# A Deep-Learning-Based Microwave Radiative Transfer Emulator for Data Assimilation and Remote Sensing

Xingming Liang, Kevin Garrett, Quanhua Liu, Eric S. Maddy, Kayo Ide, and Sid Boukabara

***Abstract*—In this article, we introduce a fully connected deep neural network algorithm to emulate the Community Cadiative Transfer Model (FCDN_CRTM) simulation of brightness temperatures (BTs) from the Advanced Technology Microwave Sounder (ATMS) channels for clear-sky cases over ocean surfaces. The FCDN_CRTM fine-tuned through three sensitivity experiments with respect to sample-size determination, model separation, and introduction of novel features toward improving the accuracy of the model. In addition to the BT simulation, we evaluated the Jacobians with respect to surface and atmospheric parameters. Atmosphere profiles from the European Centre for Medium-Range Weather Forecasts, sea surface temperature from the Canadian Meteorology Centre, and ATMS sensor data records were used as FCDN_CRTM inputs. In comparison to CRTM, the FCDN_CRTM minus CRTM mean biases were within several hundredths of a Kelvin (K), and the corresponding standard deviations (SDs) were between 0.05 and 0.15 K for all ATMS bands. The accuracies for both mean bias and SD were consistent throughout the evaluation period, which spanned approximately 1 year beyond the period of the FCDN_CRTM training dataset. Furthermore, the model Jacobians generally compared well with CRTM Jacobians in terms of surface temperature, wind speed, air temperature, and (log) water vapor. The performance of the FCDN_CRTM forward and Jacobian model indicate potential for use in data assimilation and physical retrieval systems, such as the NOAA operational Microwave Integrated Retrieval System.**

***Index Terms*—Advanced Technology Microwave Sounder (ATMS), Community Radiative Transfer Model (CRTM), data assimilation, deep learning (DL), Jacobian, remote sensing.**

Manuscript received 20 May 2022; revised 21 July 2022 and 21 August 2022; accepted 20 September 2022. Date of publication 4 October 2022; date of current version 20 October 2022. This work was supported by the National Oceanic and Atmospheric Administration (NOAA) under Grant NA20OAR4600287 and Grant NA19NES4320002 to Cooperative Institute for Satellite Earth System Studies at the Earth System Science Interdisciplinary Center (CISESS) of the University of Maryland/ESSIC. *(Corresponding author: Xingming Liang.)*

Xingming Liang is with the CISESS, University of Maryland, College Park, MD 20740 USA (e-mail: xingming.liang@noaa.gov).

Kevin Garrett and Quanhua Liu are with the Center for Satellite Applications and Research, National Environmental Satellite Data, and Information Service (NESDIS), NOAA, College Park, MD 20740 USA (e-mail: kevin.garrett@noaa.gov; quanhua.liu@noaa.gov).

Eric S. Maddy is with the Riverside Technology Inc., College Park, MD 20740 USA (e-mail: eric.maddy@noaa.gov).

Kayo Ide is with the University of Maryland, College Park, MD 20740 USA (e-mail: kayo.ide@noaa.gov).

Sid Boukabara is with The Office of Systems Architecture and Advanced Planning, NESDIS, NOAA, College Park, MD 20740 USA (e-mail: sid.boukabara@noaa.gov).

Digital Object Identifier 10.1109/JSTARS.2022.3210491

## I. Introduction

THE radiative transfer model (RTM), which is vital for satellite radiance calibration, data assimilation, and remote sensing, must efficiently and accurately simulate large volumes of satellite observations to meet performance requirements in numerical weather prediction (NWP) systems and near real-time monitoring of satellite sensors. However, RTMs are one of the most complicated and computationally expensive components in data assimilation and remote sensing. Therefore, to improve computational efficiency, several RTMs, such as the Community Radiative Transfer Model (CRTM) [1], [2], the Radiative Transfer for TOVS (RTTOV) [3], [4], and the recently developed Advanced Radiative Transfer Modeling System (ARMS) [5], parameterize atmospheric transmittance through regression models trained offline, instead of performing line-by-line calculations. However, with the development of high spatial, spectral, and temporal resolution sensors, such as the Visible Infrared Imaging Radiometer Suite (VIIRS), Advanced Technology Microwave Sounder (ATMS), Cross-track Infrared Sounder onboard the new generation Joint Polar Satellite System (JPSS), and Advanced Baseline Imager on the Geostationary Operational Environmental Satellites—R Series, demand for computational resources continues to increase in NWP systems where execution time allotted is nearly fixed to ensure timely production of forecast products. In NWP systems, such as the Global Forecast System of the National Centers for Environmental Prediction or the European Centre for Medium-Range Weather Forecasts (ECMWF), the efficiency of RTMs remains a key issue.

With rapidly evolving, cutting-edge artificial intelligence (AI), deep learning (DL), which is one of the most widespread AI methods, has made a remarkable impact on numerous fields of science and engineering. Its application in remote sensing and NWP is presently being explored [6], [7], [8] with the applications of AI in RTM growing rapidly. An initial shallow neural network (SNN)-based model, NeuroFlux, has been developed to accurately determine the longwave radiation budget from the top of the atmosphere to the surface of the Earth [9]. The SNNs have been further expanded to more complex architectures to include multiple hidden layers (deep neural network, DNNs) and improve RTM accuracy and efficiency, while demonstrating potential to replace conventional RTMs in climate models [10], [11], [12], [13], [14].

We have developed a fully connected DNN (FCDN_CRTM), with CRTM as the reference to emulate the measurements of the

M-thermal emission bands (TEB/M) for the VIIRS onboard the satellites of the JPSS and to improve RTM efficiency and accuracy [14] for global real-time data monitoring of the sensor radiometric biases through applications such as the NOAA Integrated Calibration/Validation System (ICVS) [15]. The FCDN_CRTM used upper-air profiles from the ECMWF and sea surface temperature (SST) from the Canadian Meteorology Centre (CMC) [16] and VIIRS sensor data records (SDR) as inputs to reproduce VIIRS brightness temperatures (BTs) under clear-sky conditions over the oceans globally. An FCDN-based clear-sky mask (FCDN_CSM) algorithm [17], [18] was used to identify clear-sky pixels. Preliminary results showed that both the mean bias and standard deviation (SD) of FCDN_CRTM minus CRTM simulations were only several thousandths or hundredths of a Kelvin, and the FCDN_CRTM calculations were more than 40 times faster than those of the CRTM. Overall, compared to traditional RTM, highly accurate and efficient FCDN_CRTM has been designed to monitor VIIRS TEB/M radiometric biases in real time.

The CRTM applied to microwave sensors, such as ATMS, is a critical component for the assimilation of satellite radiances in NOAA's regional and global data assimilation systems [19], physical retrieval algorithms like the operational Microwave Integrated Retrieval System (MiRS) [20], and sensor calibration/validation as with the ICVS [15]. Therefore, this article explored the applications of an FCDN in RTM emulation for microwave sensors, and extended FCDN_CRTM functionality to simulate BTs for the 22 ATMS bands. We also investigated the partial derivatives (PDs) of BTs with respect to the input surface and atmosphere parameters (i.e., model Jacobian), which are required for data assimilation. Section II discusses the FCDN model architecture. Section III describes the data and preprocessing used, while Section IV demonstrates the model training, tuning, and testing. The FCDN_CRTM BT simulation performance is covered in Section V, while Section VI discusses the FCDN_CRTM Jacobian performance. Finally, the conclusions and future work are summarized in Section VII.

## II. FCDN_CRTM Model Architecture Used for ATMS

The architecture of FCDN_CRTM initially developed for VIIRS BT simulation, comprised of three hidden layers with 384, 512, and 64 neurons in each layer, respectively. ReLU was used as the default activation function, and mean square error (MSE) was employed as the loss function. Furthermore, batch normalization (BN) [21] and regularization were introduced in the model to improve model accuracy, efficiency, and generalization [14].

The ATMS is a new generation of microwave sounders covering the frequency range of 23.8–190 GHz (in 22 bands) to provide information on the atmospheric vertical temperature and moisture profiles that are the most critical observations needed for numerical weather forecast models [22]. Since the FCDN_CRTM has been extensively tested and evaluated for use in the VIIRS TEB/M bands, we initially used its model architecture in this study for the simulation of ATMS BTs. In the meantime, the state-of-the-art hyper tuning tool—KERAS Tuner (https://www.tensorflow.org/tutorials/keras/keras_tuner)—was used to determine the performance of FCDN_CRTM. However, the input data set for the FCDN_CRTM with ATMS was different from that for VIIRS, such as input features (discussed in the next subsection), input data samples, and CRTM simulations. Therefore, owing to the different spectral regions examined by VIIRS and ATMS, model fine-tuning was required to ensure high accuracy and stability while predicting ATMS BTs. The details of model tuning and comparison before and after tuning are discussed in Sections IV and V.

TABLE I
Summary of Input Features and Output Brightness Temperatures (BTs) in the ATMS FCDN_CRTM

| Input features | | | Reference labels | |
|---|---|---|---|---|
| **Names** | **Number** | **Data source** | **Name** | **Number** |
| Secant of SCA | 1 | ATMS SDR | CRTM BTs (Ch1-Ch22) | 22 |
| Secant of SZA | 1 | ATMS SDR | | |
| Cosine of RAZ | 1 | ATMS SDR | | |
| Wind speed | 1 | ECMWF | | |
| Sea surface temperature | 1 | CMC SST | | |
| Surface pressure | 1 | ECMWF | | |
| Air temperatures | 91 | ECMWF | | |
| water vapor contents | 91 | ECMWF | | |
| Total | 188 | | Total | 22 |

SCA: satellite scan angle; SZA: satellite zenith angle; RAZ: relative azimuth angle.

The input features were based on the inputs to CRTM simulations for ATMS measurements. Most inputs were similar to the VIIRS FCDN_CRTM except for the ozone profile, which was not used in this article, as the contribution of ozone is negligible with respect to atmosphere absorption and scattering for ATMS bands. CRTM simulations for the 22 ATMS bands were used as model reference labels. Similar to the VIIRS FCDN_CRTM, atmospheric temperature and water vapor profiles, surface pressure, and wind speed (WS) were extracted from the ECMWF analysis (https://www.ecmwf.int) with a 0.25° horizontal resolution and 91 vertical layers from the surface to 0.01 hPa, which are downscaled from its original 137 vertical levels when the data is transited from ECMWF to NOAA. Four analysis files per day are acquired at 00, 06, 12, and 18 UTC and are used to collocate with ATMS SDR in time. The SST was obtained from the analysis of 0.1° resolution daily CMC SST (https://podaac.jpl.nasa.gov/dataset/CMC0.1deg-CMC-L4-GLOB-v3.0). The other parameters were determined using the ATMS geography (GEO) products, such as satellite zenith angle, satellite azimuth angle, and sensor scan angle. The total number of input features was 188, with corresponding 22 CRTM-simulated ATMS BTs used as reference labels. Table I lists the input features and reference labels of the initial model. Also the longitude and latitude from ATMS GEO products and the sensor BTs from SDR products are extracted simultaneously to assist model evaluation. Note that the CRTM is a fast and approximate RTM, which uses trained transmittance coefficients to fit a line-by-line RTM (LBLRTM) [23], [24]. Here, the use of CRTM simulations as reference labels

will be more efficient in generating the large dataset needed for FCDN_CRTM training, testing, and evaluation. However, future work will consider using LBLRTM for generating training data not only for further FCDN_CRTM validation, but particularly for FCDN_CRTM applied to hyperspectral IR sensors where some spectral regions are sensitive to multiple trace gases.

## III. Data Preprocessing and CRTM Simulations

Data preprocessing refers to collecting input data and generating reference labels for FCDN_CRTM. Dispersed data obtained from NOAA-20 for 12 days were utilized as FCDN_CRTM input data. We used data from the 26th day of each month between December 2019 and November 2020, which encompassed the atmospheric and surface conditions in all seasons. The selection of the 26th was somewhat arbitrary, but the data met criteria based on the completeness, wherein no gaps or duplications were reported in the archive of the NOAA-20 geolocation (GEO) and SDR products, as well as ECMWF and CMC SST.

