STD for both models is also small, generally below 0.04 K. Notably, the error STD of BH-RTM-L is more similar the actual STD (the square root of Σ_{RML}), especially in complex absorption channels like water vapor and ozone (the blue and purple shade area of Figure 5b), outperforming the BH-RTM-S. And BH-RTM-L has a smaller STD in the window channels. This may be attributed

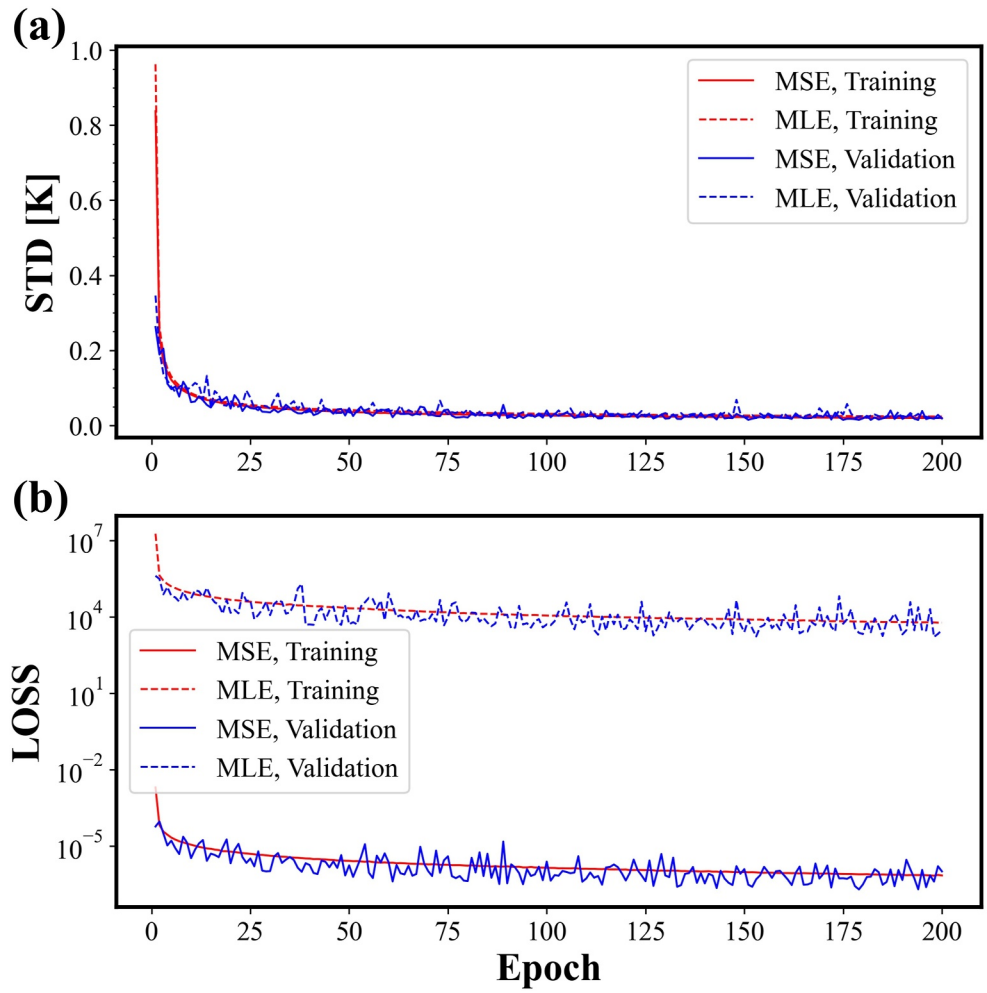

Figure 4. Model convergence process. Panel (a) shows the STD between the predictions and the labels during training, and panel (b) shows the loss reduction process. In the figures, the solid blue line represents the result of predicted by BH-RTM-S on validation set, the dashed blue line represents the result of predicted by BH-RTM-L on validation set, the solid red line represents the result of predicted by BH-RTM-S on training set, and the dashed red line represents the result of predicted by BH-RTM-L on training set.

to the larger weights assigned to the window channels during training, resulting in lower STD in these channels. This similarity suggests that the BH-RTM-L is better at capturing the true uncertainty, despite having a higher overall loss during training. In summary, while both models show similar trends in error mean and STD, the BH-RTM-L provides a more accurate reflection of the true error distribution, particularly in key absorption channels, making it potentially more reliable for practical applications. To avoid losing the uniqueness of individual profile in statistical averages, we also conducted tests on single profiles (Figure S1 in Supporting Information S1), which showed results similar to the averaged outcomes. Additionally, to further explore the generalization capability of the model, we evaluated the data set from March 1 to 5 March 2024 (Figure S2 in Supporting Information S1). These results were consistent with the aforementioned findings, remaining within an acceptable range.

To better demonstrate the generalization ability of model, we compared the predicted BT with the BT of clear-sky simulated by RTTOV on 2 March 2024 (the date not included in the original data set). Figure 6 illustrates the predicted BT and the corresponding errors for data on 2 March 2024, focusing on the channel 100 which is middle-layer temperature channel. Both BH-RTM-S and BH-RTM-L (Figures 6a and 6b) demonstrate a close match with the RTTOV results (Figure 6c). It is worth nothing that the predictions of BH-RTM-S have a mean error of −0.0335 K and a STD of 0.0508 K when compared to RTTOV (Figure 6e), indicating a low and

