Earth and Space Science

RESEARCH ARTICLE
10.1029/2020EA001402

Enhance Low Level Temperature and Moisture Profiles Through Combining NUCAPS, ABI Observations, and RTMA Analysis

Zheng Ma[1,2,3], Zhenglong Li[2], Jun Li[2], Timothy J. Schmit[4], Lidia Cucurull[5], Robert Atlas[6,7], and Bomin Sun[8]

[1]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [2]Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, Madison, WI, USA, [3]University of Chinese Academy of Sciences, Beijing, China, [4]Advanced Satellite Product Branch, Center for Satellite Applications and Research, NEDIS/NOAA, Madison, WI, USA, [5]NOAA Atlantic Oceanographic and Meteorological Laboratory, Miami, FL, USA, [6]Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD, USA, [7]NOAA Atlantic Oceanographic and Meteorological Laboratory (Director Emeritus), Miami, FL, USA, [8]I. M. Systems Group at NOAA Center for Satellite Application and Research (STAR), College Park, Suitland, MD, USA

Key Points:

- Combining NOAA Unique Combined Atmospheric Processing System and Advanced Baseline Imager observations provides improved temperature and moisture soundings
- Adding surface observations further improves low level temperature and moisture profiles
- Fusing multi-source observations is a feasible way to improve boundary layer thermodynamics for research and applications

Correspondence to:
J. Li and Z. Li,
jun.li@ssec.wisc.edu;
zhenglong.li@ssec.wisc.edu

Citation:
Ma, Z., Li, Z., Li, J., Schmit, T. J., Cucurull, L., Atlas, R., & Sun, B. (2021). Enhance low level temperature and moisture profiles through combining NUCAPS, ABI observations, and RTMA analysis. *Earth and Space Science*, *8*, e2020EA001402. https://doi.org/10.1029/2020EA001402

Received 17 AUG 2020
Accepted 25 APR 2021

Abstract Thermodynamic information from low levels in the atmosphere is crucial for operational weather forecasts and meteorological researchers. The NOAA Unique Combined Atmospheric Processing System (NUCAPS) sounding products have been proven beneficial to fill the data gap between synoptic radiosonde observations (RAOBs). However, compared with the upper troposphere, the accuracy of NUCAPS soundings in the low levels still needs improvement. In this study, a deep neural network (DNN) is applied to fuse multiple data sources to enhance the NUCAPS temperature and moisture profiles in the lower atmosphere. The network is developed by combining satellite observations, including NUCAPS sounding retrievals and high resolution geostationary satellite observations from the Advanced Baseline Imager, and surface analysis from the Real-Time Mesoscale Analysis (RTMA) as inputs, while collocated soundings from ECMWF re-analysis version 5 are used as the benchmark for the training. The performance of the model is evaluated by using the independent testing data set, data from a different year, as well as collocated RAOBs, showing improvement to the temperature and moisture profiles by reducing the root-mean-squared-error (RMSE) by more than 30% in the lower atmosphere (from 700 hPa to surface) in both clear sky and partially cloudy conditions. A convective event from June 18, 2017 is presented to illustrate the application of the enhanced low level soundings on high impact weather events. The enhanced soundings from fused data capture the large surface-based convective available potential energy structures in the preconvection environment, which is very useful for severe storm nowcasting and forecasting applications.

1. Introduction

Vertical profiles of temperature and moisture reflect important atmospheric information. In particular, the lower atmosphere, including but not limited to the boundary layer, is crucial for weather forecasting because this is where people live, and where the exchange of energy, momentum, and mass between the atmosphere and the surface occurs (Space Studies Board et al., 2019). In observing the lower atmosphere, satellite based sounders have the advantage of better spatial coverage than ground-based networks. The NOAA Unique Combined Atmospheric Processing System (NUCAPS) was developed to provide retrieved atmospheric vertical profile environmental data records (EDRs) under nonprecipitating conditions (Gambacorta, 2013), derived from the Cross-track Infrared Sounder (CrIS) and Advanced Technology Microwave Sounder (ATMS) currently onboard the Joint Polar Satellite System (JPSS) series (Suomi-NPP and NOAA-20) (Goldberg et al., 2013). For more details about the NUCAPS algorithm, readers are referred to Gambacorta et al. (2012 and 2015). The primary NUCAPS products are vertical atmospheric profiles of temperature and moisture at 100 vertical layers (Strow et al., 2003), which are the standard pressure layers of the University of Maryland Baltimore County radiative transfer algorithm; these profiles provide valuable information for forecasters and research studies. Over the years, numerous validation and application tests have been conducted to assess the performance of the retrieved temperature and moisture profiles (Feltz

et al., 2017; Nalli et al., 2013, 2018; A. Smith et al., 2015; Sun et al., 2017). During the Hazardous Weather Testbed Spring Experiments, which mainly focus on local severe storms (LSS), forecasters found NUCAPS helpful in filling spatial and temporal gaps in the atmospheric sounding system, but also reported a reduced accuracy in the lower levels near the surface. While this is typical for satellite based sounding retrievals due to complications from surface emissivity and skin temperature, as well as the coarser spatial resolution, this limitation could lead to a misrepresentation of the atmospheric thermodynamic structure and limit the application of NUCAPS sounding and derived products, such as Convective Atmospheric Potential Energy (CAPE) and Lifted Index (LI), in the storm forecasting and warning process. As a result, forecasters often had to manually modify the NUCAPS sounding lower levels based on other information and their experiences, which is time consuming and not as objective.

The lower levels of satellite sounding retrievals can be enhanced by other observations sensitive to the lower atmosphere. The legacy GOES sounder profiles employed a similar approach, using a surface analysis of temperature and dewpoint depression (Hayden, 1988). It has been shown by Smith et al. (2020) that by fusing polar hyperspectral soundings and geostationary multi-spectral soundings, the accuracy of NWP in severe weather forecasting can be improved. For NUCAPS, Bloch et al. (2019) demonstrated that the underestimation of surface-based convective available potential energy (SBCAPE) from NUCAPS products can be improved through combining the NUCAPS profile and surface observations. From the NUCAPS team, a mitigation method was implemented by developing an automatic correction scheme, in which the Real-time Mesoscale Analysis (RTMA) (De Pondeca et al., 2011) surface observations of temperature and dewpoint are used to create a well-mixed boundary layer to replace the lower levels of the NUCAPS profile (Michael & Kristin, 2017). This modified version of NUCAPS was well received by forecasters as it performed better than the original NUCAPS soundings and saved time for forecasters from manually modifying the profiles. However, as the correction scheme is based on the assumption that the boundary layer is well-mixed, two important issues remain to be solved: (a) the boundary layer is not always well-mixed, especially in the morning and evening, making it difficult to determine vertically how far away from the surface the correction should go, and (b) the low quality of satellite sounding retrievals is not limited to the boundary layer, as will be later demonstrated in this study. In this case, a more objective approach to enhance the soundings in the lower atmosphere is needed in order to improve the application of NUCAPS products in weather forecasting, especially in convective environments. In addition, bias correction methodologies are being studied to enhance the accuracy of satellite profile retrievals. A dual-regression (DR) technique (W. L. Smith Sr et al., 2012 Weisz et al., 2013) has been developed with a cloud-height-classification process to address the issue of nonlinear dependence of spectral radiance on the atmospheric variables. A further de-aliasing (DA) error correction method is developed by W. Smith Sr et al. (2015) to reduce the bias of retrievals from DR regression resulting from the limited vertical resolution, and get improved accuracy.