After collecting ECMWF, CMC SST, and ATMS SDR data for 12 days, one of primary steps in data preprocessing—generating FCDN_CRTM reference labels, namely 22 ATMS BTs—was performed using CRTM simulations. The simulations were created by interpolating both ECMWF quarter degree and CMC SST grid data to the ATMS SDR pixels in space and time for input to CRTM, along with the ATMS scan geometry.

In the preliminary experiment, we assumed that clouds were transparent in the microwave region and thus did not include cloud parameters in the CRTM simulations. Additionally, we only focused on the global ocean domain. These assumptions, which are likely to result in large simulation biases and uncertainties in some ATMS bands compared to clear-sky conditions, will be discussed later. However, excluding cloud profiles and different types of clouds reduced several hundreds of input features and simplified the design of the initial model. Consequently, the initial model could rapidly evaluate FCDN_CRTM performance after transferring from VIIRS to ATMS, determine model Jacobian performance, and assist in the development of an all-sky model for ATMS, which would include cloud parameters and land surface analysis. Nonetheless, development of the ATMS FCDN_CRTM under all-sky conditions is currently under way.

In this study, the latest CRTM version 2.3 was used to simulate ATMS BTs. Fig. 1 illustrates the global distribution of the observed ATMS BTs minus CRTM simulations of bands 1, 10, and 22 from the data obtained for 22 Feb. 2021, and their corresponding histograms were shown in Fig. 2. We observed large bias and SD in band 1 (4.403 $\pm$ 7.328 K), along with skewness, primarily because of the large uncertainty in surface emissivity and missing cloud parameters. However, for the high peaking channel, band 10, the mean bias and SD (0.444 $\pm$ 0.316 K) decreased, as this band is not sensitive to clouds or the surface. Slightly large SDs in band 22 ($-0.067 \pm 2.203$ K) may be attributed to the higher noise in the 183 GHz bands along with sensitivity to frozen hydrometeors. Although the accuracy of the CRTM simulations were significantly affected by surface emissivity, atmosphere profiles, and missing cloud parameters, we did not conduct any quality control and maintained all pixels

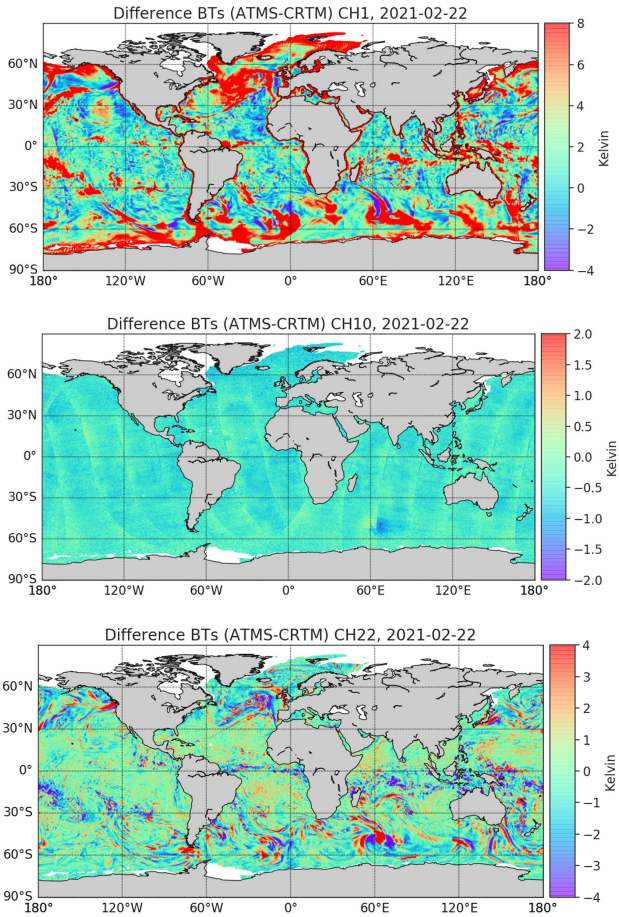


Fig. 1. Global distribution of ATMS observation minus CRTM simulation for ATMS band 1, 10, and 22.

in the global scenario to train FCDN_CRTM. This was done to determine the generalization performance of the model under any atmospheric condition.

## IV. FCDN_CRTM Tuning to Improve Model Accuracy

As discussed in Section II, we first trained and tested the basic architecture of FCDN_CRTM with ATMS data as inputs (referred to initial test), and then focused on tuning the model for sample-size determination, model separation, and introduction of novel features based on three sensitivity experiments.

As the magnitude of input data varied because of different units, data were standardized using their means ($\mu$) and SDs ($\sigma$) before they were used for model training

$$\bar{x} = \frac{x - \mu}{\sigma} \tag{1}$$

where $x$ and $\bar{x}$ stand for the model input data before and after standardization, respectively. Approximately 23 million (M) data samples were collected from the 12 days of CRTM simulations. The collected data were adequately shuffled and roughly separated into training, testing, and validation datasets in the ratio 8:1:1, wherein the number of training data samples was approximately 20 M.

The convergence status of the loss for the initial test are shown in panel (a, blue curve) of Fig. 3; and the corresponding statistics

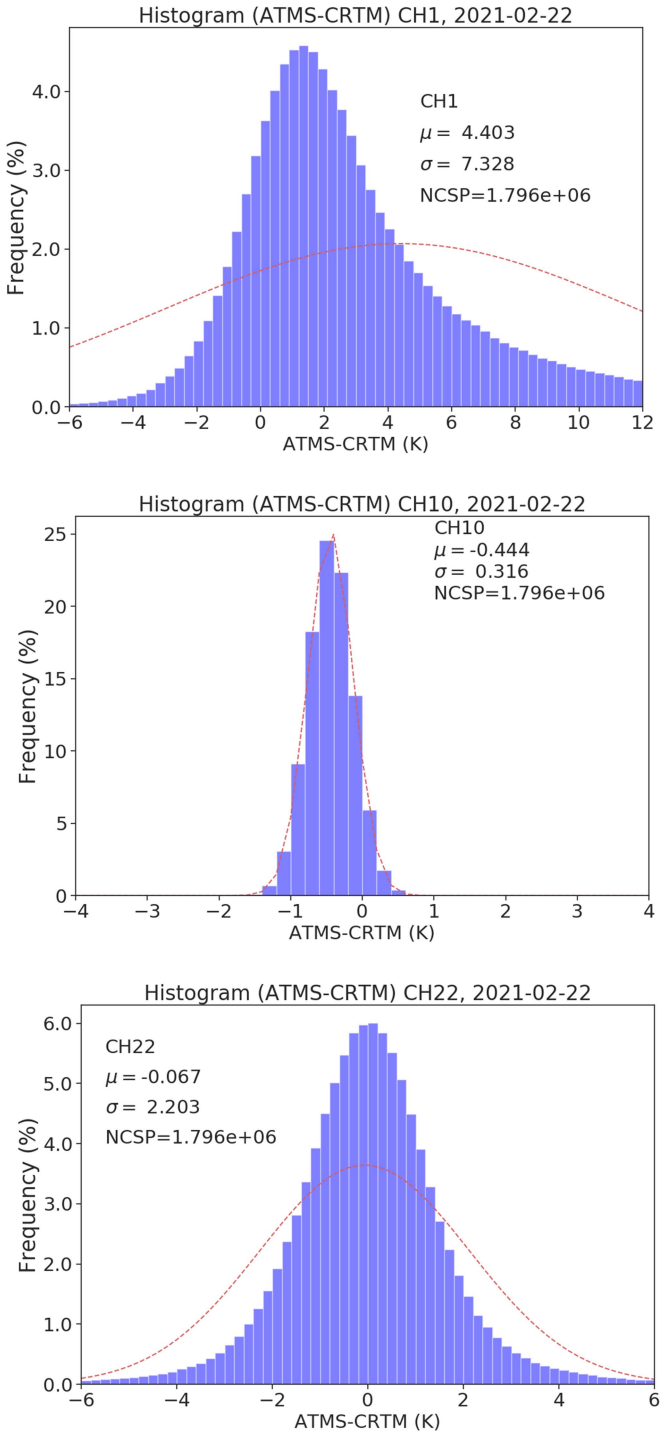


Fig. 2. Histogram of ATMS observation minus CRTM simulation for ATMS band 1, 10, and 22.

of FCDN_CRTM prediction minus CRTM simulation are shown in Fig. 4. The loss is close to coverage at about 10 million iterations. For all 22 bands, the mean biases of FCDN_CRTM minus CRTM are up to −0.15 K and SDs are within 0.2 K, indicating that the three-hidden-layer design, with BN and regularization algorithms, was still one of the good designs to predict ATMS BTs with comparison to the difference between CRTM simulation and ATMS BTs. But for high demand of sensor validation and data assimilation, model tuning is needed to further improve prediction accuracy.

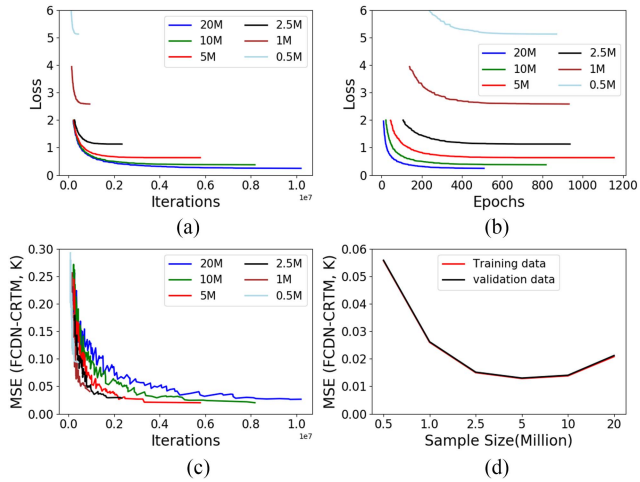


Fig. 3. Model convergence for six different training sample data sizes. (a) Selected decreasing loss as function of iterations (1024 data samples as inputs for each iteration). (b) Same as (a), but as function of epochs. (c) Same as (a), but for mean square error (MSE). (d) MSEs between the training and validation datasets for six cases.

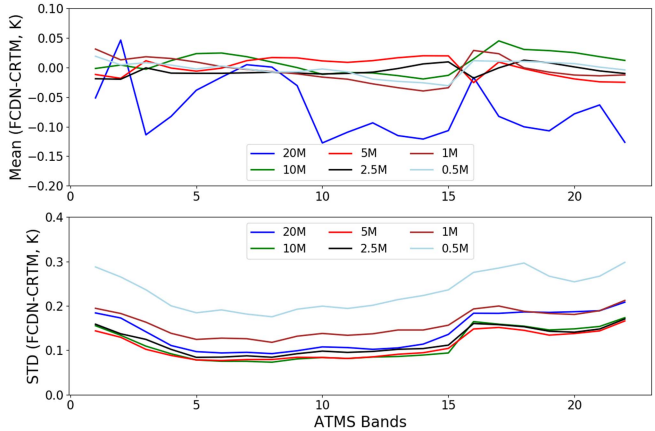


Fig. 4. Mean bias and SD of FCDN_CRTM minus CRTM simulations for the test dataset.

### A. *Sample Size Determination for Model Training*

A large number of data samples, for several months and even for an entire year, can be generated with ample resources and time. A key question is: what is the amount and representation of the data to optimize the training and model performance? Generally, too small data samples will make the model less accurate, whereas large data samples will require more time and memory space, although they will increase the accuracy and generalization of the model. Additionally, most DL models risk overfitting without adequate data representation. Since the input sample size is a considerable hyperparameter in FCDN_CRTM, a sensitivity experiment was conducted using six different sample sizes, including 0.5 M, 1 M, 2.5 M, 5 M, 10 M, and entire 20 M data samples (e.f. initial test), to train the model and determine the sample size of the data. The data samples of the former five cases were randomly selected from the 20 M training dataset.