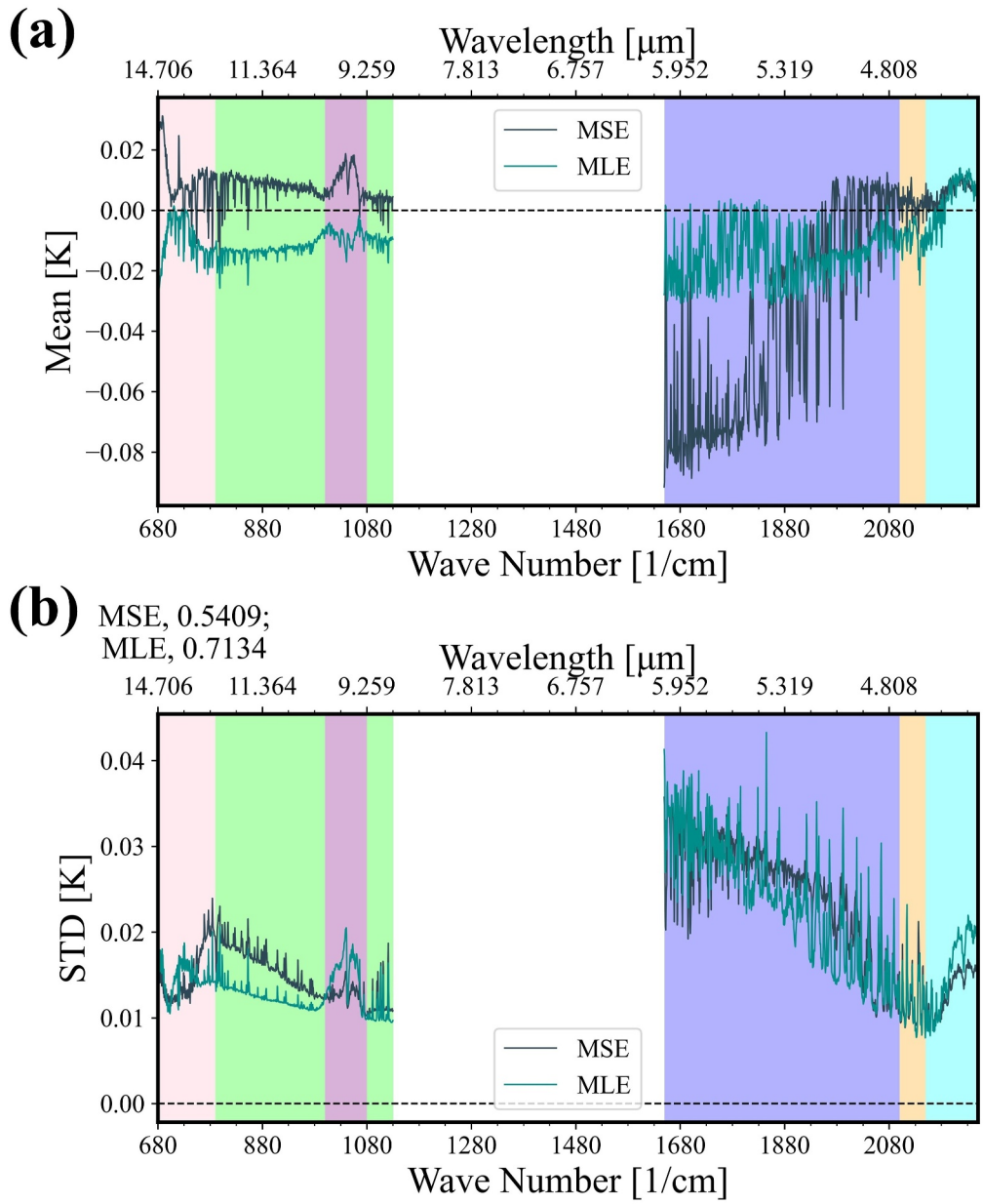

Figure 5. Evaluation results of the model. Panel (a) shows the mean error between the prediction of model and the labels, with results statistically obtained on the testing data set. The black solid line represents the mean error of BH-RTM-S, and the green solid line represents the mean error of BH-RTM-L. Panel (b) shows the error STD between the prediction of BH-RTM and the labels, also statistically obtained on the testing data set. The black solid line represents the error STD between prediction of BH-RTM-S and labels, and the green solid line represents the error STD prediction of between BH-RTM-L and labels. In the upper left corner of panel (b), the correlation between the error STD of BH-RTM-S, BH-RTM-L and the Σ_{RML} (Figure 3) is shown, respectively. The shaded areas have the same meaning as in Figure 1a.

concentrated error distribution. The BH-RTM-L (Figure 6f) shows a mean error of 0.0020 K and a STD of 0.0844 K, reflecting a better ability to reduce systematic bias. The comparison between BH-RTM-S and BH-RTM-L predictions (Figure 6d) reveals a mean error of −0.0355 K and a STD of 0.0769 K, demonstrating that both models are quite similar, but the BH-RTM-L model tends to produce predictions with slightly higher variability. In other channels with different wavelengths (there are not shown), most channels have similar outcomes. The BH-RTM-S tends to have a systematic bias that is slightly larger than that of the BH-RTM-L.

10.1029/2024JH000449

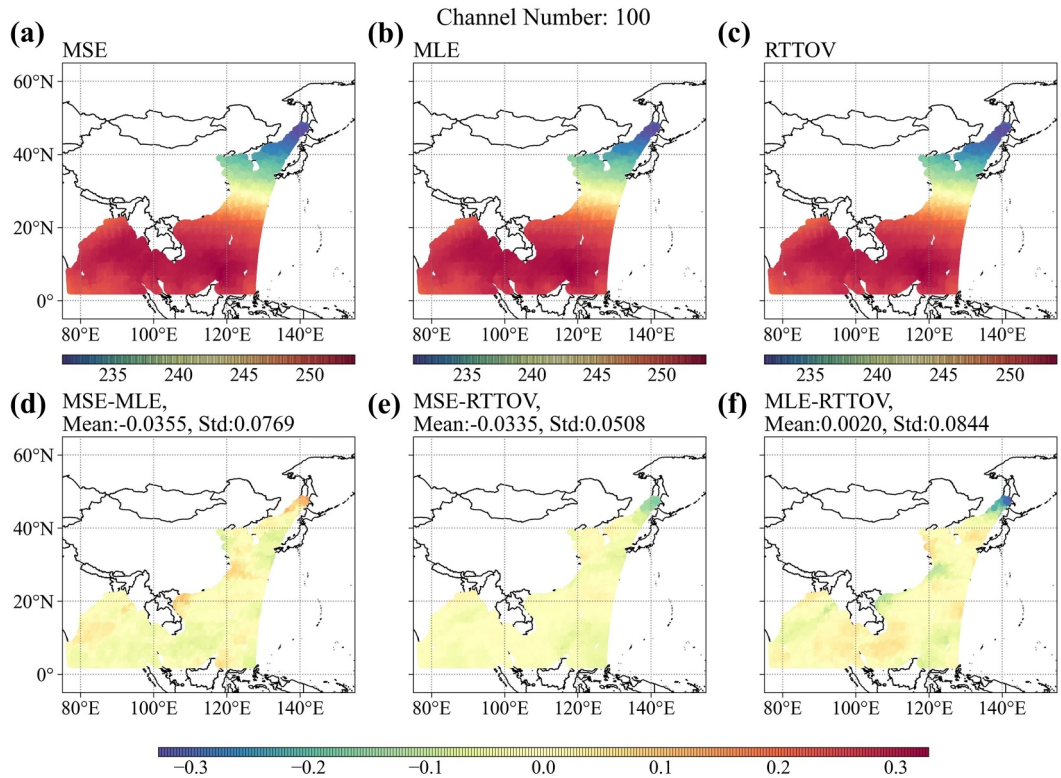

Figure 6. Clear-sky BT of channel 100 on 2 March 2024. Panel (a) shows BT simulated by BH-RTM-S, panel (b) shows BT simulated by BH-RTM-L and panel (c) shows BT simulated by RTTOV. Panel (d) shows BT of BH-RTM-S minus BH-RTM-L, panel (e) shows BT of BH-RTM-S minus RTTOV and panel (f) shows BT of BH-RTM-L minus RTTOV.

4.2. Jacobian

The Jacobian matrix of the RTM model plays a crucial role in data assimilation. This matrix helps us understand the impact of variables in each layer on simulated BT, aiding our comprehension of how assimilating BT can improve initial fields. In essence, the Jacobian matrix serves as a bridge between observational data and model state in data assimilation. By quantifying the response of the observation data to model states, it facilitates more accurate estimation of atmospheric or Earth system states, thereby advancing meteorological forecasting. Therefore, BH-RTM also needs to accurately simulate the Jacobian.

The calculation of the Jacobian is performed on the testing data set. For detailed analysis, six channels are chosen: temperature channel 100, window channel 200, ozone channel 592, water vapor channel 1,352, carbon monoxide channel 1,466, and nitrogen dioxide channel 1,600. The BH-RTM-S (Figures 7a–7f) demonstrates high similarity to RTTOV across different spectral channels. The mean and STD of the temperature and water vapor Jacobian of BH-RTM-S largely align with the RTTOV, indicating that the model can effectively capture the sensitivity of BT to atmospheric variables. For temperature Channel (Ch 100) (Figure 7a), the mean of temperature Jacobian is consistent with the RTTOV results, with highly overlapping STD. In the middle to lower layers, the water vapor Jacobian mean aligns well with RTTOV, but some differences are noted in the upper layers. Similar results have also been drawn in other channels (Figures 7b–7f). It can be observed that the layers with large values of temperature Jacobian in all selected channels are located in the middle layer, and there are significant uncertainty of temperature in these layers (the gray shaded areas in the left plot of each subplot in Figure 7). This indicates that temperature in the middle layer has a considerable impact on the final BT simulation. Additionally, the temperature Jacobian calculated by the BT-RTM-S align relatively well with RTTOV, indirectly explaining why the simulations produced by BT-RTM-S are similar to RTTOV model. For the water vapor Jacobian, although all selected channels exhibit relatively bias in the upper layers compared to RTTOV, but the uncertainty in water vapor content at these upper levels is actually smaller than other layers (the gray shaded areas in the right plot of