The first satellite of the new generation Geostationary Operational Environmental Satellite (GOES-R series), named GOES-16, was launched in November 2016. The Advanced Baseline Imager (ABI) (Schmit et al., 2005) onboard the GOES-R series has much improved capabilities over those of previous GOES imagers, beginning with a 16-channel imaging radiometer, which is 11 more channels than the previous GOES imager. ABI also has finer spatial resolution (2 km for IR channels) and faster coverage rate (10 min for full disk (FD) in mode 6, 5 min for contiguous US (CONUS), and 1 min for two independent mesoscale (MESO) regions) compared to previous GOES imagers (Schmit et al., 2017). Although GOES-R series satellites do not have a hyperspectral infrared sounder like CrIS onboard to provide atmospheric profiles with high vertical resolution, ABI does contain useful information about the lower atmosphere. The weighting functions of the ABI channels indicate that four longwave window spectral channels (11/13/14/15) and the CO_2 channel (16) are sensitive to the lower atmosphere as their weighting functions peak near the surface. Also, as ABI provides nearly continuous observations with high temporal resolution, it is capable of identifying the temporal variations of the lower atmosphere's thermodynamics. In addition, the high spatial resolution of ABI makes it possible to provide sub-footprint spatial characteristics of the NUCAPS field of regard (FOR). While the limit on spectral resolution makes it difficult to use ABI alone to provide atmospheric profiles with high vertical resolution (Schmit et al., 2008, 2019), the information from ABI can help enhance the NUCAPS soundings in the lower atmosphere.

Deep Neural Networks (DNNs) is the general term for a series of multi-layer neural networks trained with multiple hidden layers between input and output layers (Bengio, 2009; Hinton, 2006), developed on the basis of Artificial Neural Networks (ANNs). DNN has shown success in areas such as computer vision, bio-information processing, and speech recognition, etc. (Dahl et al., 2011; J. N. Liu et al., 2014; Liu et al., 2017; Seltzer et al., 2013; Tao et al., 2018). While ANN with activation functions applied can find nonlinear relationships, DNN has shown its advantage in dealing with large and complex datasets and further assisting the estimation. In the atmospheric sciences, Boukabara et al. (2019) has demonstrated that DNN can be applied to data fusion, inversion, data assimilation, and even convective event nowcasting; many studies have shown that the DNN algorithm is a powerful and extremely efficient tool when working with large amounts of data. In recent years, DNN- based algorithms have gained success in bias correction for satellite-derived products, as Tao et al. (2016) who successfully applied DNN in reducing bias and false alarms of satellite precipitation products, and Zhou & Grassotti (2020) who implemented DNN as a new approach of radiometric bias correction in MiRS system. DNN in these works have shown good capability on addressing nonlinear relationships, and high efficiency in generating predictive models with good accuracy. These capabilities make it ideal for fusing multi-source datasets to improve the lower levels of NUCAPS soundings to make it applicable for severe weather nowcasting and forecasting. Two important questions will be addressed in this study:

1. How to enhance the lower levels of NUCAPS soundings through fusing multiple data sources using DNN?
2. What impact does high-resolution geostationary satellite data and surface temperature and moisture observations have on enhancing the NUCAPS soundings, respectively?

The objective of this study is to use the ERA5 (ECMWF re-analysis version 5) reanalysis as label data in a statistical model to improve NUCAPS soundings for nowcasting applications. The enhanced NUCAPS soundings, with the focus on the lower atmosphere, are achieved through the data fusion of NUCAPS soundings, ABI clear radiances, and RTMA surface analysis in near real-time (NRT) so that they can be better used for nowcasting of LSS in preconvection environment. With this enhancement, forecasters will no longer need to manually modify the lower levels of NUCAPS soundings before using them. In this study, a DNN framework has been designed and developed to enhance the lower levels of NUCAPS (Suomi-NPP) sounding products by fusing multi-source data sets. In this study, the lower levels are defined as the levels between 0.7 and surface in sigma coordinates, or roughly 700 hPa and the surface for non-high terrain regions. A match-up data set are developed, including NUCAPS sounding products, RTMA gridded data, GOES-16 ABI radiances, and the ERA5 for April/May/June 2018. For the training, an ERA5 profile is used as the predictand, and others are used as predictors. Sensitivity tests are conducted to assess the relative impact of geostationary satellite observation from ABI and RTMA surface observations on enhancing the soundings. Validation studies against independent radiosonde observations (RAOBs) are also performed.

The data and methodologies used, and the experimental design in this study, are described in Section 2. Section 3 presents the validation results of the DNN-based data fusion models developed against ERA5 under clear-sky and partly cloudy conditions, together with the analysis on the relative contribution from the various datasets. Section 4 presents a further evaluation of the model against RAOBs collected at the NOAA Products Validation System (NPROVS, Reale et al., 2012; Sun et al. 2017). In Section 5, a high impact weather case is presented to demonstrate the application of the enhanced NUCAPS soundings. Finally, the main conclusions and future considerations are in Section 6.

2. Data and Methodologies

2.1. Data

2.1.1. NUCAPS

Three months of operational NUCAPS version 1.0 from S-NPP environmental data records (EDRs) from April to June 2018 over the CONUS region (restricted as 20°N $\sim$ 55°N, $-140° \sim -60°$) are used in this study (Q. Liu et al., 2014). Derived from the hyperspectral IR sounder (Menzel et al., 2018) CrIS and MW (microwave) sounder ATMS operating in an overlapping 3×3 footprint area, NUCAPS provides retrievals from clear to nonprecipitating cloudy conditions through procedures including cloud-clearing. The 3×3

CrIS field of view (FOV) arrays are referred to as the NUCAPS field of regard (FOR). Vertically, the profiles are retrieved at 100 fixed pressure layers from 0.016 hPa to 1,100 hPa. Those are the standard 100 layers (101 levels) for the Stand-Alone Radiative Transfer Algorithm (Hannon et al., 1996; Strow et al., 2003). The NUCAPS soundings are used as primary predictors for the DNN since they are the object to be enhanced. While this study focuses on the lower atmosphere, the levels between 700 and 200 hPa are also included, because as shown later, the tropospheric profiles are also improved. The surface pressure information from RTMA is also included as a predictor to identify the surface level. Levels below the surface are assumed to be isothermal and isohume in the training; those levels are excluded from the statistics. Due to the large size of a NUCAPS FOR (about 45 km at nadir), completely clear FORs make up a relatively small percentage, with a large percentage of them in partly cloudy regions. This study focuses on enhancing the NUCAPS soundings in lower levels under both clear and partly cloudy conditions.