Fig. 3 shows model convergence and learning curves while training the model using six different data sizes as FCDN_CRTM inputs. Panels (a) and (b) show decreasing loss

as a function of iterations and epochs. There were 1024 data sample inputs (batch size) for FCDN_CRTM in each iteration, with more data samples resulting in longer model convergence times. Furthermore, the loss converged at the smallest value for the largest sample size (20 M). Panel (c) is similar to panel (a), but indicates MSE. Panel (d) shows the MSEs between the training and validation datasets for all six cases. Note that as shown in (2) of [14], the loss function differs from MSE when the regularization is introduced in the model, which includes an L2 regularization term, in addition to MSE. The MSEs are consistent between the two datasets, indicating no significant overfitting for all cases. Interestingly, panels (c) and (d) reveal that at the end of the training period, MSE did not converge for the smallest value in the 20 M case with both training and validation datasets, although the loss functions did. Fig. 4 showed the mean biases and SDs of the FCDN_CRTM prediction minus CRTM simulation for test dataset for six cases. It is obvious that the most accurate prediction was not observed for 20 M case, but for the 5 M case, in which the mean biases are less than 0.03 K in comparison to 0.15 K in the 20 M case; and the SDs are reduced ∼0.03 K for all bands.

Therefore, we hypothesized that the 20 M case may not completely converge, and may need more time to train. However, we more suspected that the loss may not fully optimized due to the limitation of the model complexity. The large-scale data samples may require a more complex framework for training. On the other hand, the more complicated the model is, the less efficiency the model will be. Overall, the 5 M case showed the high prediction accuracy, and is the best fit for the FCDN_CRTM training among the six cases. Hereafter, we use 5 M data samples to train the model and continue to do latter sensitivity experiments, and leave the exploring more complicated model for future work.

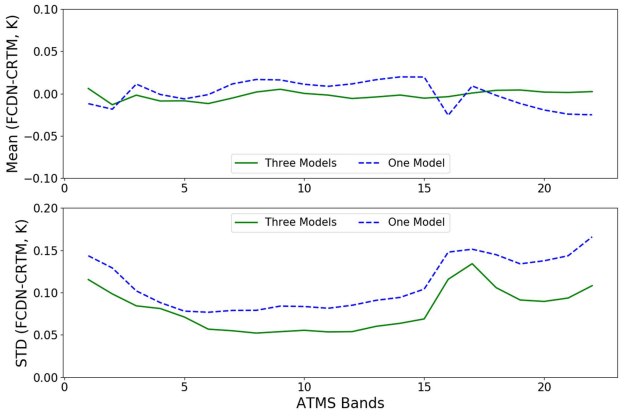


Fig. 5. Mean bias (upper) and SD (lower) of FCDN_CRTM minus CRTM between the one-model and three-model design.

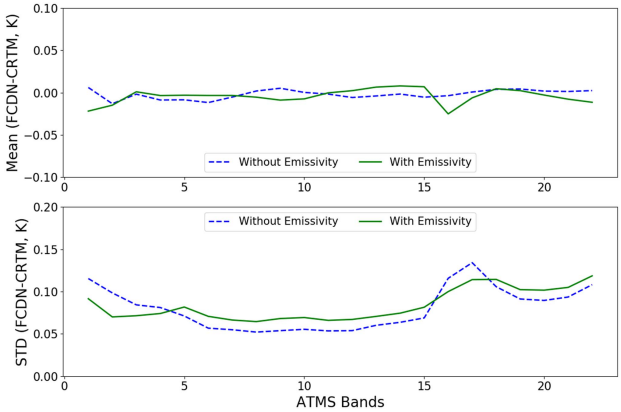


Fig. 6. Mean bias (upper) and SD (lower) of FCDN_CRTM minus CRTM between the models with and without emissivity as an input feature.

### *B. Three-Model Design Based on Band Radiation Properties*

The 22 ATMS bands are sensitive to input data at varying degrees. Bands 5–15 are sensitive to oxygen absorption and therefore atmospheric temperature, but not the surface or low clouds, and are referred to as temperature sounding bands (T-bands). Bands 1–4 and 16–17 are surface bands (S-bands) and more sensitive to the surface and clouds. Bands 18–22 are humidity sounding bands (H-bands) and more sensitive to water vapor and frozen hydrometeors. If the bands are trained using one model, with different sensitivities to the input data, they may interact with, and limit, model accuracy. To identify and minimize these effects on model accuracy, we trained the three band groups separately (three-model design).

Fig. 5 shows the mean biases (upper) and SDs (lower) of FCDN_CRTM minus CRTM between one-model and three-model designs. We observed that the three-model design reduced SD by up to 0.05 K compared to the one-model design.

### *C. Introducing Emissivity as a Novel Feature*

Emissivity is a key parameter used to estimate surface emission and reflection in RTMs. We previously discussed that the CRTM simulation was affected by the uncertainty of surface emissivity, particularly for S-bands. In CRTM, ocean surface emissivity is band-dependent and calculated using WS and direction, SST, and salinity, with several geophysical angles of the sensor based on the complex FAST microwave emissivity model [25], [26], [27], which are related to ocean wave facets and foam distribution. However, WS, SST, and geophysical parameters of the sensor alone may not represent the emissivity of the 22 bands to accurately predict BTs in the FCDN_CRTM model. Therefore, we performed a sensitivity experiment with emissivity as a novel feature to determine model accuracy. We retained the three-model design and supplemented the model with the emissivity of the 22 bands as new input features, which were precalculated using CRTM. Fig. 6 compares the results of the models with or without emissivity. The mean biases of FCDN_CRTM minus CRTM were comparable, but SDs decreased for all S-bands when emissivity was included as a novel feature.

## V. Model Evaluation

Based on the sensitivity experiments, a final design of the FCDN_CRTM was implemented using 5 M data samples for training, using the three-model design instead of one, and introducing surface emissivity as a novel feature for the S-band

TABLE II
MEAN BIAS AND SD OF FCDN_CRTM MINUS CRTM FOR TEST DATA FOR ONE-MODEL AND THREE-MODEL

| Channel | One-model | | Three-model | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| CH1 | –0.012 | 0.144 | –0.022 | 0.092 |
| CH2 | –0.018 | 0.129 | –0.015 | 0.070 |
| CH3 | 0.011 | 0.102 | 0.001 | 0.071 |
| CH4 | –0.001 | 0.088 | –0.003 | 0.074 |
| CH5 | –0.006 | 0.078 | –0.008 | 0.071 |
| CH6 | –0.001 | 0.077 | –0.012 | 0.057 |
| CH7 | 0.012 | 0.079 | –0.005 | 0.055 |
| CH8 | 0.017 | 0.079 | 0.002 | 0.052 |
| CH9 | 0.016 | 0.084 | 0.005 | 0.054 |
| CH10 | 0.011 | 0.083 | 0.0 | 0.055 |
| CH11 | 0.009 | 0.081 | –0.002 | 0.053 |
| CH12 | 0.012 | 0.085 | –0.006 | 0.054 |
| CH13 | 0.017 | 0.091 | 0.004 | 0.060 |
| CH14 | 0.020 | 0.094 | –0.002 | 0.064 |
| CH15 | 0.020 | 0.104 | –0.005 | 0.069 |
| CH16 | –0.026 | 0.148 | –0.025 | 0.100 |
| CH17 | 0.009 | 0.151 | –0.006 | 0.114 |
| CH18 | –0.002 | 0.145 | 0.004 | 0.106 |
| CH19 | –0.012 | 0.134 | 0.004 | 0.091 |
| CH20 | –0.019 | 0.138 | 0.002 | 0.090 |
| CH21 | –0.024 | 0.144 | 0.001 | 0.094 |
| CH22 | –0.025 | 0.166 | 0.002 | 0.108 |

model. In this section, we present a detailed evaluation of the FCDN_CRTM accuracy, efficiency, and robustness (namely, generalization). Since the one-model design is more computationally efficient despite altered accuracy as it does not require an additional computation for emissivity, it can be used for high efficiency applications, such as the evaluation of model Jacobian, which will be discussed in next section.

### A. *Model Accuracy for Test Data*

Table II shows the mean bias and SD of FCDN_CRTM minus CRTM for test data after FCDN_CRTM reconstruction for both one-model and three-model. The mean biases were within 0.025 K for all bands and were comparable between the two models. However, the SDs for three-model and were smaller by 0.03–0.06 K than one-model. The SD values were between 0.05 K and 0.11 K. The T-band performed the best, (0.05–0.07 K), followed by H-band (0.09–0.108 K) and S-band (0.07–0.114 K). The highest accuracy in T-bands is mainly due to their less sensitivity to complicated surface emissivity and low cloud, as discussed in last section. On the contrary, the SDs are slightly worse for more surface-sensitive S-bands. For H-bands, the SDs are partially affected by the more uncertainty in water vapor contents. Nevertheless, the FCDN_CRTM showed much higher accuracy after reconstruction for all three-band groups than the initial test. In the rest of this section, the three-model was used to evaluate the FCDN_CRTM performance.

Fig. 7 illustrates the global distribution of FCDN_CRTM minus CRTM and corresponding histograms for the smallest and largest SDs among the 22 bands, i.e., for bands 8 and 17, respectively. The global distributions were quite uniform for all

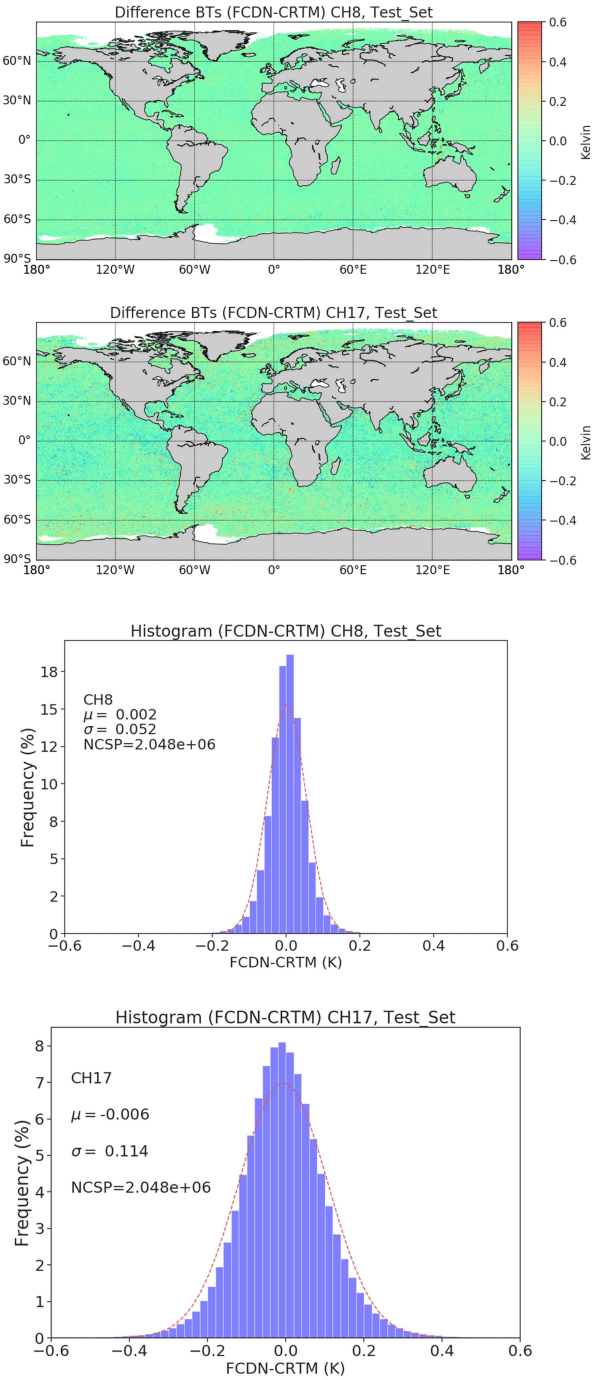


Fig. 7. Global distribution of FCDN_CRTM minus CRTM (upper two panels) and corresponding histograms (bottom two panels) for the smallest and largest SDs—Band 8 and Band 17.

bands, with most of the values close to 0. The histograms were nearly Gaussian, with slight noise for band 17, indicating that the model did not have significant outliers. The global distributions for other bands were similar to those of bands 8 or 17. Overall, the FCDM_CRTM was highly accurate at predicting BTs and emulating CRTM simulations.