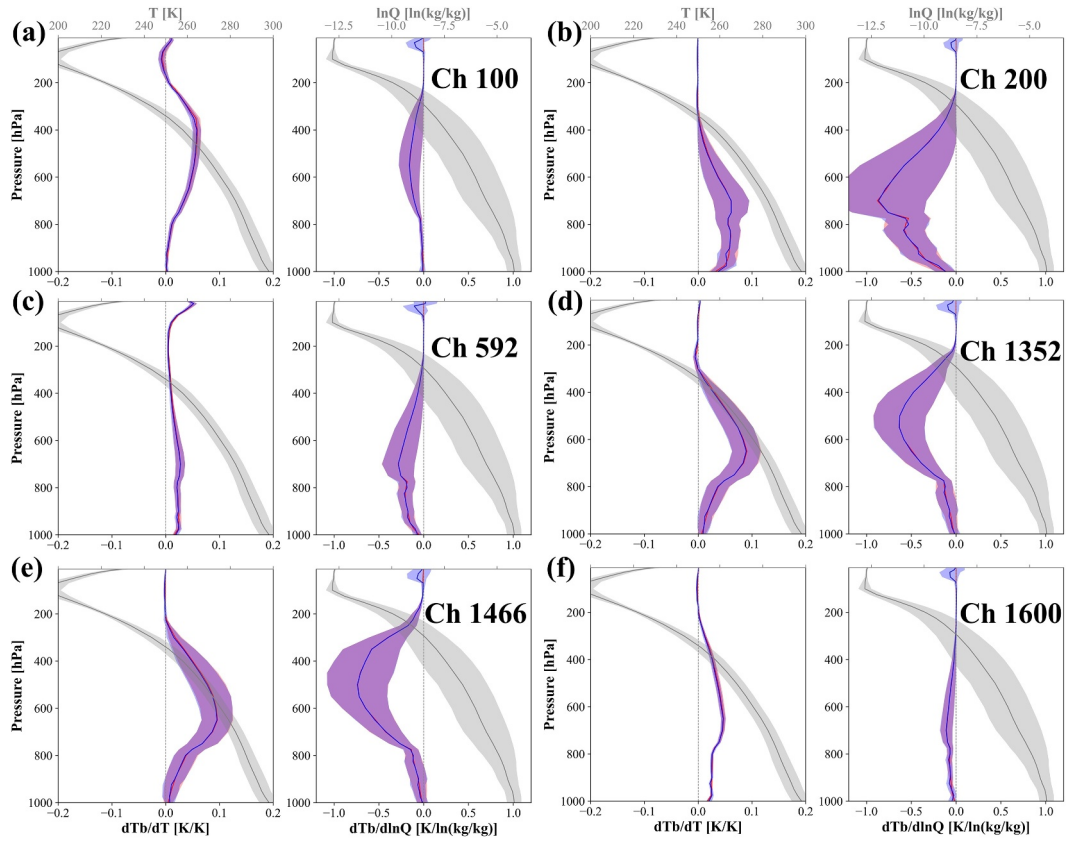

Figure 7. Temperature Jacobian and water vapor Jacobian calculated by BH-RTM-S. Each subplot, Panels (a)–(f), contains two panels: the left panel shows the temperature Jacobian, while the right panel displays the water vapor Jacobian. The solid gray line represents the average temperature or water vapor profile, derived from the testing data set, and the gray shading indicates the standard deviation (STD) of temperature or water vapor, also calculated from the testing data set. The blue solid line represents the average temperature or water vapor Jacobian calculated by BH-RTM-S, with the blue shading showing the corresponding STD. The red solid line and red dashed line represent the average and STD of the temperature and water vapor Jacobian calculated by RTTOV, respectively. Panel (a) through (f) correspond to Channel 100, 200, 592, 1,352, 1,466, and 1,600, respectively.

each subplot in Figure 7). This suggests that this layer has a lesser impact on the final simulated brightness temperatures.

To more clearly illustrate the characteristics of the operator for hyperspectral instruments, the temperature and water vapor Jacobians for all 1,682 channels in GIIRS are shown. Figures 8a and 8d are calculated using RTTOV, Figures 8b and 8e are simulated by BH-RTM-S, and Figures 8c and 8f show the differences between the two. For the water vapor Jacobian (Figure 8c), the error is mainly observed in the upper atmosphere of water vapor channels, with values ranging from −0.04 to 0.08 $K/\ln(\mathrm{kg}/\mathrm{kg})$. For the temperature Jacobian (Figure 8f), the error is primarily around 200 hPa, with values between −0.003 and 0.003 K/K. Compared to the magnitude of the Jacobians themselves, the error remains within about 5%.

The BH-RTM-L (Figures 9a–9f) shows larger difference with RTTOV compared to the difference between BH-RTM-S and RTTOV. A similar conclusion is found across all channels (Figure 10). However, the water vapor Jacobian (Figure 10c) is more pronounced at higher levels, while the temperature Jacobian error (Figure 10f) is more evident at middle and lower levels. Although the water vapor Jacobian error is relatively large compared to its magnitude, it remains within about 10%. The possible reason is that the goal of BH-RTM-L is to estimate the true BT, rather than labels (RTTOV), and optimize model parameters by maximizing the probability of labels. This method has advantages in dealing with complex error structures, but it also introduces some complexity and uncertainty, especially in the case of ideal assumptions for the error distribution, which may lead to lower similarity to labels (RTTOV) in Jacobian calculations. Overall, both BH-RTM-S and BH-RTM-L capture the

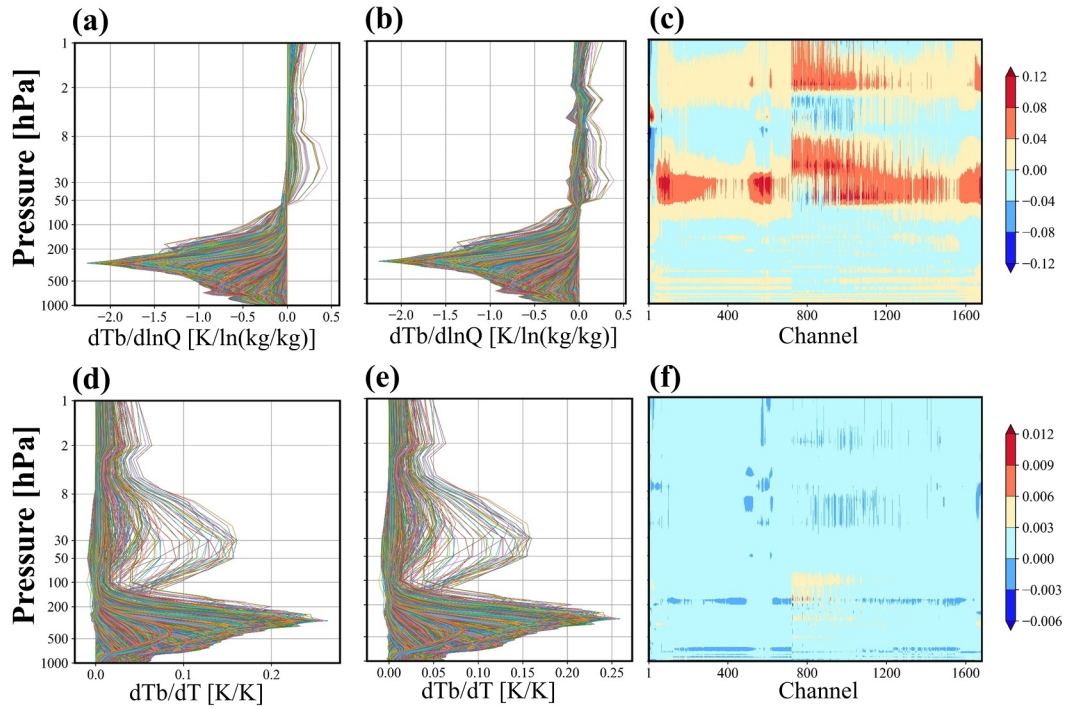

Figure 8. Temperature Jacobian (d)–(f) and water vapor Jacobian (a)–(c) of 1,682 channels calculated by BH-RTM-S. Panels (a) and (d) are calculated by RTTOV. Panels (b) and (e) are calculated by BH-RTM-S. Panels (c) and (f) are difference between RTTOV and BH-RTM-S. The profile data set used in the calculation is from the testing data set.