2.1.2. ABI

The GOES-16 ABI radiance observations over CONUS are used to provide additional information to help enhance the NUCAPS profiles. Of the 16 ABI bands, the mean brightness temperature (BT) and standard deviation (STD) within NUCAPS sub-FOR from seven IR bands (8, 9, 10, 13, 14, 15, and 16, or 6.2, 6.9, 7.3, 10.3, 11.2, 12.3, and 13.3 μm) are used as predictors in this study, covering water vapor channels (band 8, 9, and10), longwave window channels (band 13, 14, and15) and the CO_2 channel (band 16). As depicted by their respective weighting functions (See Figure 4 in Schmit et al., 2008), these channels are either sensitive to the lower level atmosphere, or contain information about tropospheric water vapor absorption. The operational ABI cloud mask products (Heidinger & Straka, 2013) are applied to identify the pixels that are cloud contaminated. In this study, the most recent ABI observations in the CONUS sector prior to the NUCAPS observation time are used to represent the characteristics observed from ABI bands, which means the time difference between ABI and NUCAPS is no more than 5 min.

2.1.3. RTMA

The RTMA is a NOAA/NCEP analysis/assimilation system for near-surface weather conditions using the NCEP/EMC Gridpoint Statistical Interpolation system in 2-DVar mode to assimilate conventional and satellite-derived observations. The RTMA currently runs in four domains, namely, CONUS, Alaska, Puerto Rico, and Hawaii National Digital Forecast Database grids, where observations originating from synoptic, aviation routine weather report, Mesoscale Network (Mesonet), ship, buoy, tide gauge, and Coastal-Marine Automated Network stations are assimilated (De Pondeca et al., 2011). In this study, the hourly 2-m temperature and dewpoint analysis from the 2.5 km CONUS-RTMA data set are used as two predictors to provide important atmospheric information near the surface. As the near surface atmospheric thermodynamics are highly correlated to the low level atmosphere, the RTMA gridded data is expected to add value via enhancing the accuracy of profiles in the lower atmosphere.

2.1.4. ERA-5

The ERA-5 data set, which is the successor to ERA-Interim, is the fifth generation reanalysis generated by the European Center for Medium-Range Weather Forecasts (ECMWF). It uses 4D-Var data assimilation in CY41R2 of ECMWF's Integrated Forecast System, with 137 hybrid sigma/pressure (model) levels in the vertical, with the top level at 0.01 hPa. Compared with ERA-Interim, ERA-5 has enhanced horizontal and vertical resolution, as well as an enhanced output frequency (Hoffmann et al., 2019; Service (C3S), 2017; Tarek et al., 2020). The data set used in this study is from reanalysis products that are interpolated to 37 fixed pressure levels from 1 to 1,000 hPa, with a spatial resolution of 0.25° and a temporal resolution of 1 h. The ERA-5 data set is used as the output in the training process, and as references to provide truth values in the validation process. It is important to point out that ERA5 is arguably one of the better reanalysis data sets in terms of accuracy, spatial resolution, and temporal resolution. While RAOBs would serve better to be used as the predictand, the limited sample size due to NUCAPS' local pass time, is not enough for training. They will be used for independent validation instead.

2.1.5. Radiosonde Data

RAOBs used in the validation are collected by the NPROVS (Reale et al., 2012; Sun et al., 2017), supported by the NOAA JPSS and operated at the NOAA NESDIS Office of Satellite Applications and Research

starting in 2008. NPROVS provides routine data access, collocation, and inter-comparison of multiple satellite temperature and water vapor sounding product suites and NWP model profiles matched with global conventional and special radiosonde observations. The collocation approach is to select the "single closest" sounding from each product suite anchored to the RAOB launch location.

Special RAOBs include primarily the JPSS-funded dedicated sonde observations and Global Climate Observing System Global Reference Upper Air Network (GRUAN, Bodeker et al., 2016) sonde observations. The sondes used in the study, covering April/May/June of 2017 and 2018, are primarily the dedicated launches that were coincidental with an S-NPP overpass. They were launched at the Southern Great Plains (SGP), Oklahoma, a Department of Energy Atmospheric Radiation Measurement program site; the Howard University Beltsville Center for Climate System Observation site in Beltsville, Maryland, a GRUAN site; and Boulder, CO, a GRUAN site. Note, those dedicated sondes were not assimilated into NWP, constituting a valuable data source for satellite data calibration and validation.

To obtain a reasonable number of RAOBs for use in this validation effort, RAOBs that were launched within 2 h of an S-NPP overpass are utilized. As a result, some RAOBs launched at synoptic times from SGP (which are also collected at NPROVS) are included.

RAOBs used in this study include three types: Vaisala RS92 corrected or processed with standard operational processing (34% of total RAOBs), Vaisala RS92 processed with GRUAN data processing (Dirksen et al., 2014) (22%), and Vaisala RS41 with standard operational processing (44%).

RAOB temperature observations may suffer from solar radiative heating in the lower stratosphere, but that is not an issue since the validation is cut off at 200 hPa. The RAOB humidity observations tend to have a dry bias in the upper troposphere particularly during the daytime and for RS92 with standard operational processing. Use of RS92 and RS41 in satellite hyperspectral sounding products can be found in Sun et al. (2017, 2018).

2.2. Methodologies

2.2.1. Pre-Processing

As shown in Table 1, there are five main sources of data used in this study, of which NUCAPS, ABI, and RTMA are the three main input data sets for the DNN, and ERA-5 serves as the output. RAOBs from NPROVS are used as the independent validation data set. A match-up data set, including the predictors from the output and all three input sources, is created during the pre-processing step. Quality flags for the NUCAPS data set are applied to keep only those FORs with good quality from the IR + MW retrievals. All other data sources are collocated to the NUCAPS location and time. For the collocation of ABI, the most recent CONUS observation times prior to NUCAPS soundings are used for the temporal collocation, while spatially all the pixels within each NUCAPS FOR (or the 3 by 3 CrIS FOVs) are searched with the collocation tool developed by the Atmosphere Science Investigator-led Processing Systems team at SSEC (Nagle & Holz, 2009). After that, the ABI cloud mask product helps to eliminate pixels that are cloud contaminated. Through the area weighted average of clear sky ABI pixels, the average BT and STD of BTs within the FOR are obtained, with STD representing the homogeneity of ABI BTs within the FOR. For each NUCAPS FOR, the percentage of clear area is calculated as the ratio of the clear ABI pixels to the total ABI pixels. Note that an ABI pixel with a larger local zenith angle (LZA) covers a larger area than those with smaller LZAs. This percentage characterizes the averaged clear sky BT's representativeness of the spatial characteristics of the FOR, and is used as an additional predictor in the training process under partly cloudy conditions. For analysis fields from RTMA and ERA-5, temporal and spatial interpolations are performed to collocate to the observation time and location of the NUCAPS FORs. After the hourly gridded data set is interpolated spatially to the NUCAPS FORs, a temporal interpolation is performed on each FOR using the nearest two hours' data for that FOR to obtain the final collocated value. The match-up data set has 426966 samples.