### B. *Model Robustness*

To determine the stability of the model, the FCDN_CRTM was used to predict ATMS BTs for six days independent from

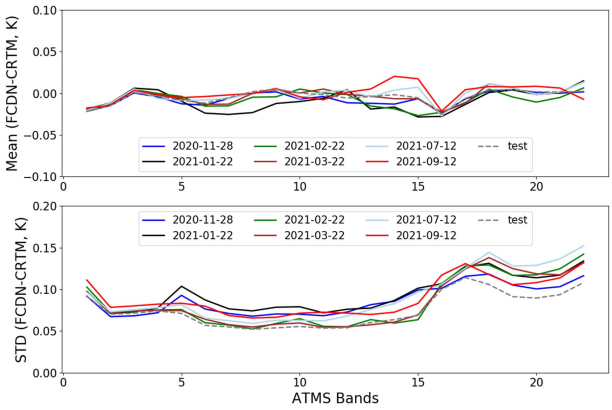


Fig. 8. FCDN_CRTM simulation results for six days. The test data are also overlaid in the figure to assist comparison of the accuracy.

the training set, including 28 Nov. 2020, 22 Jan. 2021, 22 Feb. 2021, 22 Mar. 2021, 12 Jul. 2021, and 12 Sep. 2021, which covered about 1 year after the end of the training period.

Fig. 8 shows the results of FCDN_CRTM predictions for the six days. The test data are also overlaid in the figure to assist the comparison of prediction accuracy. The mean biases of FCDN_CRTM minus CRTM were approximately 0 and comparable to test data for all six days. In contrast, the SDs were comparable or slightly larger than those of test data, with amplitudes within 0.02 K for S-bands, 0.03 K for T-bands, and 0.05 K for H-bands. The smallest SD was 0.052 K for 22 Feb. 2021 (band 8) and the largest SD was 0.152 K for 12 Jul. 2021 (band 22), which were separated from the training period by eight months.

Fig. 9 illustrates the global distribution of FCDN_CRTM minus CRTM and corresponding histograms for the smallest (band 8) and largest SDs (band 22). Similar to test data, the global distribution was uniform and approximately 0 for the smallest SD. Although some noise exists in band 22, such as at high latitudes of the Southern Hemisphere and in some coastal regions, maybe due to the higher noise in this band, and cloud contamination and more uncertainty of the water vapor contents in these areas, most of the global distribution is uniform and close to 0, and the corresponding histograms exhibit a Gaussian distribution. There were approximately 50 outliers out of 2-M samples, with mean biases greater than 2 K, which necessitate further investigation. Overall, the FCDN_CRTM results for the six days indicated that the model was robust and could accurately predict BTs for all ATMS bands beyond the training period.

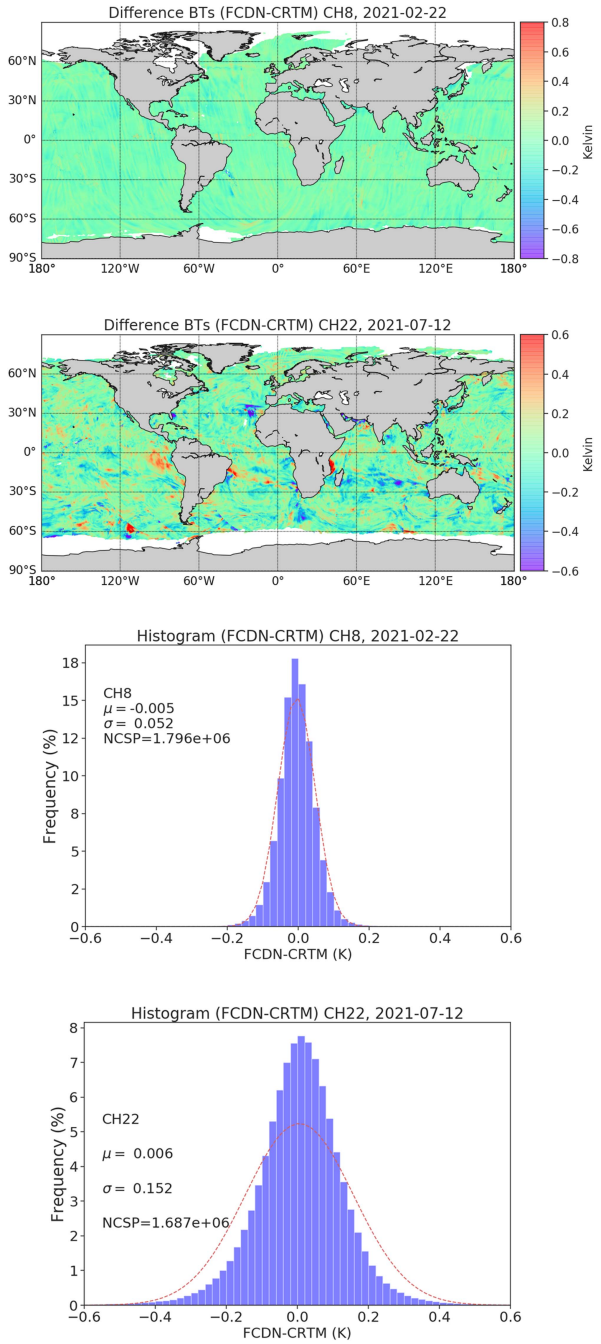


Fig. 9. Global distribution of FCDN_CRTM minus CRTM (upper two panels) and corresponding histograms (bottom two panels) for the smallest SD (22 Feb. 2021, Band 8) and the largest SDs (12 Jul. 2021, Band 22) among the independent data testing.

### C. Comparison With ATMS Measurements

Ultimately the objective of the FCDN_CRTM simulate the observed BTs of ATMS, matching or exceeding the performance of CRTM.

Table III shows the global mean bias and SD of ATMS minus CRTM and ATMS minus FCDN_CRTM for 22 Feb. 2021, for all 22 bands. Both the mean bias and SD between FCDN_CRTM and CRTM were consistent for the temperature and water vapor sounding bands as well as the surface bands. The differences in mean bias and SD between FCDN_CRTM and CRTM were only up to 0.015 and 0.014 K, respectively. This further suggested that FCDN_CRTM results were close to CRTM results, when compared with ATMS SDR. As indicated in Section III, all mean biases and SDs for S- and H-bands were slightly larger than those for T-bands, suggesting that these bands were more sensitive to surface emissivity and clouds.

Fig. 10 demonstrates that the global distributions of ATMS minus FCDN_CRTM and ATMS minus CRTM were comparable for bands 3 and 5 on 22 Feb. 2021. Although the global distribution of the BT differences is quite different between

TABLE III
MEAN BIAS AND SD OF ATMS MINUS CRTM AND ATMS MINUS FCDN_CRTM ON 22 FEB. 2021

| Channel | ATMS-CRTM | | ATMS-FCDN_CRTM | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| CH1 | 4.403 | 7.323 | 4.425 | 7.321 |
| CH2 | 5.817 | 9.574 | 5.830 | 9.575 |
| CH3 | 2.287 | 4.915 | 2.283 | 4.915 |
| CH4 | 0.339 | 2.740 | 0.339 | 2.739 |
| CH5 | −0.369 | 0.999 | −0.365 | 1.000 |
| CH6 | −0.203 | 0.442 | −0.187 | 0.449 |
| CH7 | −0.410 | 0.245 | −0.395 | 0.252 |
| CH8 | −0.629 | 0.218 | −0.624 | 0.225 |
| CH9 | −0.754 | 0.274 | −0.749 | 0.285 |
| CH10 | −0.444 | 0.316 | 0.449 | 0.324 |
| CH11 | −0.282 | 0.425 | −0.282 | 0.431 |
| CH12 | −0.130 | 0.459 | −0.129 | 0.463 |
| CH13 | −0.199 | 0.684 | −0.183 | 0.687 |
| CH14 | −0.097 | 0.980 | −0.078 | 0.982 |
| CH15 | −0.113 | 1.615 | −0.087 | 1.617 |
| CH16 | 4.596 | 7.694 | 4.619 | 7.684 |
| CH17 | −0.686 | 8.147 | −0.675 | 8.139 |
| CH18 | −1.980 | 5.543 | −1.985 | 5.542 |
| CH19 | −1.407 | 4.259 | −1.402 | 4.258 |
| CH20 | −1.036 | 3.334 | −1.026 | 3.337 |
| CH21 | −0.658 | 2.595 | −0.653 | 2.604 |
| CH22 | −0.067 | 2.201 | −0.074 | 2.212 |

bands 3 and 5, due to different band group, for the same band, the ATMS minus FCDN_CRTM are well consistent with ATMS minus CRTM. This further verified the accuracy of FCDN_CRTM with respect to the reference CRTM.

### D. Model Efficiency

One important benefit of AI is increased efficiency in performing calculations. For CRTM on a Linux server running (24 3.4 GHz Intel Xeon CPUs), the simulation required approximately 900 s to simulation 2 M ATMS clear-sky scenes. In contrast, the total processing time for the FCDN_CRTM with multiple CPUs was only 8 s for the one-model design and 17 s for the three-model design, which were 112 times and 53 times faster than that CRTM, respectively. Liang and Liu [14] reported that the increase in efficiency for FCDN_CRTM was primarily due to the DNN architecture, rather than increasing the number of CPU resources. Further improvement in efficiency is expected for all-sky radiance simulations, where CRTM scattering calculations require more CPU time compared with clear-sky scenes.

## VI. MODEL JACOBIAN

Variational data assimilation and remote sensing algorithms not only require accurately simulated BTs but also PDs of the BTs with respect to atmospheric and surface geophysical parameters, which are collectively called Jacobians. Jacobians are required to estimate the gradient of the cost function during minimization in data assimilation [19]. The FCDN_CRTM can automatically produce model Jacobians during model training. Therefore, the model Jacobians can be obtained as a direct output

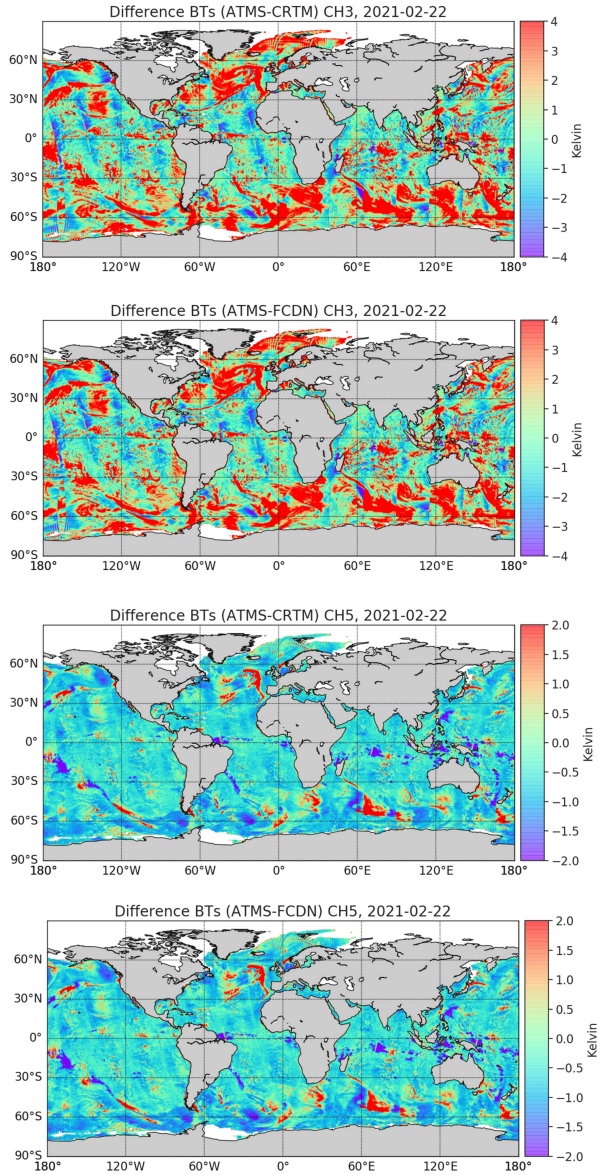


Fig. 10. Global distribution of mean bias of ATMS minus FCDN_CRTM and ATMS minus CRTM for ATMS band 3 (upper two panels) and band 5 (bottom two panels) on 22 Feb. 2021.

of the FCDN_CRTM, as opposed to requiring a separate ML model, or the need to write more complex tangent linear and adjoint methods as implemented in CRTM. Through evaluation of the FCDN_CRTM Jacobian, the possibility of using an RTM emulator in data assimilation applications can be assessed, and the simulated BTs can be further validated (i.e., explainable AI).