sensitivity of simulated BT to atmospheric variables. The accuracy of the model still needs to be validated on individual profile to demonstrate the practical utility of the model. We select a profile from 1 March 2024, and compute the Jacobian (Figure 11). Here, we present the results for the temperature channels 81 and 110, the window channels 200, and the water vapor channels 1,104 and 1,352. As seen in the results, for the temperature Jacobian, the two ML models' Jacobian calculations at mid-levels are close to RTTOV, while there are slight fluctuations at higher and lower levels. For the water vapor Jacobian, both models provide relatively accurate simulations at mid to upper levels, though slight fluctuations are still observed at the lower levels. It shows that the temperature channels, water vapor channels and window channels are able to capture the sensitivity of brightness temperature to the atmosphere at some extent, providing a promising approach for ML-based assimilation of satellite observations.

4.3. Physical Mechanism

As mentioned above, each time step of the BiLSTM corresponds to a vertical layer in the profile. By capturing complex relationships in both upward and downward directions, it predicts BT. This approach is similar to RTM.

We use the BH-RTM-S for interpretability analysis. For the BiLSTM model, the forward and backward LSTM each output a hidden state sequence (hereafter referred to as $LSTM_F$ and $LSTM_B$), representing the information flow from the surface to the top layer and from the top layer to the surface, respectively. For each layer, there exists a hidden vector $H_n, n = 1, 2, \ldots, 37$. As shown in Figure 12, to study how the model captures the relationships between layers, the hidden vectors $H_n, n = 1, 2, \ldots, 37$ are saved. Then each hidden vector H_n is decoded to obtain the final output. For each channel, we can obtain a vector with a length of 37, where the last value represents the BT of this channel.

We have selected 6 channels, namely channels 22, 96, 127, 726, 1,109, and 1,402, which correspond to high, middle, and low-layer temperature channels, as well as high, middle, and low-layer water vapor channels, respectively. As shown in Figures 13a–13f, the 37-layer output of each channel has a good correspondence with the forward and reverse weighting function (WF), which represents the radiative contribution of the atmosphere at

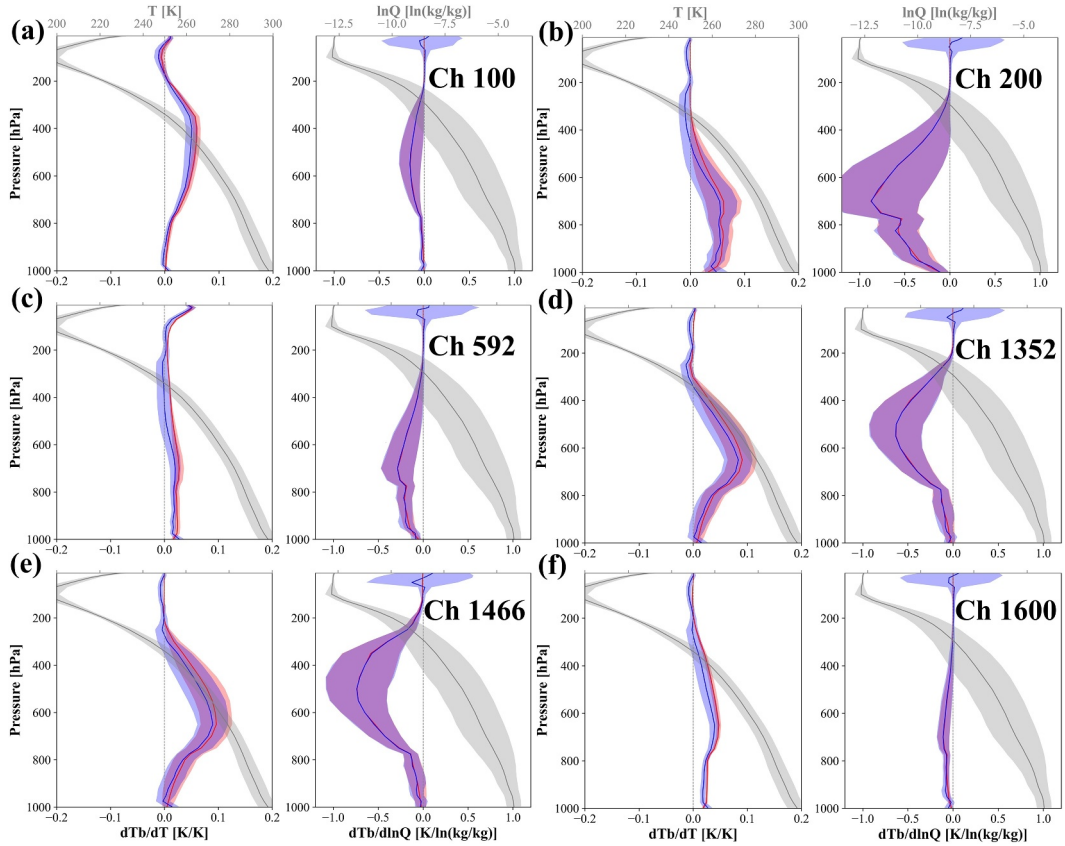

Figure 9. Same as Figure 7, but the Jacobian calculated by BH-RTM-L.

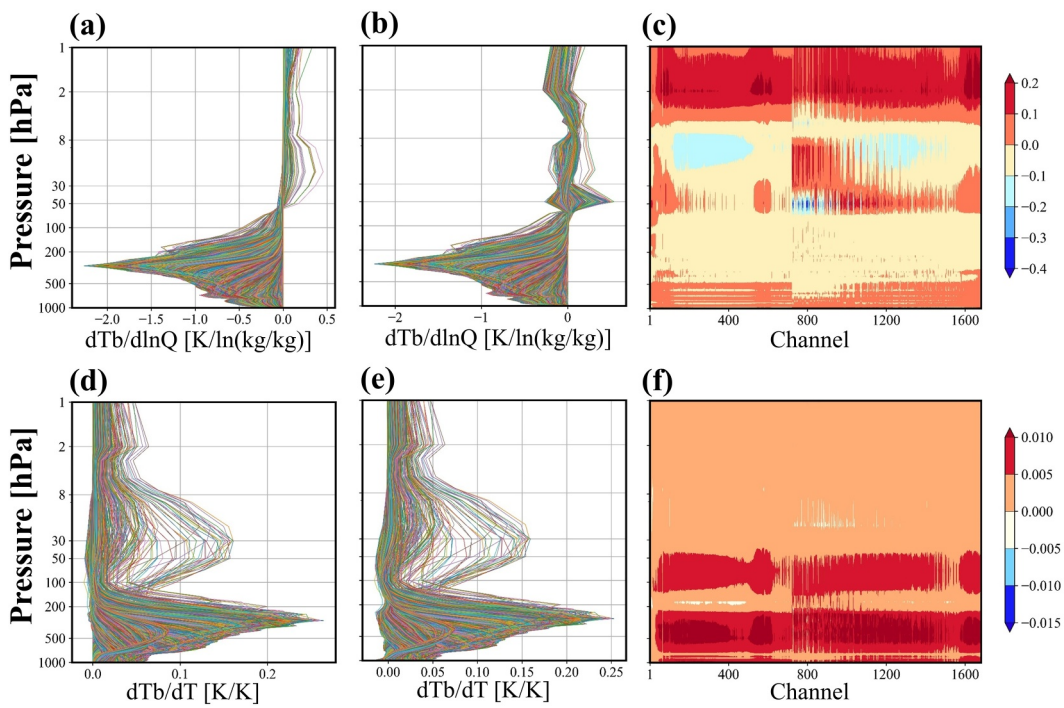

Figure 10. Same as Figure 8, but the Jacobian calculated by BH-RTM-L.