2.2.2. DNN

A DNN framework is designed and developed in this study to enhance the lower levels of NUCAPS temperature and moisture profiles by fusing data from NUCAPS soundings, sub-FOR ABI observations, and

Table 1
Data and Variables Used for Training and Validation in This Study

Data usage	Data source	Variable	Unit
Predictors	NUCAPS	Temperature (from 200 hPa to surface)	K
		Water vapor mixing ratio (from 200 hPa to surface)	$g{\cdot}kg^{-1}$
		Surface pressure	hPa
	ABI	Brightness temperatures and their Sub-FOR ABI homogeneity (Bands: 6.2, 6.9, 7.3, 10.3, 11.2, 12.3, and 13.3 μm)	K
		Pixel satellite zenith angle: sec(θ)	N/A
		Clear pixel percentage (for partly cloudy): clear_pixel_area/total_pixel_area × 100%	%
	RTMA	Surface 2-m temperature	K
		Surface 2-m dewpoint	K
Predictands	ERA5	Temperature (from 200 hPa to surface)	K
		Water vapor mixing ratio (from 200 hPa to surface)	$g{\cdot}kg^{-1}$
Validation	RAOB	Temperature (from 200 hPa to surface)	K
		Water vapor mixing ratio (from 200 hPa to surface)	$g{\cdot}kg^{-1}$

ABI, Advanced Baseline Imager; ERA5, ECMWF re-analysis version 5; NUCAPS, NOAA Unique Combined Atmospheric Processing System; RAOB, radiosonde observations.

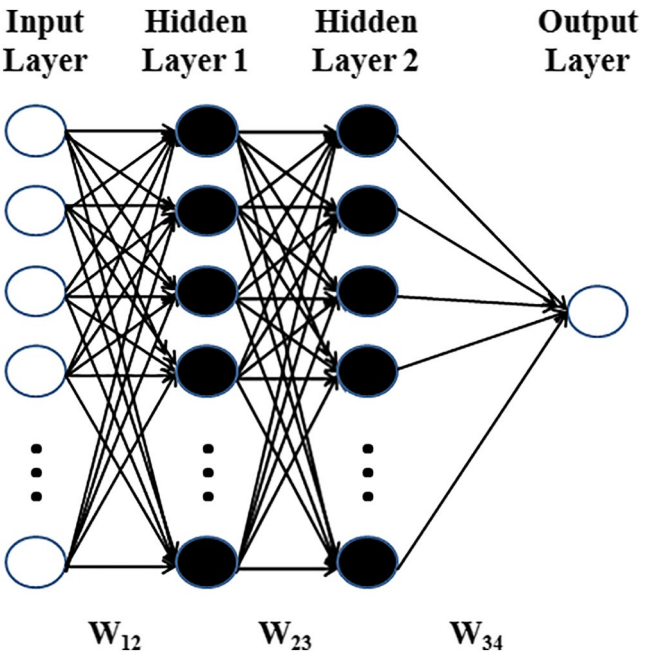

Figure 1. The four-layer, fully connected DNN framework is used in this study. The first layer is the input layer where the predictor variables are input. The second and third layers are hidden layers where nonlinear regressions take place. The fourth layer is the output layer which directly links to the target value the model is to predict. W_{ij} stands for the weights of connections between different neurons from layer i to layer j, and works with activation functions to determine various input-output relationships. DNN, deep neural networks.

surface observation from RTMA. Similar to Tao et al. (2016, 2018), a four-layer, fully connected neural network is used, as shown in Figure 1. The network is comprised of neurons organized in layers, and the connections between the neurons throughout the layers. A neuron receives inputs from the connections to the previous layer, summarizes them, and produces an output through an activation function before passing the output to the next layer. The first layer in a network serves as the input layer, and the last layer serves as the output layer for the final prediction. The layers between the input and output layers are called hidden layers. These are critical for the model's nonlinearity because of the application of activation functions. The relationships between input and output are determined by the parameters assigned to the connections. These parameters (weights) are randomly initialized at the beginning and modified automatically through the training of the input and output samples. A loss function is defined to assess the performance of the network throughout the training. When the output of the loss function exceeds the preset threshold, indicating the expected accuracy is not met, back propagation starts and the weights of the connections for each layer are modified. Forward and backward iterations continue until the loss function is reduced to a limited range.

For this network, two hidden layers are used between the input and output layers. This number is determined with reference to previous works by Tao et al. (2016, 2018) and Zhou and Grassotti (2020), after a few tests, it showed that the improvement of adding more layers for this task is not substantial enough when compared with the increased cost of computation and risk of overfitting. The input layer consists of all the predictors for the training, including the NUCAPS temperature/moisture values for all levels 200 hPa and below, ABI variables, and RTMA variables mentioned in Table 1, with each predictor serving as a neuron. The number of neurons in each of the hidden layers is set to 100 after several tests.

Table 2
Valid Sample Size for Clear Sky Training and Validation at Each Pressure Level

Pressure level (hPa)	Sample size for training	Sample size for validation	Pressure level (hPa)	Sample size for training	Sample size for validation
650 and levels above	71,696 for each level	17,925 for each level	850	66,016	16,497
700	71,648	17,914	875	63,313	15,820
750	71,298	17,824	900	60,690	15,126
775	70,769	17,682	925	56,955	14,230
800	69,252	17,312	950	51,658	12,958
825	67,754	16,947	975	37,281	9331

There is only one neuron in the output layer representing the regression value of temperature/moisture at a specific ERA fixed pressure level. The enhanced profiles are formed with the outputs from all the trained models for different levels combined together vertically. Before adding them to the training framework, a zero-centered normalization is applied to each predictor to eliminate the difference in order of magnitude. A weight initialization method developed by He et al. (2015), known as He-et-al Initialization, is applied to the network to enhance the efficiency of the training process. The application of this method makes the gradient descend faster and more efficiently, thereby improving the performance of model training to some extent. The newly developed Exponential Linear Unit (ELU) activation function (Clevert et al., 2016) is applied to each of the layers except the output layer, where no activation function is used. As an activation function, ELU is very similar to rectified linear units (ReLU) (Nair & Hinton, 2010) except when dealing with negative inputs. Both in identity function form for non-negative inputs, ELU differentiates from ReLU as it can produce negative outputs which allow them to push mean unit activations closer to zero, similar to batch normalization but with lower computational complexity. According to Pedamonti's experiments on classification task (Pedamonti, 2018), the ELU activation has shown better performance in accuracy over ReLU. In this study, the loss function selected during the training process is mean squared error. In addition, we use the Adam Optimizer (Kingma & Ba, 2017), an adaptive learning rate optimization algorithm designed specifically for DNN training, which proves to be well suited for problems that have large amounts of data and/or parameters. The match-up data set is randomly split into a training set containing 80% of the data and an independent validation set containing the remaining 20% using the random splitting tool from Scikit-learn (Pedregosa et al., 2011), for which the training set is used as inputs in the training process. The trained models are tested with both the training and the independent validation data sets to ensure that there is no overfitting. Only when the prediction accuracy of the training set and validation set appear to be close to each other is the model regarded as stable.