### A. Model Jacobian Equation

Assume that BTs $y$ were mapped from input $x$ using the FCDN_CRTM model $H$ (2) and the input $x$ was standardized before it was input to the model (1), then the PD of $y$ to $x$ can be represented as (3), based on the derivative chain rule. A detailed proof of (3) and the later (4) was showed at the Appendix A

$$y = H(x) \tag{2}$$

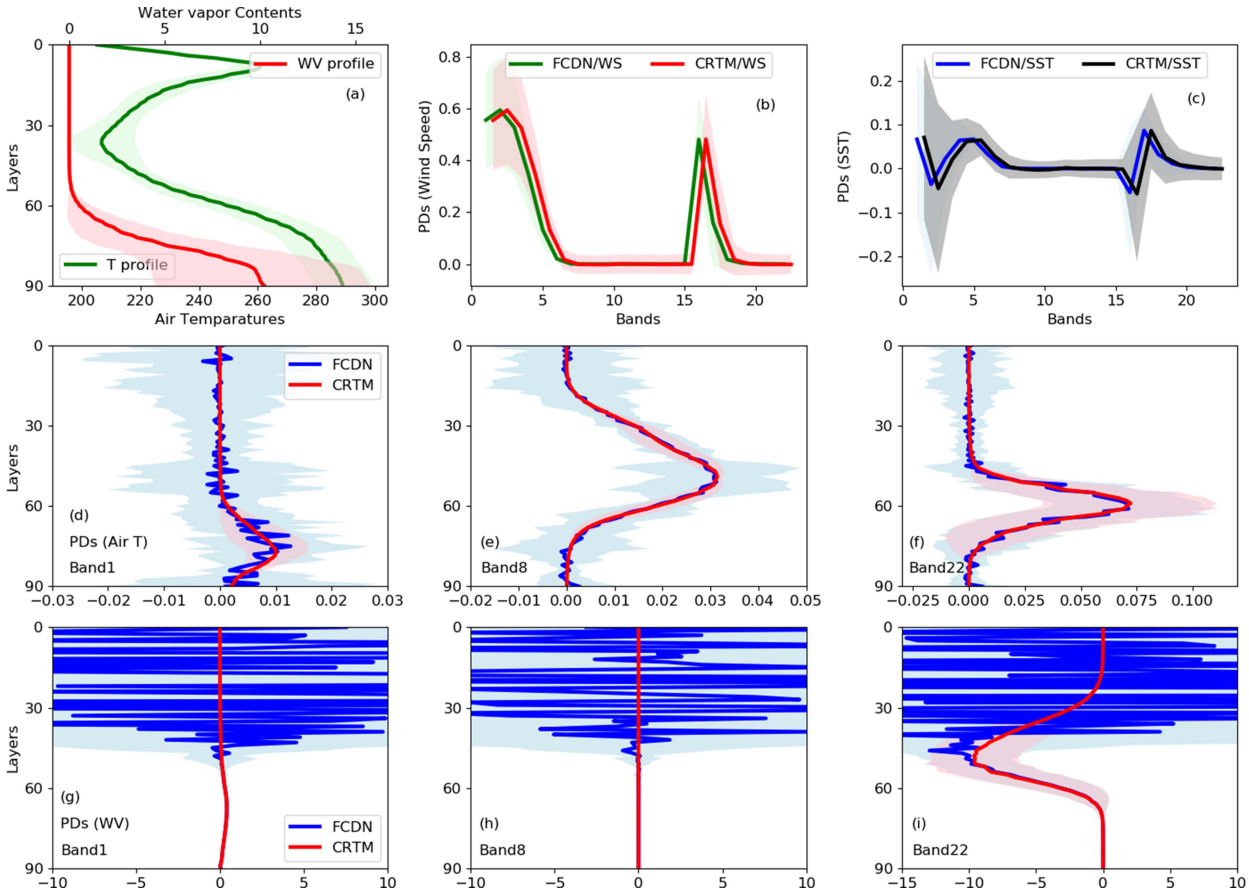


Fig. 11. Summary of the FCDN_CRTM Jacobian compared to CRTM. (a) Profiles of air temperature and water vapor contents. (b) Partial derivatives with respect to wind speed. (c) Partial derivatives with respect to sea surface temperature. (d–f) Partial derivatives with respect to air temperature profile for band 1, 8, and 22. (g–i) Partial derivatives with respect to water vapor content profile for band 1, 8, and 22. The curves are the mean of all samples with $\pm\sigma$ in the filled areas. The curves (b) and (c) are slightly shifted in $x$-axis to clearly distinguish partial derivatives between CRTM and FCDN_CRTM.

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial \bar{x}} \frac{\partial \bar{x}}{\partial x} = \frac{1}{\sigma_x} \frac{\partial y}{\partial \bar{x}}. \tag{3}$$

Thereafter, $\frac{\partial y}{\partial \bar{x}}$ can be automatically calculated from the FCDN_CRTM model (Appendix B). As input $x$ includes 188 features for 22 output BTs, we can obtain a $188 \times 22$ Jacobian matrix for each sample. However, producing all Jacobian matrices for 2 M test data samples requires large memory resources. Therefore, we randomly selected 5000 profiles to evaluate the FCDN_CRTM Jacobian in comparison to CRTM Jacobians.

Although the three-model design exhibited more accuracy than the one-model design in BT space, the efficiency was reduced by a factor of two. Additionally, emissivity reduced the efficiency of the FCDN_CRTM model at both training and testing, as the model incurred an additional computational cost. For data assimilation, with high demand for model efficiency, it is necessary to balance the efficiency and accuracy of the model. Furthermore, an offline analysis indicated that there was not significantly different in Jacobian between one- and three-model. Therefore, in this section, we first selected the more efficient one-model design to determine the model Jacobian, then we further improved model accuracy based on the evaluation of the Jacobians. For the FCDN_CRTM clear-sky model, the PDs of the 22 ATMS bands were considered with respect to four parameters, including surface WS, SST, atmospheric temperature at 91 layers (AT), and atmospheric water vapor content at 91 layers (WV).

### B. Evaluation of Model Jacobian

Fig. 11 summarizes the FCDN_CRTM Jacobian compared to that of the CRTM. The profiles of temperature and water vapor are shown in panel (a) to assist with the model Jacobian evaluation. The curves in Panels (b) $\partial y/\partial(x_{\mathrm{ws}})$ and (c) $\partial y/\partial(x_{\mathrm{sst}})$ are slightly shifted along the $x$-axis to clearly distinguish the PDs between CRTM and FCDN_CRTM. Both curves reveal consistency between CRTM and FCDN_CRTM for all bands. Moreover, most PDs for T- and H-bands were approximately 0 due to insensitivity of these bands to the surface. All $\partial y/\partial(x_{\mathrm{ws}})$ were positive for S-bands, indicating that BTs increased with increasing WS.

Panels (d)–(i) show the $\partial y/\partial(x_{at})$ and $\partial y/\partial(x_{wv})$. Bands 1, 8, and 22 were selected to determine the accuracy of the Jacobians for the three band groups. The PDs for other bands are not shown as their PDs are similar to those of the selected bands. The curves of $\partial y/\partial(x_{at})$ were mostly consistent with CRTM, with slight oscillations in the vertical for all bands. The oscillation for the S-band was the largest, likely due to the large uncertainty of the surface emissivity. The oscillations became smaller for H-bands and were minimal for T-bands.

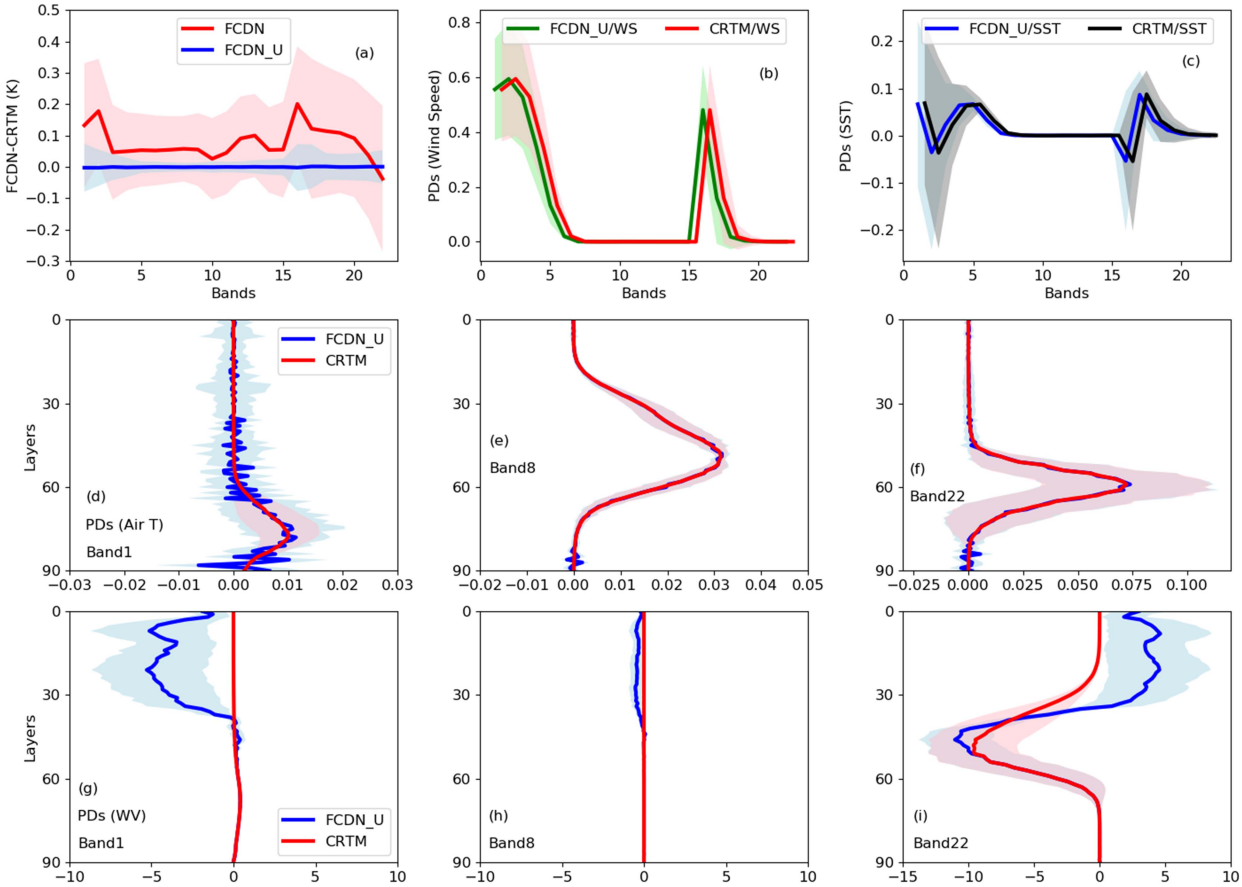


Fig. 12. Summary of the updated FCDN_CRTM Jacobian compared to CRTM. Same as Fig. 11, but using updated FCDN_CRTM. (a) BT differences between the original or updated FCDN_CRTM and CRTM. (b–i) same as Fig. 11(b)–(i), but for FCDN_U.