10.1029/2024JH000449

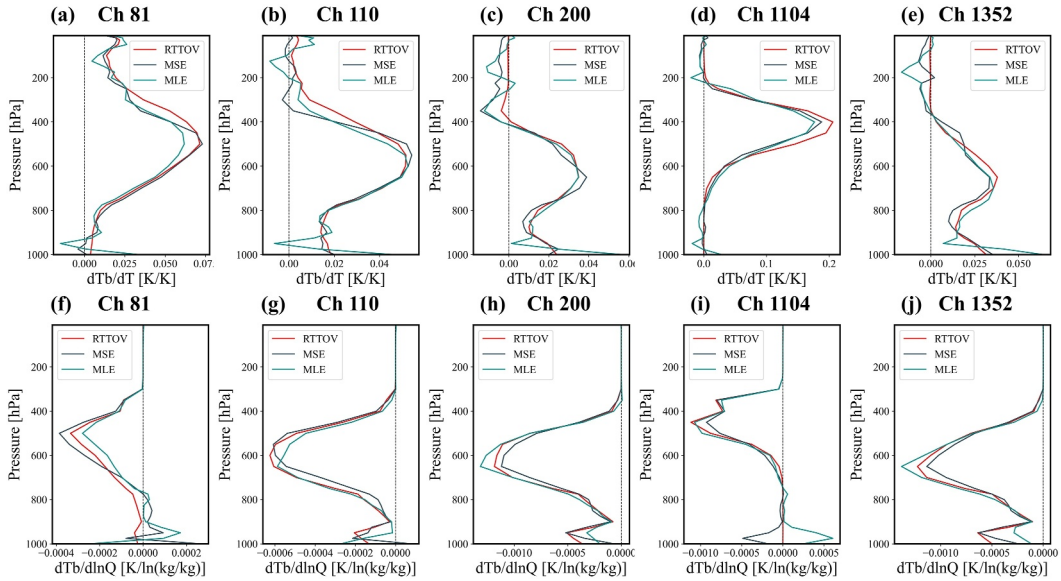

Figure 11. Temperature Jacobian (a)–(e) and water vapor Jacobian (f)–(j) of one profile. Panels (a) and (f) correspond to channel 81, panels (b) and (g) to channel 110, panels (c) and (h) to channel 200, panels (d) and (i) to channel 1,104, and panels (e) and (j) to channel 1,352.

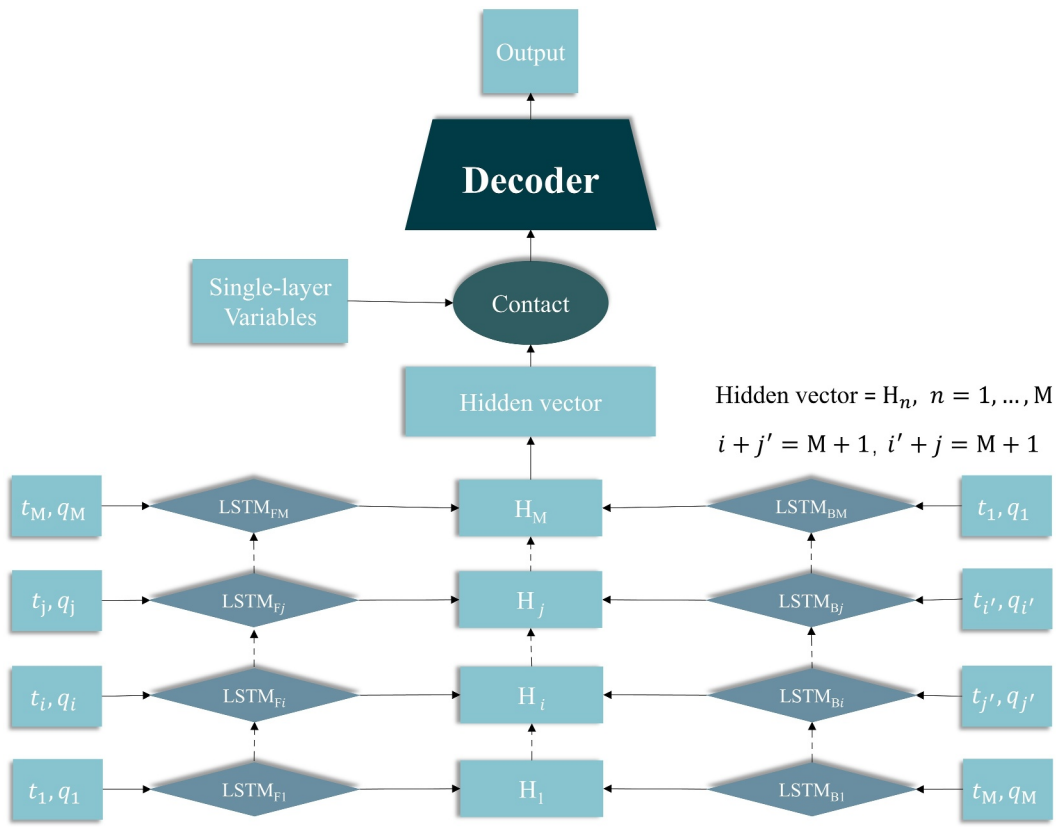

Figure 12. Model Framework about the layer-by-layer output of the model. The "M" in the figure represents the number of layers of the variable. In this context, $M = 37$.