2.3. Sensitivity Experiment Design

In order to understand the relative contribution from different data sources in enhancing the low level soundings, two sets of sensitivity experiments are conducted. The first set of experiments focuses on understanding NUCAPS, ABI, RTMA, and their combined effects. The other set, through predictor denial experiments, focuses on understanding ABI variables, that is, the relative impact of different channels and their respective sub-FOR homogeneities. Statistical analysis and comparisons of individual profiles from these experiments are presented in the following sections.

3. Results and Analysis

3.1. Statistical Validation Results

3.1.1. Validation in Clear Skies

As mentioned in Section 2, the collocated samples are randomly separated into a training set consisting of 80% of the samples, and an independent testing set consisting of the remaining 20%. Table 2 shows the clear-sky valid sample sizes for each level. Since we only use above surface values, the valid sample size turns out

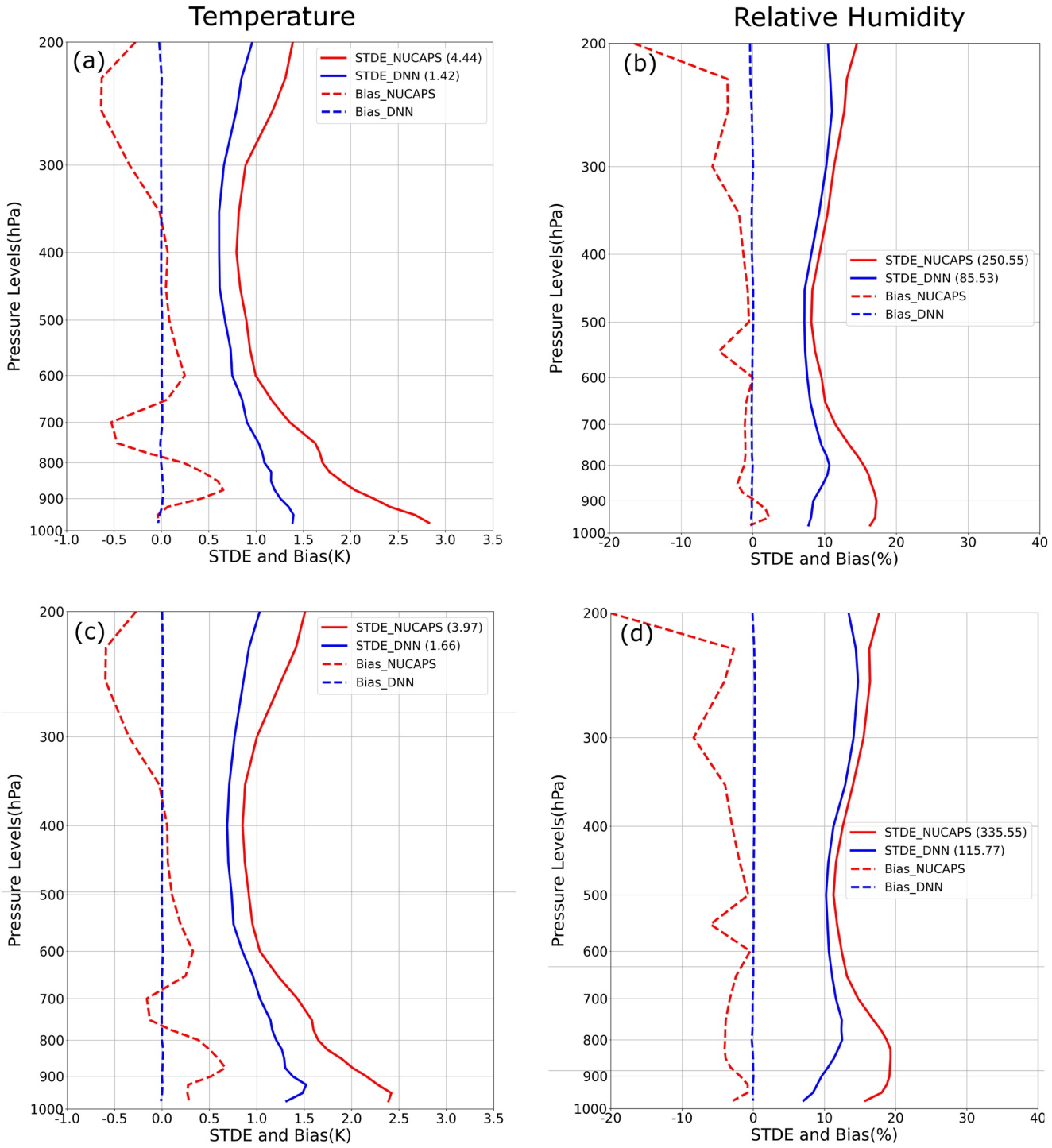

Figure 2. Statistics of Standard Deviation of Error (STDE, solid lines) and mean bias (dashed lines) vertical profiles for (a), (c) temperature, and (b), (d) relative humidity (RH) of NUCAPS (red lines) and outputs from DNN data fused model (blue lines) versus ERA5 at different pressure levels under (a), (b) clear-sky and (c), (d) partly cloudy conditions from the testing data set. Shown in the brackets are the values of mean variances (MV) = $RMSE^2$ of lower levels from 700 hPa to the surface for temperature and RH, respectively. DNN, deep neural networks; RMSE, root-mean-squared-error.

to be smaller when the pressure level is closer to the ground. To maintain the statistical consistency and significance, levels with a sample size smaller than 50% are not used in the statistics. The standard deviation of error (STDE) and mean bias between the predicted results from data fusion and the corresponding ERA5 values for temperature and relative humidity are shown in Figure 2. For humidity, the variable output by

Table 3
Valid Sample Size for Partly Cloudy Training and Validation at Each Pressure Level

Pressure level (hPa)	Sample size for training	Sample size for validation	Pressure level (hPa)	Sample size for training	Sample size for validation
650 and levels above	269,883 for each level	67,462 for each level	850	250,999	62,761
700	269,657	67,411	875	243,434	60,937
750	268,403	67,127	900	236,984	59,341
775	266,511	66,660	925	227,548	56,922
800	261,779	65,516	950	208,101	52,112
825	256,574	64,206	975	152,110	38,168

the model is water vapor mixing ratio, and quality control is applied in the form of post-processing by fixing any negative result to zero. Then the relative humidity (RH) used for validation is calculated with the predicted mixing ratio and the ERA5 temperature (truth) to ensure the validation of moisture is independent from the error introduced by temperature. From the results in clear skies (Figures 2a and 2b), the accuracy of NUCAPS soundings in the lower atmosphere is significantly reduced compared to the upper levels. The temperature STDE is larger than 2.0 K for levels 850 hPa and below, and the RH STDE is larger than 15% for levels 800 hPa and below. Significant improvements can be seen from the enhanced soundings. For both the temperature and moisture (RH) profiles, the mean biases are reduced to a very small amount (almost 0) from 975 to 200 hPa compared to the original NUCAPS profiles; substantial STDE reductions are also found throughout the whole vertical profile. Significant improvements can be seen in the lower levels, especially for 850 hPa and below, where the reduction in RMSE is over 30% for both temperature and RH profiles. Also shown in Figure 2 is the mean variance ($RMSE^2$) of the lower levels (defined as the vertical mean of the variance of lower levels from 700 hPa to surface) from the enhanced results and the original NUCAPS; the variances are significantly reduced by 68.1% for temperature and 65.9% for RH. Note that $RMSE^2$ is used instead of STD^2 to show the overall improvement. It is worth noting that the STD values shown in Figure 2 are significantly smaller than those of ERA5 (4–9 K for temperature and 20%–30% for RH), which characterize the natural variation of ERA5 temperature and moisture. These validation results using the independent validation data set indicate that the DNN based fusion model for combining NUCAPS, ABI and RTMA is working as expected to improve the NUCAPS sounding profiles in both temperature and moisture profiles, especially in the lower levels.