However, the uncertainties of PDs were considerably large compared to those of the CRTM for all bands, particularly for S-bands, followed by H-bands and T-bands, which was similar to the pattern of PD oscillation. As shown in panel (g)–(i), the $\partial y/\partial(x_{wv})$ were consistent with CRTM below layer 40; they were close to 0 for S-bands and T-bands and showed a notable increase with CRTM for H-bands. However, unrealistic oscillations and noise prevailed above layer 50 (below 200 hPa) for all bands. The oscillation reached several tens and the uncertainty went up to several hundreds compared to near 0 for CRTM PDs. This suggests that the model Jacobian could not be used for data assimilation, and also indicated that the FCDN_CRTM requires further improvement.

### C. Improving Model Jacobian

Evaluation of the FCDN_CRTM Jacobian revealed that two issues needed to be addressed—the first issue was the large oscillation and uncertainty in $\partial y/\partial(x_{wv})$ above layer 50 and second that was the large uncertainty in $\partial y/\partial(x_{at})$, particularly for S-bands. The first issue was likely due to the standardization of layer water vapor content when used as input for FCDN_CRTM. As shown in Fig. 11(a), both mean and SD in the water vapor for the layers above 50 were extremely close to 0. Any biases or random error in PDs generated from the FCDN_CRTM were unrealistically enlarged when they were divided by the SD of the water vapor content (3). Therefore, using absolute water vapor content instead of their standardization as input may reduce the large oscillations and uncertainty of PDs above layer 50. Subsequently, to minimize the uncertainty in PDs of atmospheric temperature, we may eliminate the BN algorithm from the model. Based on the Tensorflow document[1], BN alters target-source independence as it normalizes data across the batch dimension. However, removing BN may reduce model accuracy or generalization [14], and using absolute water vapor content instead of their standardization as model inputs may further affect model accuracy. To compensate the possible loss, we used the standardization of output, the BTs at 22 ATMS bands, and modified (3)–(4), as follows:

$$\frac{\partial y}{\partial x} = \frac{\sigma_y}{\sigma_x}\frac{\partial \bar{y}}{\partial \bar{x}} \quad \text{if } x \text{ was not WV}$$
$$\frac{\partial y}{\partial x} = \sigma_y \frac{\partial \bar{y}}{\partial \bar{x}} \quad \text{if } x \text{ was WV}. \tag{4}$$

As demonstrated in the previous section, using surface emissivity as a model input could improve model accuracy for S-bands. However, extra computing time and other resources are required to calculate emissivity, which further lowers model efficiency. Therefore, emissivity was not used for Jacobian evaluation, but it may be a good option for other applications. Fig. 12

[1][Online]. Available: https://www.tensorflow.org/guide/advanced_autodiff

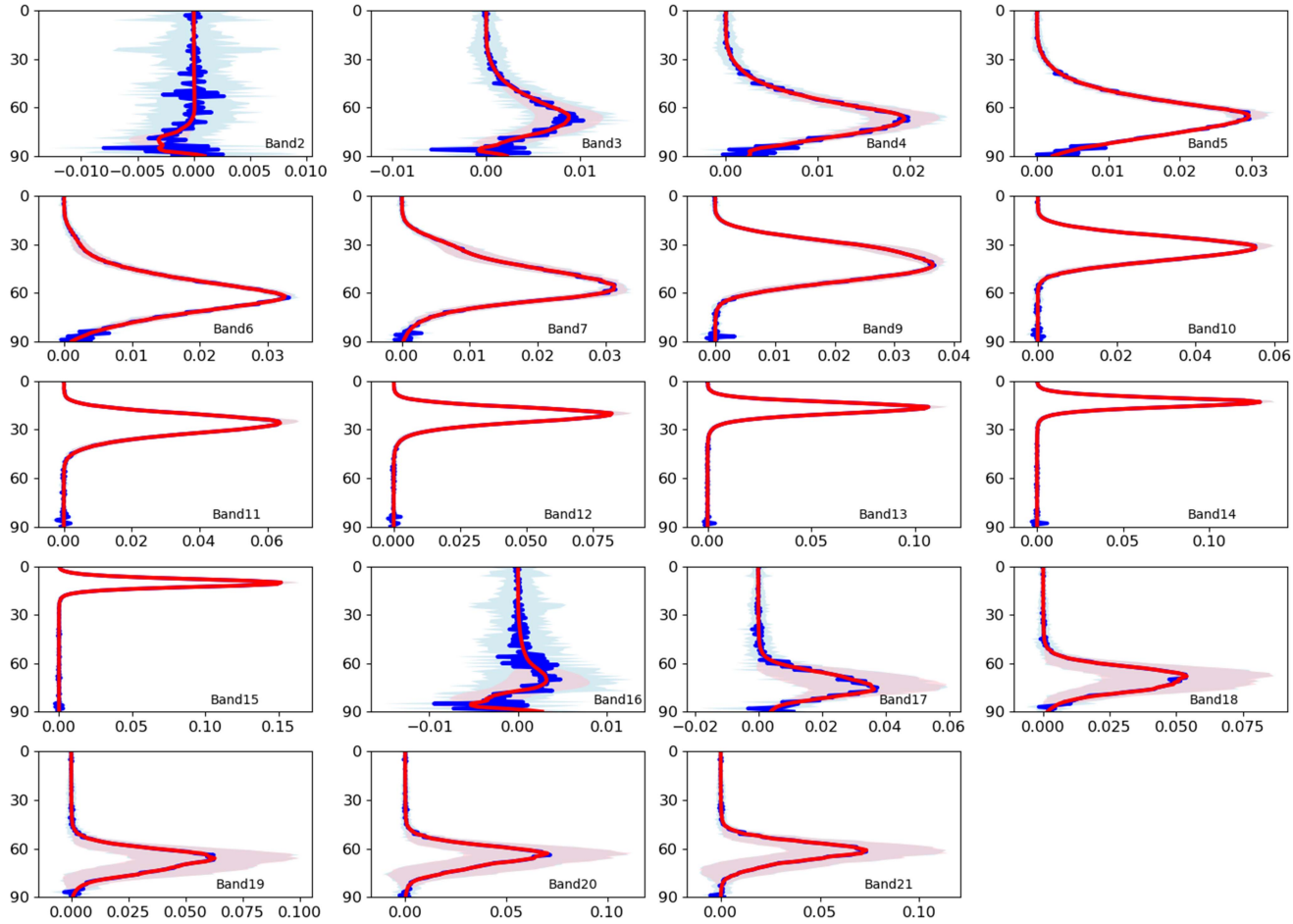


Fig. 13. PDs to 91-layer air temperatures between CRTM and FCDN_CRTM for the bands not showed in Fig. 12.

shows the updated FCDN_CRTM (FCDN_U) BTs and Jacobian compared to CRTM. Both mean and SD for the FCDN_U BTs were closer to 0 and much smaller than those of FCDN_CRTM for all bands, indicating that using standardization of output further improved model accuracy Although the uncertainties of the $\partial y/\partial(x_{\text{ws}})$ and $\partial y/\partial(x_{\text{sst}})$ in S-bands remained unchanged, those in T- and H-bands decreased and both were close to 0. Additionally, the uncertainties for $\partial y/\partial(x_{at})$ were significantly reduced for all bands, particularly for T- and H-bands, which were closer to the CRTM Jacobians. Furthermore, the unrealistic oscillation and uncertainty in the $\partial y/\partial(x_{wv})$ above layer 50 were significantly reduced after eliminating the standardization of water vapor content. However, the oscillation and uncertainty still remained quite large and deviated from the CRTM Jacobian, particularly for S- and T-bands, in which most CRTM PDs were close to 0.

Figs. 13 and 14 show $\partial y/\partial(x_{at})$ and $\partial y/\partial(x_{wv})$ of the other 19 bands, which are not shown in Fig. 12. We observed that for the same band group, the patterns of the PD curves were similar. Similar to Fig. 12(g)–(i), Fig. 14 revealed the large deviations from the CRTM Jacobian above layer 50 for all bands. Some of these issues have been addressed in previous studies [28], [29], [30]. For an NN model, the data on pressure level are partially lost when the model inputs map its outputs using a nonlinear projection, and the shape for $\partial y/\partial(x_{wv})$ is more pronounced. This issue needs further investigation. However, many data assimilation systems do not use absolute water vapor content in the state vector. For instance, the NOAA MiRS uses the logarithm of the water vapor content to avoid negative water vapor in the retrieval when projecting back to absolute space. Therefore, we assessed the Jacobian for the logarithm of water vapor content, $\partial y/\partial(\log(x_{wv}))$, which is equal to $wv * \partial y/\partial(\log(x_{wv}))$. Fig. 15, by demonstrating $\partial y/\partial(\log(x_{wv}))$ for all bands, revealed that all oscillations above layer 50 were removed, making the PDs consistent with those of CRTM in these layers. As the absolute water vapor contents above layer 50 are close to 0, the $\partial y/\partial(\log(x_{wv}))$, obtained by multiplying absolute WV, were certainly close to 0 in these layers, making the curves consistent with those of CRTM. However, as shown in the Fig. 15, $\partial y/\partial(\log(x_{wv}))$ increased the oscillation and noise near the surface. This was because the absolute water vapor content near the surface was several orders of magnitude larger than that at higher layers, indicating that any bias and noise in $\partial y/\partial(x_{wv})$ would be enlarged after multiplying it with the large magnitude of the WV near the surface to obtain $\partial y/\partial(\log(x_{wv}))$. Despite these small artifacts, further investigations are planned to assess application of FCDN_CRTM in MiRS.

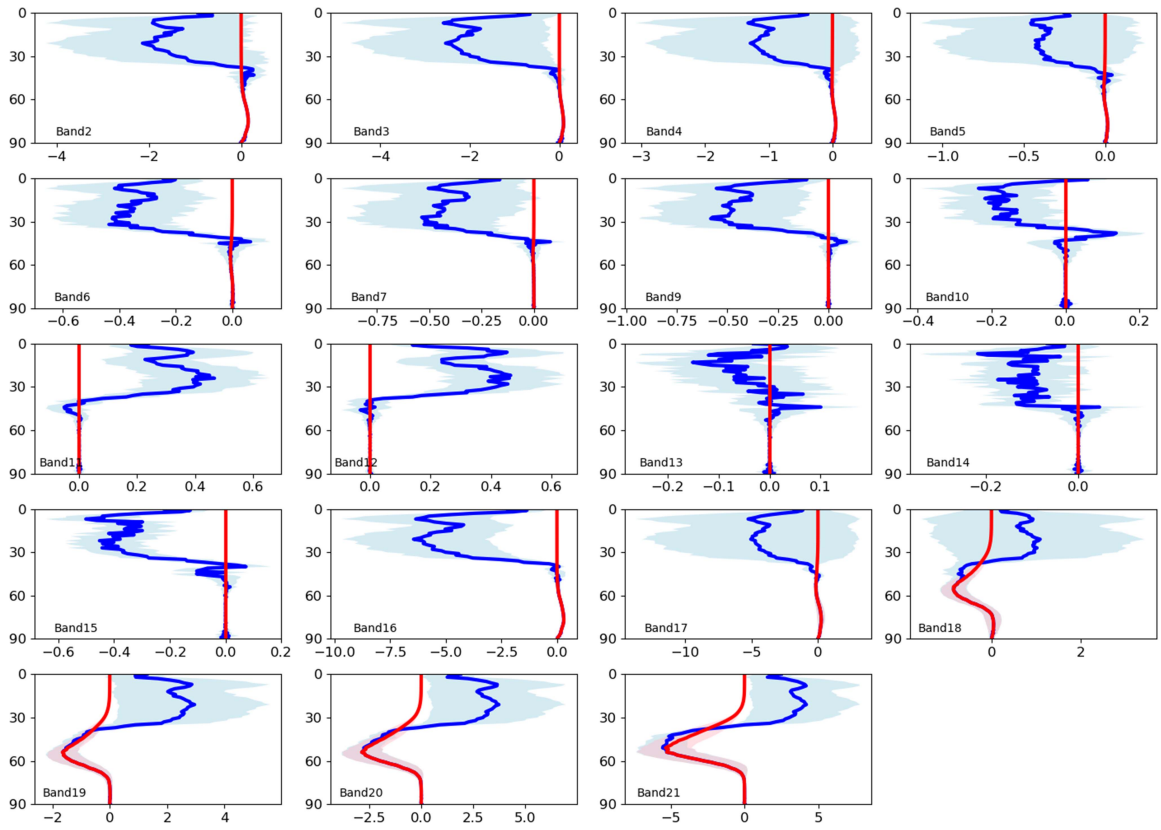


Fig. 14. PDs for water vapor at 91 pressure layers between CRTM (red) and FCDN_CRTM (blue) for the bands not showed in Fig. 12 (bands 2–7 and 9–21).