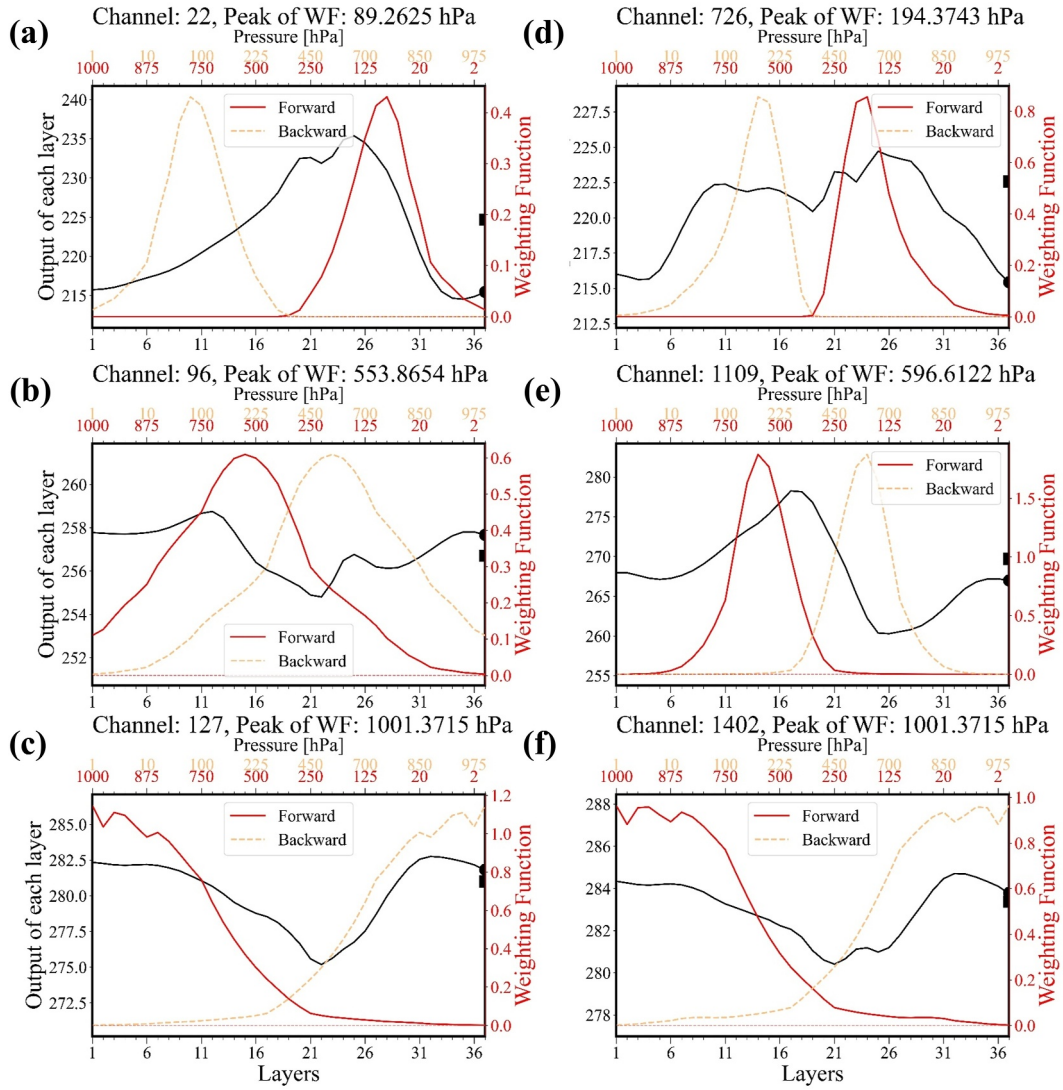

Figure 13. The layer-by-layer outputs of the model. The black line represents the output results at each layer of the model, with the transverse and longitudinal coordinates corresponding to the bottom and left sides of the figure, respectively. The solid red line indicates the forward weight function (WF), which follows the sequence from lower to higher layers, while the dashed orange line represents the backward weight function. The transverse and longitudinal coordinates for both the red and orange lines correspond to the top and right sides of the figure, respectively. The black filled circles correspond to the simulated brightness temperature (BT) results from the RTTOV model for the corresponding channels, which are statistically obtained from the test set. The black square dots represents the result of the dot product between the weights (which are obtained by adding the forward and backward weight functions and normalizing them) and the layer-by-layer outputs (represented by the black line). The values of black filled circles and black square dots correspond to values on the left axis. Panels (a)–(c) represent temperature channels for the high, middle, and low layers, respectively, while panels (d)–(f) represent water vapor channels for the high, middle, and low layers.

different altitudes to satellite instruments. And the red dot, which indicates the outcome of the dot product calculation between the weight that is obtained by adding the forward and backward WF and normalizing and the layer-by-layer output (black line), is closely aligned with the purple dot, representing the BT simulated by RTTOV. The good correspondence between WF and output and close proximity between purple dot and red dot demonstrate that the model has the ability to capture and integrate important information from different levels, assigning higher weights to layers with larger WF, thereby enabling accurate BT prediction. The results of BH-RTM-L are similar.

Table 2
Computation Time (in Seconds) for Forward and Jacobian Calculations Across 1,682 FY-4B GIIRS Channels

	Forward		Jacobian	
	Profile set[a]	Per profile[b]	Profile set	Per profile
RTTOV (CPU-1000)[c]	16.21	12.11	190.14	131.99
ML model (CPU-1000)	3.09	40.71	5,134	3,714
ML model (GPU-1000)	0.14	1.27	13.95	95.17
ML model (GPU-20000)	0.38	24.75	246	2,100

[a]All profiles in one calculation. [b]Each calculation for one profile. [c]Using CPU computation, the data set consists of 1,000 profiles.

4.4. Computation Efficiency

The computational efficiency of the model is very important for its application to data assimilation. We measure computation time using two methods: processing all profiles in a single calculation and processing each profile individually, as both modes are applied in current practices. The GPU used is an NVIDIA H100 GPU, while CPU (Inter(R) Xeon(R) Platinum 8369B CPU) computations are performed using a single thread. We test profile data sets of various sizes. Due to CPU memory limitations, RTTOV encountered memory overflow when processing too many profiles simultaneously for FY-4B GIIRS. Therefore, a data set of 1,000 profiles was used for comparison (For detailed information, please refer to Table 2). The results show that for RTTOV, calculating one profile at a time is slightly faster. Conversely, for the ML model, processing all profiles in one calculation proved to be more efficient. When computing the Jacobian for 1,000 profiles on the CPU for ML model, the computation time was significantly extended. This is attributed to two factors: First of all, the ML model is relatively large since it uses low-dimensional inputs (82) to simulate outputs for 1,682 channels, inherently increasing the model's size. Second, the large number of channels necessitates sequential computation of the Jacobian, which consumes considerable time. This aligns with the findings of Liu et al. (Q. Liu & Liang, 2023), which indicate that the computation time for the Jacobian is directly proportional to the number of channels. Using GPU computation, the efficiency improved substantially. Under the first mode (all profiles in one calculations), the forward computation efficiency is enhanced by two orders of magnitude compared to RTTOV, while the Jacobian computation efficiency improved by one order of magnitude. Under the second mode (each calculation for one profile), the GPU still outperformed RTTOV. As the size of data set increased to 20,000, the GPU's forward computation efficiency becomes even more evident, while the decrease in Jacobian computation efficiency is proportional to the increase in data set size.

5. Conclusion and Discussion

The RTM model calculation for the hyperspectral instrument the FY-4B GIIRS consumes a lot of computing resources. Previous ML work for infrared instruments has been done to attempt addressing above problem, but it is either not end-to-end or not for hyperspectral and it does not account for the uncertainty of labels. This article constructed an end-to-end ML model for the FY-4B GIIRS, which uses encoder-decoder architecture. The encoder uses BiLSTM, and the decoder uses MLP. Two models are trained, BH-RTM-L and BH-RTM-S, using the likelihood function of MLE and MSE as their respective loss function. BH-RTM-S specifically targets minimizing the discrepancy between predicted values and the RTTOV labels, while the goal of BH-RTM-L is to estimate the true BT given the error probability distribution, which assigns different weights for different channels.

The results of model evaluation showed that both models demonstrated good accuracy, with the absolute value of the maximum difference between the two models and the label not exceeding 0.1 K, and the STD not exceeding 0.04 K. Furthermore, BH-RTM-L had a closer mean error to 0 in most bands, while the error uncertainty of BH-RTM-L was more similar to the true error uncertainty. This indicated that the predicted results of BH-RTM-L were more reliable. And similar results could be obtained by selecting a day 2 March 2024 that was not included in the original data set. In addition, the accuracy of Jacobian was evaluated. The temperature Jacobian and water vapor Jacobian calculated by the BH-RTM-S in various bands were in good agreement with the RTTOV, while the Jacobian calculated by BH-RTM-L showed a slight deviation from the RTTOV calculations. Overall, the Jacobian calculated by both models was comparable to the results calculated by RTTOV, highlighting the potential of these models for satellite data assimilation. The high accuracy of these models could be attributed, in part, to its BiLSTM architecture, which could bidirectionally capture the information between different atmospheric layers. The layer-by-layer output demonstrated that the model effectively captured the contribution of each layer to the final brightness temperature, providing a degree of interpretability.

The Infrared Atmospheric Sounder Interferometer - New Generation (IASI-NG) will be embarked on the second generation European Meteorological Polar-orbiting Satellite (MetOp-SG), which continues and improves the IASI mission in the next decades in the field of operational meteorology, climate monitoring, and characterization

of atmospheric composition. IASI-NG has a satellite channel number of 16,921 and a spectral sampling of 0.125 cm^{-1}. Hyperspectral data will bring huge burden to data storage and computing resources, but the operator proposed in this paper has the potential to overcome this problem to some extent (Bermudo et al., 2022). The calculation efficiency of BH-RTM is about 0.38 s for 20,000 samples, and the time for calculating Jacobian of 1,682 channels is about 4 min for 20,000 samples. All of the above calculations are performed on an NVIDIA H100 GPU, which has high computational efficiency and does not require a large amount of computational resources. It is expected to provide a new approach for improving the operational efficiency of satellite data assimilation.