3.1.2. Validation in Partly Cloudy Skies

The DNN-based data fusion model in partly cloudy sky conditions is similar to that in clear skies. All the predictors used in clear skies are also used under partly cloudy conditions, with one modification and one addition. In clear skies, all ABI variables are included for the whole NUCAPS FOR since it is completely clear, while in partly cloudy conditions, all ABI variables are included for the clear portion of the NUCAPS FOR. Hence, the clear sky ABI observations still enhance the low levels of NUCAPS soundings. In addition, to account for the representativeness of the clear sky ABI observations, the percentage of the area covered by clear ABI pixels for each FOR is also included as a predictor. If no clear pixel exists within a NUCAPS FOR, it is omitted. A few experiments have also been conducted to determine the lowest percentage clear portion before excluding the FORs, as with the least representative ABI information they could possibly result in no positive impact on the training. The validation experiment using FORs with ABI clear percentages less than 1% still shows a slight improvement over the experiment with all other predictors other than from the ABI. As the sample size is small for these low clear-percentage samples, it is not statistically significant to determine the optimal threshold of the clear-percentage. The threshold used in this study is empirically set to 1%. The impact on the sounding yield is insensitive to this setting.

Table 3 lists the sample size for each level. The validation results using ERA-5 are shown by Figures 2c and 2d. The reduction in errors for temperature and moisture profiles shows that the data fusion model in partly cloudy conditions can enhance the NUCAPS soundings in the lower levels under partly cloudy conditions, in both bias and STDE. The improvement in the variances of the lower levels is 58.1% for temperature and 65.5% for RH, which is comparable to the values from clear skies. This result indicates that the model

developed through data fusion is able to take the useful information from clear sky ABI and RTMA to improve the low level NUCAPS soundings in partly cloudy skies. There is also improvement in the NUCAPS profiles above the lower levels. It is interesting that NUCAPS in partly cloudy conditions appears to have smaller STDE in the lower atmosphere than in clear skies. However, note that the samples are different, which invalidates such a conclusion.

3.2. Relative Impacts of Observations on Enhanced Soundings

Impact studies in the form of Observation System Experiment (OSEs) are conducted under clear-sky conditions to understand the relative contribution of ABI and RTMA on reducing the uncertainties in NUCAPS soundings in lower levels. Three additional observation denial experiments are carried out:

- DNN_NUCAPS is for the experiment including only NUCAPS profiles as predictors
- DNN_ABI is for the experiment including only ABI variables as predictors
- DNN_NUCAPS_ABI is for the experiment including both NUCAPS and ABI predictors, but excluding the surface observations from RTMA.

For consistency, all three experiments use the same training set (80%) and validation set (20%). DNN_NUCAPS_ABI_surf means that all the data sources are included, same as the predicted results from Figure 2. As shown in Figure 3, significant improvements can be seen at all levels from DNN_NUCAPS over NUCAPS. From Figure 2, the NUCAPS temperature bias is much smaller than the STDE, so the improvement shown in Figure 3 is mostly due to a reduction in STDE. This improvement is from the deep learning alone, without additional information, indicating the power of DNN to take advantage of the correlations between different levels in NUCAPS soundings and make improvements. Further improvements from DNN_NUCAPS_ABI and DNN_NUCAPS_ABI_surf reflect the added value in enhancing the lower level soundings with additional fused data. When ABI information is added into the training, the improvements in accuracy can be seen for both temperature and moisture profiles. It is worth noting that the improvements are more significant at lower levels, which is consistent with our expectations as the ABI window channels are most sensitive to the lower atmosphere where the NUCAPS sounding retrievals show reduced accuracy. When RTMA is added to the training, the accuracy is further improved at the low levels. However, no obvious improvement above 600 hPa for temperature and above 800 hPa for moisture is noticed compared to the contributions from ABI. This result indicates that while RTMA could provide near-surface information that is highly correlated with the lower level atmosphere, the vertical range of this correlation is limited. Figure 3 shows that there are three major sources contributing to improving the NUCAPS soundings: the DNN itself, the ABI, and the surface observations from RTMA. The mean variance of the lower levels from 700 hPa to the surface is also shown in Figure 3 for each experiment along with the original NUCAPS. Based on the results for clear sky conditions (shown by Figures 3a and 3b), the DNN itself contributes the largest variance reductions of 55.4% for temperature and 51.7% for RH; the ABI, with four bands sensitive to the lower atmosphere, has moderate variance reductions of 7.2% for temperature and 10.2% for RH; and the RTMA, with only near surface information, contributes the smallest variance reductions of 5.5% for temperature and 3.9% for moisture.

The experiments on observations are also carried out for FORs under partly cloudy sky conditions. The statistical STDE profiles of temperature and relative humidity are shown in Figures 3c and 3d, respectively. Similar to the results for clear skies, contributions from the DNN itself, the ABI, and the RTMA can be seen by comparing the different experiments. The mean variance reductions are 40.1% from the DNN, 7.1% from the ABI, and 10.9% from the RTMA for temperature, and 48.4% from the DNN, 8.6% from the ABI, and 8.4% from the RTMA for relative humidity. Again, the DNN itself has the largest contribution, and the contribution from ABI and RTMA are both substantial in the lower levels. But the relative contribution from ABI is smaller than that in clear skies, while the contribution from RTMA is relatively larger. As the RTMA surface observations are all-sky products, it appears that the contribution of ABI in partly cloudy conditions is lower than that in clear skies. This effect can also be seen from the ABI alone experiment, which is the only profile that differs greatly from the clear-sky conditions. As shown in Figure 3c, the STDE of temperature from ABI alone is smaller than that of the original NUCAPS for near-surface levels, but larger than NUCAPS for all levels above. Compared with the results in Figure 3a, the level where ABI alone results start exceeding NUCAPS is much lower in Figure 3c. Similar results can be seen from the comparison of RH between Fig-