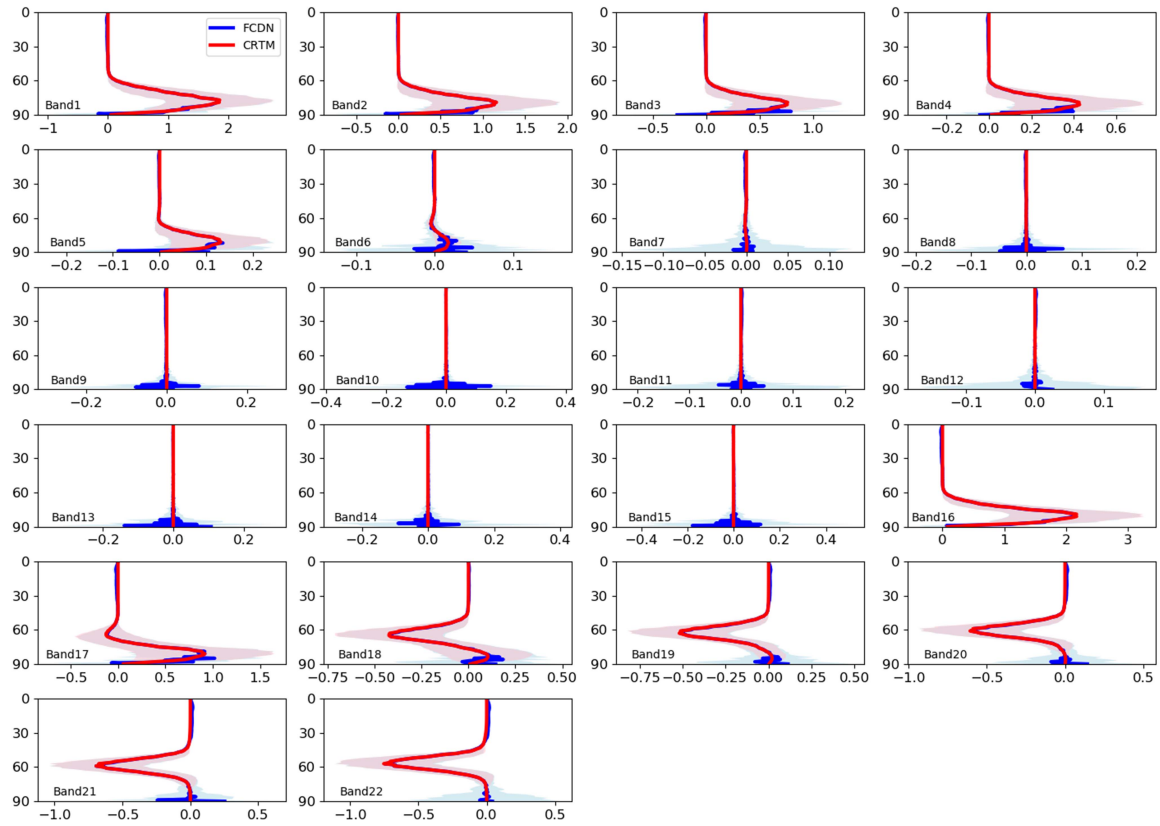


Fig. 15. Same as in Fig. 14, but showing PDs for the logarithm of water vapor for all ATMS bands.

## VII. Conclusion

In this article, we proposed an initial FCDN_CRTM to explore the efficiency and accuracy of emulating the RTM of ATMS BTs in the 22 microwave bands for clear-sky, ocean scenes. The input features were extracted from ECMWF analysis, CMC SST, and ATMS SDR products, while the CRTM simulated BTs were used as model reference labels. The FCDN_CRTM training dataset spanned 12 days over 1 year to encompass varying atmospheric and surface conditions. An additional dataset of six days, encompassing 1 year after the end of the training period, was used to determine the stability and generalization of the model. Simultaneously, the Jacobians of the 22 bands were evaluated with respect to surface state and atmosphere profiles to further validate the BT simulations, and explore the use of the model in data assimilation and physical inversion algorithms.

The initial FCDN_CRTM was fine-tuned through three sensitivity experiments with respect to sample-size determination, model separation, and introduction of novel features, toward improving the accuracy of the model. The reconstructed model was well evaluated on model accuracy, efficiency, and long-term stability. The FCDN_CRTM simulations were in good agreement with those of CRTM. The mean bias and SD were within 0.025 K and 0.05–0.15 K, respectively, for all ATMS bands. The SD for T-bands were the smallest, followed by H- and S-bands. The global distributions of FCDN_CRTM minus CRTM were uniform and their corresponding histogram distributions were Gaussian. The high accuracy was consistent when the FCDN_CRTM was applied to data outside of the training period. Upon comparison with real ATMS observations, both the mean bias and SD of FCDN_CRTM and CRTM were consistent with each other for all bands. Furthermore, the FCDN_CRTM simulations were 112 times (one-model) and 53 times (three-model) faster compared to CRTM simulations.

To determine the possible applications of the model in data assimilation and further improve its accuracy, we evaluated the FCDN_CRTM Jacobian. The initial evaluation identified two main issues in the FCDN_CRTM Jacobian associated with vertical oscillations and large uncertainty in $\partial y/\partial(x_{wv})$ above layer 50, and large uncertainty in $\partial y/\partial(x_{at})$. An updated FCDN_CRTM, including training the water vapor in log space, significantly minimized both issues and further improved model accuracy. The results and performance both in terms of accuracy and efficiency, indicate that the FCDN_CRTM may be tested as a forward and Jacobian operator in data assimilation.

## Appendix A
## Proof of (3) and (4)

In this appendix, we give a detailed derivation for (3) and (4). When we standardize the input data $x$, (1) can be rewritten as

$$\bar{x} = \frac{x - \mu_x}{\sigma_x}. \tag{A1}$$

$$\text{We can get} \quad \frac{\partial \bar{x}}{\partial x} = \frac{1}{\sigma_x}. \tag{A2}$$

Based on the chain rule, (3) can be proved as

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial \bar{x}} \frac{\partial \bar{x}}{\partial x} = \frac{1}{\sigma_x} \frac{\partial y}{\partial \bar{x}}. \tag{A3}$$

$$\text{If y is also standardized as } \bar{y} = \frac{y - \mu_y}{\sigma_y} \tag{A4}$$

$$\text{then } \frac{\partial \bar{y}}{\partial y} = \frac{1}{\sigma_y}. \tag{A5}$$

So (4) can be obtained as

$$\frac{\partial y}{\partial x} = \frac{\sigma_y}{\partial \bar{y}} \frac{\partial \bar{y}}{\partial \bar{x}} \frac{\partial \bar{x}}{\partial x} = \sigma_y \frac{\partial \bar{y}}{\partial \bar{x}} \frac{1}{\sigma_x} = \frac{\sigma_y}{\sigma_x} \frac{\partial \bar{y}}{\partial \bar{x}} \ (\text{if } x \text{ was not WV}). \tag{A6}$$

However, in the updated model FCDN_U, we did not standardize WV, so

$$\frac{\partial y}{\partial x} = \frac{\sigma_y}{\partial \bar{y}} \frac{\partial \bar{y}}{\partial x} = \sigma_y \frac{\partial \bar{y}}{\partial x} \ (\text{if } x \text{ was not WV}). \tag{A7}$$

## Appendix B
## Calculation of FCDN_CRTM Jacobian

For general multilayer networks with $N-1$ nonlinear layers and one linear output layer and without BN included in all layers, the nonlinear layer $i$ can be defined as follows:

$$Z_i = W_{i-1}\, A_{i-1} + B_i, \ (0 < i < N) \tag{A8}$$

$$A_i = G\,(Z_i) \tag{A9}$$

$$A_0 = X. \tag{A10}$$

Here, $W_{i-1}$ is the weight matrix associated with layers $i-1$ and $i$, and $B_i$ is the bias matrix. $A_i$ and $Z_i$ are the outputs in layer $i$ with and without nonlinear activation function G constraint. $A_i$ will be input data matrix $X$ in the input layer when $i = 0$. The output layer $N$ calculates output $Y$ as

$$Y \ = A_N \ = W_{N-1}\, A_{N-1} + B_N. \tag{A11}$$

Thus, the model Jacobian matrix can be calculated as

$$\begin{aligned} \frac{\partial Y}{\partial X} &= W_{N-1} \frac{\partial A_{N-1}}{\partial X} \\ &= W_{N-1}\, G'\,(Z_{N-1})\, W_{N-2} \frac{\partial A_{N-2}}{\partial X} \\ &= W_{N-1}\, G'\,(Z_{N-1})\, W_{N-2} G'\,(Z_{N-2}) \\ &\quad W_{N-3} \ldots\ G'\,(Z_1)\, W_0 \\ &= W_{N-1} \prod_{i\,=\,N-2}^{0} G'\,(Z_{i+1})\, W_i. \end{aligned} \tag{A12}$$

Equation (A12) calculates Jacobian for general multilayer networks. The FCDN_CRTM model include three nonlinear layers, thus $N = 4$. Since ReLU was used as an activation function in FCDN_CRTM, $G'(Z_i)$ is a matrix containing only 0 and 1. The algorithm has been included in the FCDN_CRTM model to calculate the PDs of output BTs to each input variable.

## Acknowledgment

The CRTM was provided by the CRTM team at NOAA/STAR. The authors thank Christopher Grassotti and the NOAA/STAR

MiRS team, and Changyong Cao from the NOAA/STAR VIIRS SDR team for technical discussions and suggestions. The views, opinions, and findings contained in this report are those of the authors and should not be construed as an official NOAA or U.S. Government position, policy, or decision.

## REFERENCES


[1] X. Liang , A. Ignatov, and Y. Kihai, "Implementation of the community radiative transfer model (CRTM) in advanced clear-sky processor for oceans (ACSPO) and validation against nighttime AVHRR radiances," *J. Geophys. Res.*, vol. 114, 2009, Art. no. D06112, doi: 10.1029/2008JD010960.

[2] Q. Liu and S. Boukabara, "Community radiation transfer model (CRTM) applications in supporting the suomi national polar-orbiting partnership (SNPP) mission validation and verification," *Remote Sens. Environ.*, vol. 140, pp. 744–754, 2014.

[3] R. Saunders, M. Matricardi, and P. Brunel, "An improved fast radiative transfer model for assimilation of satellite radiances observations," *Quart. J. Roy. Meteorol. Soc.*, vol. 125, pp. 1407–1425, 1999, doi: 10.1256/smsqj.55614.

[4] R. Saunders et al., "A comparison of radiative transfer models for simulating atmospheric infrared sounder (AIRS) radiance," *J. Geophys. Res.*, vol. 112, 2007, Art. no. D01S90, doi: 10.1029/2006JD007088.

[5] F. Z. Weng, X. W. Yu, Y. H. Duan, J. Yang, and J. J. Wang, "Advanced radiative transfer modeling system (ARMS): A new-generation satellite observation operator developed for numerical weather prediction and remote sensing applications," *Adv. Atmos. Sci.*, vol. 37, no. 2, pp. 131–136, 2020, doi: 10.1007/s00376-019-9170-2.