Although the accuracy and computational efficiency of BH-RTM demonstrate its potential, several issues remain. The most important is that BH-RTM is targeted at clear-sky over ocean, which can lead to the inability to observe on land. If it is applied to satellite data assimilation as observational operator in the future, it will miss a large number of observations. Additionally, assumptions in the statistics of true errors, which the errors are unbiased and the error covariance matrix is diagonal, may introduce inaccuracies. Future work should consider more complex error distributions. Despite these limitations, BH-RTM is successfully simulated FY-4B GIIRS data and can accurately characterize the Jacobian matrix, which has the potential to be applied to other satellite instruments in the future. Currently, artificial intelligence (AI) meteorological models are developing rapidly, and some studies have begun to consider building assimilation systems for AI models (Li, Han, Li, et al., 2024; Xiao et al., 2023; Xu et al., 2024). This research can provide radiation transfer operators for AI models, allowing them to assimilate FY-4B GIIRS data, thereby providing strong support for AI to perform cyclic assimilation forecasting independently of traditional NWP, further promoting the application of AI in the field of meteorology.

Appendix A: Estimation of Error Distribution

To estimate the error distribution between the labels (RTTOV) and the true BT, first, we collect the error mean μ_{RML} and covariance matrix Σ_{RML} between RTTOV and the LBL model. Given the small error, it is approximately $\mu_{RML} = 0$ here. We assume that the error ϵ_{RML} follow Gaussian distribution with a mean of 0 and a diagonal covariance Σ_{RML}, that is,

$$\epsilon_{RML} = y_R - y_L \sim \mathcal{N}(0, \Sigma_{RML}) \tag{A1}$$

Then we mention two more assumptions:

1. The random variables representing the error ϵ_R between RTTOV and the true BT, and the error ϵ_L between LBL and the true BT, are assumed to be mutually independent. Both are assumed to follow a Gaussian distribution with a mean of 0 and a diagonal covariance, which assumes that the errors across different channels of FY-4B GIIRS are mutually independent, that is,

$$\epsilon_R = y_t - y_R \sim \mathcal{N}(0, \Sigma_R) \tag{A2}$$

$$\epsilon_L = y_t - y_L \sim \mathcal{N}(0, \Sigma_L) \tag{A3}$$

   where y_R is the BT simulated by RTTOV and y_L is the BT simulated by LBL.
2. Assuming that the LBL model is more accurate compared to RTTOV, it implies that the STD of the error between LBL and the true BT is smaller than that between RTTOV and the true BT, meaning that the diagonal of Σ_R are much more than the diagonal of Σ_L.

Based on the these assumptions, we consider the error distribution between RTTOV and LBL to be the same as that between RTTOV and the true BT, that is, $\Sigma_{RML} \approx \Sigma_R$, indicating that the former uncertainty approximates the latter. Later on, we believe that Σ_R is equal to Σ_{RML}.

Data Availability Statement

The ERA5 reanalysis data is available at https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5 (Hersbach et al., 2020). FY-4B GIIRS L1 data is available at https://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx.

Acknowledgments
This study was jointly supported by the National Natural Science Foundation of China under Grant (U2442219) and National Key R&D Program of China (2022YFC3004004). We thank the editor and the reviewers for their useful feedback that improved this paper.

References

Arking, A., & Grossman, K. (1972). The influence of line shape and band structure on temperatures in planetary atmospheres. *Journal of the Atmospheric Sciences*, *29*(5), 937–949. https://doi.org/10.1175/1520-0469(1972)029<0937:tiolsa>2.0.co;2

Bermudo, F., Jurado, E., Penquer, A., & Le Fèvre, C. (2022). Infrared atmospheric sounding interferometer new generation (IASI-NG): Program development status. In *IGARSS 2022 - 2022 IEEE international geoscience and remote sensing symposium* (pp. 6479–6482). https://doi.org/10.1109/IGARSS46834.2022.9884369

Bouttier, F., & Courtier, P. (2002). Data assimilation concepts and methods march 1999. Meteorological training course lecture series. *ECMWF*, *718*, 59.

Cardinali, C., & Prates, F. (2009). Forecast sensitivity to observations (FSO) as a diagnostic tool: Monitoring the impact of observations on the short-range forecast. In *Proceedings of the ECMWF workshop on diagnostics of data assimilation system performance (ECMWF, 2009)*.

Chevallier, F., Chéruy, F., Scott, N., & Chédin, A. (1998). A neural network approach for a fast and accurate computation of a longwave radiative budget. *Journal of Applied Meteorology and Climatology*, *37*(11), 1385–1397.

Eresmaa, R., Letertre-Danczak, J., Lupu, C., Bormann, N., & McNally, A. P. (2017). The assimilation of cross-track infrared sounder radiances at ECMWF. *Quarterly Journal of the Royal Meteorological Society*, *143*(709), 3177–3188. https://doi.org/10.1002/qj.3171

Feng, J., Qin, X., Wu, C., Zhang, P., Yang, L., Shen, X., et al. (2022). Improving typhoon predictions by assimilating the retrieval of atmospheric temperature profiles from the fengyun-4a's geostationary interferometric infrared sounder (GIIRS). *Atmospheric Research*, *280*, 106391. https://doi.org/10.1016/j.atmosres.2022.106391

Fu, Q., & Liou, K. (1992). On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *Journal of the Atmospheric Sciences*, *49*(22), 2139–2156. https://doi.org/10.1175/1520-0469(1992)049<2139:otcdmf>2.0.co;2

Geer, A., Bormann, N., Lonitz, K., Weston, P., Forbes, R., & English, S. (2019). Recent progress in all-sky radiance assimilation. *ECMWF Newsl*, *8*, 20–25.

Geer, A. J., Migliorini, S., & Matricardi, M. (2019). All-sky assimilation of infrared radiances sensitive to mid-and upper-tropospheric moisture and cloud. *Atmospheric Measurement Techniques*, *12*(9), 4903–4929. https://doi.org/10.5194/amt-12-4903-2019

Graham, R. M., Hudson, S. R., & Maturilli, M. (2019). Improved performance of ERA5 in arctic gateway relative to four global atmospheric reanalyses. *Geophysical Research Letters*, *46*(11), 6138–6147. https://doi.org/10.1029/2019gl082781

Han, W., Yin, R., Li, J., Shen, X., Wang, H., Wang, J., et al. (2023). Assimilation of geostationary hyperspectral infrared sounders (GEOHIS): Progresses and perspectives. In *Numerical weather prediction: East Asian perspectives* (pp. 205–216). Springer.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis [Dataset]. *Wiley Online Library*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Kratz, D. P. (1995). The correlated k-distribution technique as applied to the AVHRR channels. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *53*(5), 501–517. https://doi.org/10.1016/0022-4073(95)00006-7

Lacis, A. A., & Oinas, V. (1991). A description of the correlated k distribution method for modeling nongray gaseous absorption, thermal emission, and multiple scattering in vertically inhomogeneous atmospheres. *Journal of Geophysical Research*, *96*(D5), 9027–9063. https://doi.org/10.1029/90jd01945

Le, T., Liu, C., Yao, B., Natraj, V., & Yung, Y. L. (2020). Application of machine learning to hyperspectral radiative transfer simulations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *246*, 106928. https://doi.org/10.1016/j.jqsrt.2020.106928

Li, Y., Han, W., & Duan, W. (2024). Dynamic channel selection based on vertical sensitivities for the assimilation of fy-4a geostationary interferometric infrared sounder targeted observations. *Quarterly Journal of the Royal Meteorological Society*, *150*(763), 1–17. https://doi.org/10.1002/qj.4760