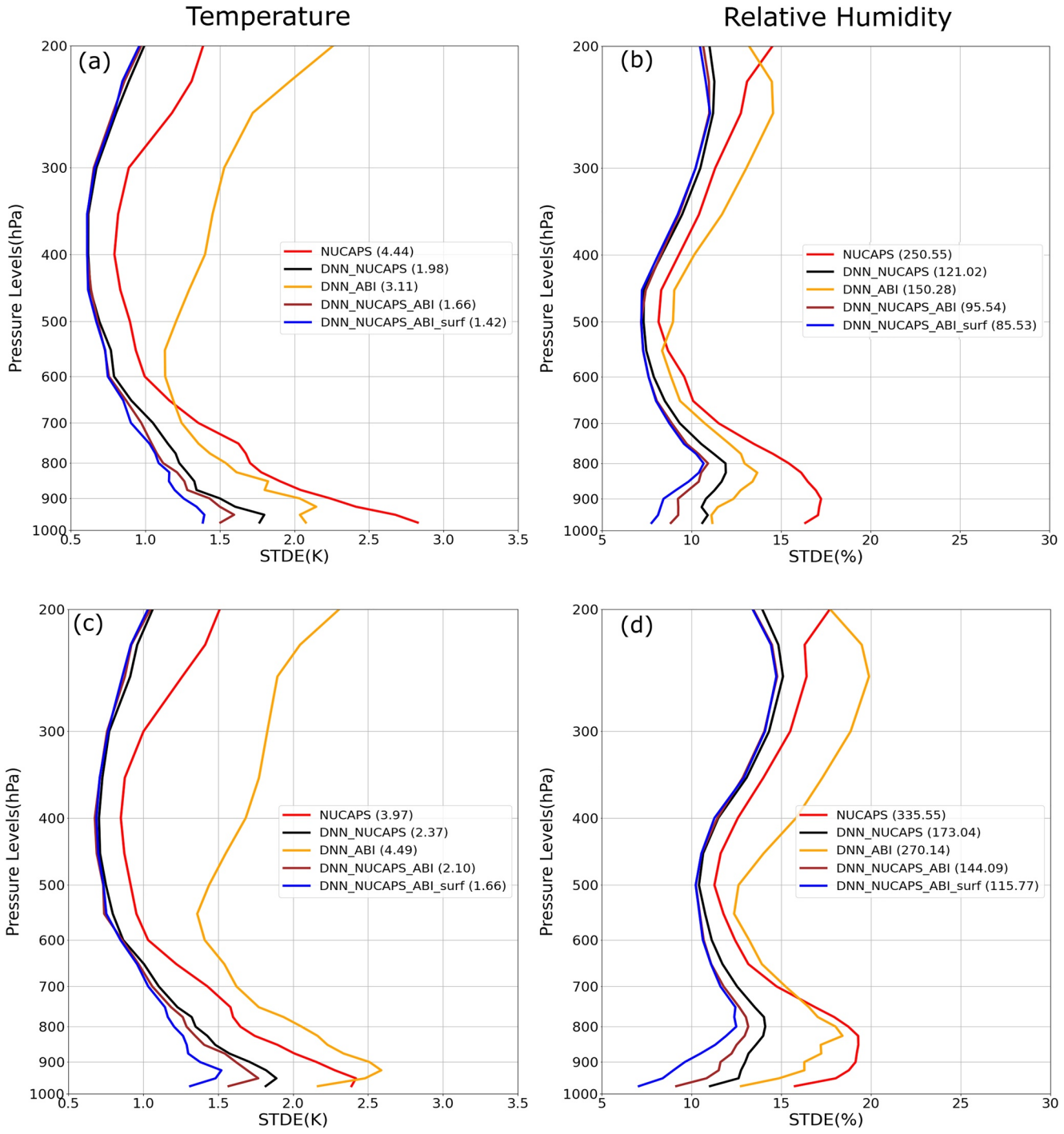

Figure 3. The STDE vertical profiles for (a), (c) Temperature and (b), (d) Relative humidity (RH) of (red) NUCAPS and outputs from various experiments versus ERA-5. (Black) DNN_NUCAPS is for the experiment including only NUCAPS profiles as predictors. (Orange) DNN_ABI is for the experiment including only ABI related predictors. (Brown) DNN_NUCAPS_ABI is the experiment including both NUCAPS and ABI predictors, and (Blue) DNN_NUCAPS_ABI_surf is the experiment in which all data sources are included. The randomly selected 20% independent validation data set from April/May/June 2018 for (a), (b) clear sky samples, and (c), (d) partially cloudy samples are used. Shown in the brackets are the values of mean variances (MV) of lower levels from 700 hPa to surface for temperature and RH, respectively. ABI, Advanced Baseline Imager; NUCAPS, NOAA Unique Combined Atmospheric Processing System; STDE, standard deviation of error.

ures 3b and 3d. This outcome is due to the fact, that the performance of ABI alone under cloudy conditions is not as good as that in clear skies. Both temperature and moisture show an increased RMSE in partly cloudy conditions. The degradation in partly cloudy conditions most likely comes from the reduced representativeness of the mean clear sky ABI observations within NUCAPS FORs. As the clear parts are seldom evenly distributed within an FOR, when the clear portion only covers a part of FOR, using such clear-sky information to represent the entire FOR will introduce some uncertainties. However, this reduced representativeness of the ABI clear observations still has a positive impact on enhancing the NUCAPS soundings in Figures 3c and 3d with the help of the DNN algorithm.

Included as a sensitivity experiment, the DNN_ABI test, with only ABI BTs and the corresponding STDs from the seven channels described in Table 1, along with the LZA ($\sec(\theta)$) included as predictors, shows an interesting error profile compared with the original NUCAPS. For temperature, smaller STDEs are seen in the lower levels and exceed NUCAPS before it becomes larger with height and much larger than NUCAPS in the higher levels. The RH STDE profile shows similar behavior. These results indicate that the ABI does not contain as much sounding information as CrIS and ATMS; even with a powerful DNN, the upper troposphere still sees larger uncertainty. More importantly, in the lower levels, NUCAPS does not show any advantage from the high vertical resolving power from CrIS, possibly due to complications from the surface. It is therefore possible to use ABI observations to improve the lower levels of NUCAPS soundings.

Three types of ABI bands are used in this study: the three water vapor (WV) bands (6.2, 6.9, and 7.3 μm), the single CO_2 band (13.3 μm), and the three window bands (10.3, 11.2, and 12.3 μm). It is reasonable to use the three window bands in this study, since they are mostly sensitive to the lower atmospheric temperature and moisture. The use of WV bands, mostly sensitive to middle and upper troposphere moisture, and the CO_2 band, mostly sensitive to lower atmosphere temperature but peaking higher than the window bands, is not as straightforward. In order to investigate the relative contributions of ABI's WV and CO_2 bands on improving the soundings, three experiments are conducted by denying those ABI bands from the predictors. To eliminate the effect from the other predictors, only the predictors from ABI are included. Figure 4 shows the relative impact of ABI channels on the DNN retrieval performance. Drawn from the temperature profiles, the CO_2 band has a much larger contribution to temperature when compared to the three WV channels. This is consistent with the fact that this CO_2 band is mostly sensitive to the lower atmosphere temperature while the three water vapor bands are mostly sensitive to the mid-to-upper troposphere moisture (see Figure 4 in Schmit et al., 2008). For relative humidity, it is interesting to see that the CO_2 band contribution is substantial at levels below 800 hPa, while water vapor channels have more impact on the levels above that. Not only is the CO_2 band sensitive to lower level moisture, it is more sensitive to lower level moisture than all window bands except the 12.3 μm (Li et al., 2020). Using the experiment where only the ABI window bands are used (no WV and CO_2 bands), the ABI WV bands has moderate variance reductions of 8.2% for temperature and 4.8% for RH in the lower levels; and the CO_2 band has more significant variance reductions of 60.5% for temperature and 42% for moisture.