[6] C. K. Sønderby et al., "Metnet: A neural weather model for precipitation forecasting," Mar. 2020, *arXiv:2003.12140*.

[7] J. A. Weyn, D. R. Durran, R. Caruana, and N. Cresswell-Clay, "Subseasonal forecasting with a large ensemble of deep-learning weather prediction models," *Earth Space Sci. Open Arch.*, vol. 55, no. 7, 2021, Art. no. e2021MS002502, doi: 10.1002/essoar.10506235.1.

[8] E. S. Maddy and S. A. Boukabara, "MIIDAPS-AI: An explainable machine-learning algorithm for infrared and microwave remote sensing and data assimilation preprocessing—Application to LEO and GEO sensors," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8566–8576, 2021, doi: 10.1109/JSTARS.2021.3104389.

[9] F. Chevalier, F. Chéruy, N. A. Scott, and A. Chédin, "A neural network approach for a fast and accurate computation of a longwave radiative budget," *J. Appl. Meteorol.*, vol. 37, pp. 1385–1397, 1998.

[10] N. Westing, B. Borghetti, and K. C. Gross, "Fast and effective techniques for LWIR radiative transfer modeling: A dimension-reduction approach," *Remote Sens.*, vol. 11, 2019, Art. no. 1866, doi: 10.3390/rs11161866.

[11] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, Y. T. Hou, S. J. Lord, and A. A. Belochitski, "Accurate and fast neural network emulations 395 of model radiation for the NCEP coupled climate forecast system: Climate simulations and seasonal predictions," *Monthly Weather Rev.*, vol. 138, pp. 1822–1842, 2009, doi: 10.1175/2009MWR3149.1.

[12] B. D. Bue et al., "Neural network radiative transfer for imaging spectroscopy," *Atmos. Meas. Techn.*, vol. 12, no. 4, pp. 2567–2578, 2019, doi: 10.5194/amt-12-2567-2019.

[13] S. Mishra and R. Molinaro, "Physics informed neural networks for simulating radiative transfer," *J. Quantitative Spectrosc. Radiat. Transfer*, vol. 270, 2021, Art. no. 107705, doi: 10.1016/j.jqsrt.2021.107705.

[14] X. Liang and Q. Liu, "Applying deep learning to clear-sky radiance simulation for VIIRS with community radiative transfer model—Part 2: Model architecture and assessment," *Remote Sens.*, vol. 12, 2020, Art. no. 3825.

[15] X. Liang et al., "Monitoring of VIIRS ocean clear-sky brightness temperatures against CRTM simulation in ICVS for TEB/M bands," in *Proc. SPIE, Earth Observ. Syst. XXII*, 2017, Art. no. 104021S, doi: 10.1117/12.2273443.

[16] B. Brasnett, "The impact of satellite retrievals in a global sea-surface- temperature analysis," *Quart. J. Roy. Meteorol. Soc.*, vol. 134, pp. 1745–1760, 2008, doi: 10.1002/qj.319.

[17] X. Liang, Q. Liu, B. Yan, and N. Sun, "A deep learning trained clear-sky mask algorithm for VIIRS radiometric bias assessment," *Remote Sens.*, vol. 12, 2020, Art. no. 78.

[18] X. Liang and Q. Liu, "Applying deep learning to clear-sky radiance simulation for VIIRS with community radiative transfer model—Part 1: Develop AI-based clear-sky mask," *Remote Sens.*, vol. 13, 2021, Art. no. 222.

[19] H. Wang, T. J. Fuller-Rowell, R. A. Akmaev, M. Hu, D. T. Kleist, and M. D. Iredell, "First simulations with a whole atmosphere data assimilation and forecast system: The January 2009 major sudden stratospheric warming," *J. Geophys. Res.*, vol. 116, 2011, Art. no. A12321, doi: 10.1029/2011JA017081.

[20] S.-A. Boukabara et al., "MiRS: An all-weather 1DVAR satellite data assimilation and retrieval system," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 9, pp. 3249–3272, Sep. 2011, doi: 10.1109/TGRS.2011.2158438.

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[22] E. Kim, C.-H. J. Lyu, K. Anderson, R. V. Leslie, and W. J. Blackwell, "S-NPP ATMS instrument prelaunch and on-orbit performance evaluation," *J. Geophys. Res.: Atmos.*, vol. 119, pp. 5653–5670, 2014, doi: 10.1002/2013JD020483.

[23] S. A. Clough et al., "Atmospheric radiative transfer modeling: A summary of the AER codes, short communication," *J. Quantitative Spectrosc. Radiat. Transfer*, vol. 91, pp. 233–244, 2005.

[24] S. A. Clough, M. J. Iacono, and J.-L. Moncet, "Line-by-line calculation of atmospheric fluxes and cooling rates: Application to water vapor," *J. Geophys. Res.*, vol. 97, pp. 15761–15785, 1992.

[25] S. English and T. Hewison, "A fast generic millimeter-wave emissivity model," *Proc. SPIE*, vol. 3503, pp. 288–300, 1998.

[26] Q. Liu, F. Weng, and S. J. English, "An improved fast microwave water emissivity model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1238–1250, Apr. 2011, doi: 10.1109/TGRS.2010.2064779.

[27] X. Liang and A. Ignatov, "Validation and improvements of daytime CRTM performance using AVHRR IR 3.7 um band," in *Proc. 13th AMS Conf. Atmos. Radiat.*, 2010, Art. no. P2.3. [Online]. Available: https://ams.confex.com/ams/pdfpapers/170593.pdf

[28] F. Aires, M. Schmitt, A. Chedin, and N. Scott, "The 'Weight smoothing' regularization of MLP for Jacobian stabilization," *IEEE Trans. Neural Netw.*, vol. 10, no. 6, pp. 1502–1510, Nov. 1999.

[29] F. Aires, C. Prigent, and W. B. Rossow, "Neural network uncertainty assessment using Bayesian statistics with application to remote sensing: 3 network Jacobians," *J. Geophys. Res.*, vol. 109, 2004, Art. no. D10, doi: 10.1029/2003JD004175.

[30] F. Chevallier and J.-F. Mahfouf, "Evaluation of the Jacobians of infrared radiation models for variational data assimilation," *J. Appl. Meteorol. Climatol.*, vol. 40, pp. 1445–1461, 2001.


**Xingming Liang** received the B.S. degree in construction machinery from the Jilin University, Changchun, China, in 1992, and the M.S. degree in information science and the Ph.D. degree in remote sensing from the Saga University, Saga, Japan, in 2002 and 2005, respectively.

From 2007 to 2016, he was a Research Scientist with the Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA, and developed the NOAA leading sea surface temperature (SST) product—Advanced Clear-Sky Processor over Ocean and SST real-time monitoring system—Monitoring of Infrared Clear-sky Radiance over Ocean for SST with NOAA Center for Satellite Applications and Research SST team. From 2016 to 2019, he was a Senior Scientist with the Earth Resources Technology Inc. and supported VIIRS calibration and validation and NOAA Integrated Calibration/Validation System. Since 2019, he has been an Assistant Research Scientist with the Cooperative Institute for Satellite Earth System Studies, University of Maryland, College Park, MD, USA, and working on artificial intelligence applications in remote sensing and data assimilation and supporting the NOAA microwave integrated retrieval system and the community radiative transfer model. His research interests include radiative transfer model, machine learning, SST and aerosol remote sensing, and satellite observation monitoring and validation.

**Kevin Garrett** received the M.S. degree in atmospheric sciences from the Texas A&M University, College Station, TX, USA, in 2007.

He is currently a Physical Scientist with the NOAA/NESDIS Center for Satellite Applications and Research, College Park, MD, USA, and is the Federal Manager for the Community Radiative Transfer Model. Additionally, he serves as the NESDIS Representative to the U.S. Joint Center for Satellite Data Assimilation (JCSDA) Executive Team. Previously, he worked as a Contractor Task Lead on the NOAA Microwave Integrated Retrieval System and supported the JCSDA, leading the optimization of SNPP ATMS and DMSP SSMI/S data assimilation, and implementation of new passive microwave satellite data including GPM GMI, GCOM-W1 AMSR2, and Megha-Tropiques SAPHIR into the NOAA Global Data Assimilation System/Global Forecast System. His research interests include all-sky radiative transfer and applications in passive microwave remote sensing, data fusion, and data assimilation.

**Quanhua Liu** received the B.S. degree from the Nanjing University of Information Science and Technology (former Nanjing Institute of Meteorology), Nanjing, China, in 1982, the master's degree in physics from the Chinese Academy of Science, Beijing, China, in 1984, and the Ph.D. degree in meteorology and remote sensing from the University of Kiel, Kiel, Germany, in 1992.

He was a Senior Research Scientist with the University of Maryland, College Park, MD, USA, working as a Co-Chair of the Community Radiative Transfer Model in the USA. He is currently a Physical Scientist with the NOAA/NESDIS Center for Satellite Applications and Research. He is leading NOAA ATMS sensor data record calibration and NOAA Microwave Integrated Retrieval System that retrieves 10 environmental data records. His research interests include radiative transfer models, satellite products, and sensor calibration and climate studies.

**Eric S. Maddy** received the B.S. degree in physics and the B.A. degree in mathematics from the Frostburg State University, Frostburg, MD, USA, in 2001, and the M.S. and Ph.D. degrees in atmospheric physics from the University of Maryland Baltimore County, Baltimore, MD, USA, in 2003 and 2007, respectively.

He is currently an expert on traditional and exploratory (machine learning and artificial intelligence) polar and geostationary IR/MW remote sounding, data assimilation, and data fusion. He has more than 18 years of experience working with the National Oceanic and Atmospheric Administration Center for SaTellite Applications and Research leading projects on the development of carbon trace gases retrievals from operational hyperspectral sounders, fusing algorithms, and products from multiple instrument platforms (CrIS/ATMS/VIIRS, IASI/AMSU-A MHS/AVHRR), and leading the development of artificial intelligence versions of those algorithms.

**Kayo Ide** received the B.S. degree in aeronautics from the Nagoya University, Nagoya, Japan, and the M.S. and Ph.D. degrees in aeronautics from the California Institute of Technology, Pasadena, CA, USA.

She is currently an Associate Professor with joint appointment with the Department of Atmospheric and Oceanic Science, Department of Mathematics, Earth System Sciences Interdisciplinary Center, and Institute of Physical Science and Technology, University of Maryland, College Park, MD, USA, where she teaches geophysical dynamics, applied mathematics, scientific computing, and data assimilation. She also leads an interdisciplinary scientific prediction group and collaborates with scientists and practitioners with the NASA, Navy, and NOAA Laboratories.

**Sid Boukabara** received the Ph.D. degree in remote sensing from the Denis Diderot University, Paris, France, in 1997.

After the Ph.D. degree, he joined the Atmospheric Environmental Research Inc., until 2005. He then joined the NOAA/NESDIS Center for Satellite Applications and Research, where he has led several efforts notably the development of variational algorithms and serving as a Deputy Director and then the Acting Director of the U.S. Joint Center for Satellite Data Assimilation, leading several advanced satellite data assimilation activities. Between 2015 and 2017, he served as a Senior Policy Advisor to the Commerce Department's Assistant Secretary for Environmental Observing Systems and Prediction, providing scientific perspectives to decision- and policy-making. He has 24 years of experience accumulated between the private sector and the U.S. Federal Government. He is currently a Principal Scientist with the NESDIS's Office of Systems Architecture and Advanced Planning supporting the effort to design and optimize the Nation's next-generation space and ground architecture. He also collaborates with Government, academic, and private-sector partners to explore innovative ways to exploit environmental data, with a view to enhance the satellite processing systems of the future. In particular, he is an expert in exploring new numerical approaches, such as artificial intelligence machine learning and computer vision to improve the Nation's usage of satellite data in NOAA's environmental prediction systems. His research interests include satellite data assimilation, radiative transfer, variational systems, and providing mechanisms to evaluate and optimize the science value of global observing systems.