Li, Y., Han, W., Li, H., Duan, W., Chen, L., Zhong, X., et al. (2024). Fuxi-en4dvar: An assimilation system based on machine learning weather forecasting model ensuring physical constraints. *Geophysical Research Letters*, *51*(22), e2024GL111136. https://doi.org/10.1029/2024GL111136

Liang, X., Garrett, K., Liu, Q., Maddy, E. S., Ide, K., & Boukabara, S. (2022). A deep-learning-based microwave radiative transfer emulator for data assimilation and remote sensing. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *15*, 8819–8833. https://doi.org/10.1109/jstars.2022.3210491

Liang, X.-M., Ignatov, A., & Kihai, Y. (2009). Implementation of the community radiative transfer model in advanced clear-sky processor for oceans and validation against nighttime AVHRR radiances. *Journal of Geophysical Research*, *114*(D6), D06112. https://doi.org/10.1029/2008jd010960

Liu, C., Yang, P., Nasiri, S. L., Platnick, S., Meyer, K. G., Wang, C., & Ding, S. (2015). A fast visible infrared imaging radiometer suite simulator for cloudy atmospheres. *Journal of Geophysical Research: Atmospheres*, *120*(1), 240–255. https://doi.org/10.1002/2014jd022443

Liu, C., Yao, B., Natraj, V., Weng, F., Le, T., Shia, R.-L., & Yung, Y. L. (2020). A spectral data compression (SDCOMP) radiative transfer model for high-spectral-resolution radiation simulations. *Journal of the Atmospheric Sciences*, *77*(6), 2055–2066. https://doi.org/10.1175/jas-d-19-0238.1

Liu, Q., & Liang, X. (2023). Physics constraint deep learning based radiative transfer model. *Optics Express*, *31*(17), 28596–28610. https://doi.org/10.1364/oe.493818

Liu, X., Smith, W. L., Zhou, D. K., & Larar, A. (2006). Principal component-based radiative transfer model for hyperspectral sensors: Theoretical concept. *Applied Optics*, *45*(1), 201–209. https://doi.org/10.1364/ao.45.000201

Liu, X., Zhou, D., Larar, A., Smith, W., Schluessel, P., Newman, S., et al. (2009). Retrieval of atmospheric profiles and cloud properties from IASI spectra using super-channels. *Atmospheric Chemistry and Physics*, *9*(23), 9121–9142. https://doi.org/10.5194/acp-9-9121-2009

McNally, T., Bonavita, M., & Thépaut, J.-N. (2014). The role of satellite data in the forecasting of hurricane sandy. *Monthly Weather Review*, *142*(2), 634–646. https://doi.org/10.1175/mwr-d-13-00170.1

Meador, W., & Weaver, W. (1980). Two-stream approximations to radiative transfer in planetary atmospheres: A unified description of existing methods and a new improvement. *Journal of the Atmospheric Sciences*, *37*(3), 630–643. https://doi.org/10.1175/1520-0469(1980)037<0630:tsatrt>2.0.co;2

Moncet, J.-L., Uymin, G., Lipton, A. E., & Snell, H. E. (2008). Infrared radiance modeling by optimal spectral sampling. *Journal of the Atmospheric Sciences*, *65*(12), 3917–3934. https://doi.org/10.1175/2008jas2711.1

Natraj, V., & Spurr, R. J. (2007). A fast linearized pseudo-spherical two orders of scattering model to account for polarization in vertically inhomogeneous scattering–absorbing media. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *107*(2), 263–293. https://doi.org/10.1016/j.jqsrt.2007.02.011

Niu, Z., Zhang, L., Han, Y., Dong, P., & Huang, W. (2023). Performances between the fy-4a/GIIRS and fy-4b/GIIRS long-wave infrared (LWIR) channels under clear-sky and all-sky conditions. *Quarterly Journal of the Royal Meteorological Society*, *149*(754), 1612–1628. https://doi.org/10.1002/qj.4473

Saunders, R., Matricardi, M., & Brunel, P. (1999). An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quarterly Journal of the Royal Meteorological Society*, *125*(556), 1407–1425. https://doi.org/10.1256/smsqj.55614

Shi, G., Xu, N., Wang, B., Dai, T., & Zhao, J. (2009). An improved treatment of overlapping absorption bands based on the correlated k distribution model for thermal infrared radiative transfer calculations. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *110*(8), 435–451. https://doi.org/10.1016/j.jqsrt.2009.01.008

Stegmann, P. G., Johnson, B., Moradi, I., Karpowicz, B., & McCarty, W. (2022). A deep learning approach to fast radiative transfer. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *280*, 108088. https://doi.org/10.1016/j.jqsrt.2022.108088

Su, M., Liu, C., Di, D., Le, T., Sun, Y., Li, J., et al. (2023). A multi-domain compression radiative transfer model for the fengyun-4 geosynchronous interferometric infrared sounder (GIIRS). *Advances in Atmospheric Sciences*, *40*(10), 1844–1858. https://doi.org/10.1007/s00376-023-2293-5

Tetzner, D., Thomas, E., & Allen, C. (2019). A validation of ERA5 reanalysis data in the southern Antarctic Peninsula—Ellsworth land region, and its implications for ice core studies. *Geosciences*, *9*(7), 289. https://doi.org/10.3390/geosciences9070289

Ukkonen, P. (2022). Exploring pathways to more accurate machine learning emulation of atmospheric radiative transfer. *Journal of Advances in Modeling Earth Systems*, *14*(4), e2021MS002875. https://doi.org/10.1029/2021ms002875

Wang, C., Yang, P., Baum, B. A., Platnick, S., Heidinger, A. K., Hu, Y., & Holz, R. E. (2011). Retrieval of ice cloud optical thickness and effective particle size using a fast infrared radiative transfer model. *Journal of Applied Meteorology and Climatology*, *50*(11), 2283–2297. https://doi.org/10.1175/jamc-d-11-067.1

Weng, F., & Liu, Q. (2003). Satellite data assimilation in numerical weather prediction models. Part I: Forward radiative transfer and Jacobian modeling in cloudy atmospheres. *Journal of the Atmospheric Sciences*, *60*(21), 2633–2646. https://doi.org/10.1175/1520-0469(2003)060<2633:sdainw>2.0.co;2

Weng, F., Yu, X., Duan, Y., Yang, J., & Wang, J. (2020). *Advanced radiative transfer modeling system (arms): A new-generation satellite observation operator developed for numerical weather prediction and remote sensing applications*. Springer.

Xiao, Y., Bai, L., Xue, W., Chen, K., Han, T., & Ouyang, W. (2023). Fengwu-4dvar: Coupling the data-driven weather forecasting model with 4D variational assimilation. *arXiv preprint arXiv:2312.12455*.

Xu, X., Sun, X., Han, W., Zhong, X., Chen, L., & Li, H. (2024). Fuxi-da: A generalized deep learning data assimilation framework for assimilating satellite observations. *arXiv preprint arXiv:2404.08522*.

Yang, J., Zhang, Z., Wei, C., Lu, F., & Guo, Q. (2017). Introducing the new generation of Chinese geostationary weather satellites, fengyun-4. *Bulletin of the American Meteorological Society*, *98*(8), 1637–1658. https://doi.org/10.1175/bams-d-16-0065.1

Yin, R., Han, W., Gao, Z., & Li, J. (2021). 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. *Geophysical Research Letters*, *48*(15), e2021GL093672. https://doi.org/10.1029/2021gl093672

Yin, R., Han, W., Wang, H., & Wang, J. (2022). Impacts of FY-4A GIIRS water vapor channels data assimilation on the forecast of "21· 7" extreme rainstorm in Henan, China with CMA-MESO. *Remote Sensing*, *14*(22), 5710. https://doi.org/10.3390/rs14225710

Zhang, S., Zheng, D., Hu, X., & Yang, M. (2015). Bidirectional long short-term memory networks for relation classification. In *Proceedings of the 29th pacific Asia conference on language, information and computation* (pp. 73–78).