3.3. Validation Against ERA-5 From 2017

The combination of NUCAPS, ABI, and RTMA with the DNN model has been shown to enhance the lower level temperature and moisture profiles when validated using the 20% independent validation set from 2018. It remains unclear how the model performs using another more independent data set, that is, data from 2017. The data for April/May/June 2017 is collocated following the same pre-processing procedure. With the ABI cloud mask data for April 1–19 missing, the total sample size is about 210,000, substantially less than that from 2018. As with the 2018 data set, the 2017 data set is separated into clear-sky and partly cloudy conditions, respectively, and model validations are performed accordingly. See Table 4 for the valid sample size for each level for both clear sky and partly cloudy conditions, respectively. Based on Tables 2–4, the partly cloudy region has a sample size about 4 times that of clear skies. It is therefore essential to extend the DNN model to partly cloudy conditions to increase the data yield.

The validation results for clear sky and partly cloudy samples are shown in Figures 5a and 5b, and Figures 5c and 5d, respectively. In general, combining NUCAPS, ABI, and RTMA with the DNN shows improvements in the profiles, with smaller STDE and mean bias compared with the original NUCAPS. Similar to the results for 2018, the improvements from the DNN based fusion model appear to be more significant

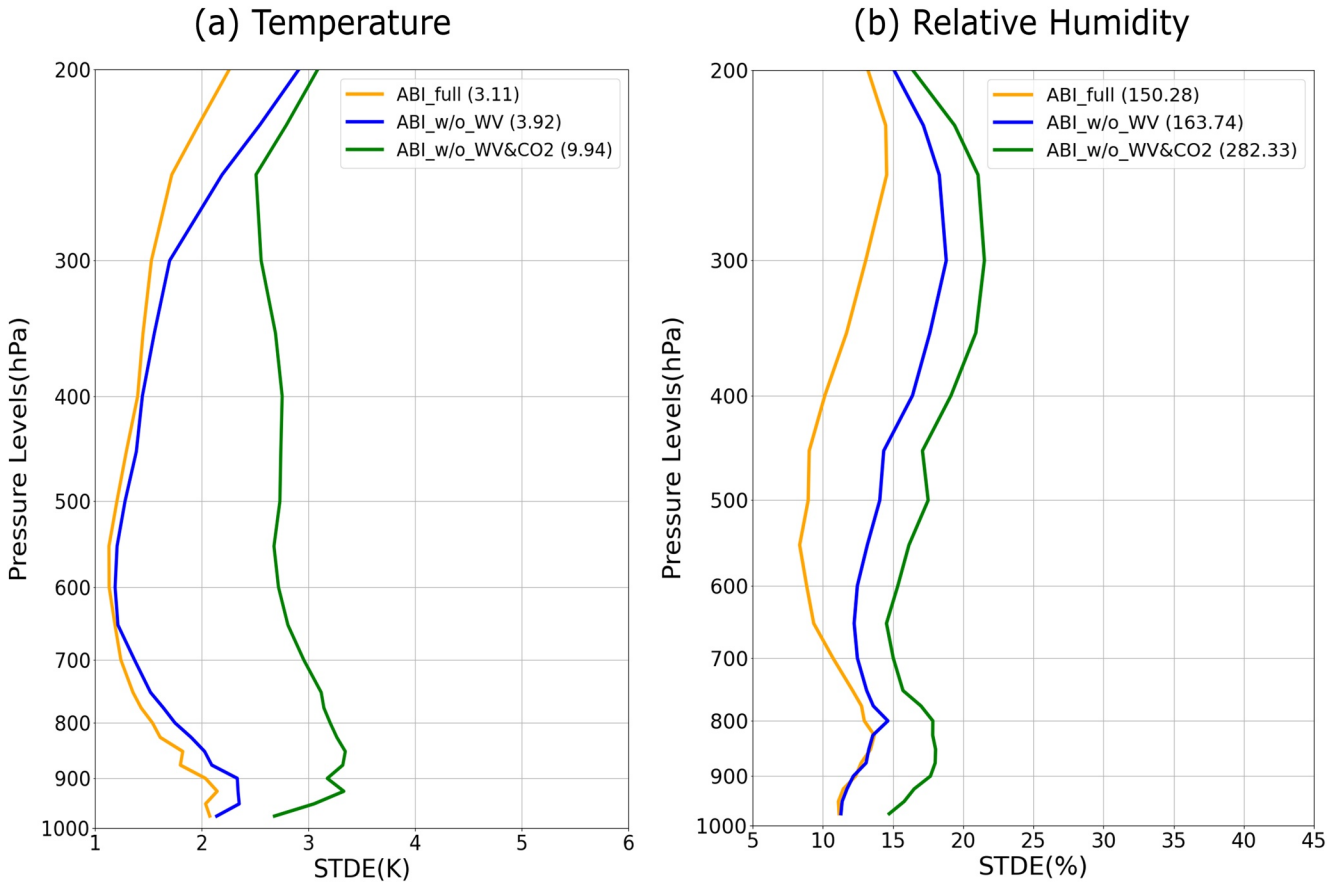

Figure 4. The STDE vertical profiles for (a) temperature and (b) relative humidity from various experiments versus ERA5. (Orange) ABI_full is for the experiment including only all seven ABI channels mentioned in this study. (Blue) ABI_w/o_WV is for the experiment without water vapor channels (6.2, 6.9 and 7.3 μm). (Green) ABI_w/o_WV&CO_2 is for the experiment without water vapor channels and the CO_2 channel (13.3 μm). Shown in the brackets are the values of mean variances (MV) of lower levels from 700 hPa to the surface for temperature and RH, respectively. ABI, Advanced Baseline Imager; RH, relative humidity; STDE, standard deviation of error.

in the lower atmosphere. The variance reductions in the lower levels are 43.0% for temperature and 52.2% for RH in clear skies, and 45.8% and 50.8%, respectively, in partly cloudy conditions. These numbers are not as profound as those from the validation using the 2018 data set, but still quite significant. In addition, for the upper levels, especially the levels above 500 hPa, the improvements appear not as substantial as (even neutral for RH) those from 2018. This result makes sense as the validation is performed with a completely independent data set from a different year, and the sample sizes are almost as large as those used for training. It is important to point out that the validation using the independent 2017 data set is necessary

Table 4
Valid Sample Size for Clear Sky and Partly Cloudy Validation for 2017 at Each Pressure Level

Pressure level (hPa)	Clear sky sample size	Partly cloudy sample size	Pressure level (hPa)	Clear sky sample size	Partly cloudy sample size
650 and levels above	35,400 for each level	176,867 for each level	850	33,025	167,362
700	35,372	176,769	875	31,939	163,483
750	35,188	176,075	900	31,041	159,667
775	34,923	175,123	925	29,900	153,004
800	34,293	172,564	950	27,426	137,772
825	33,634	170,101	975	20,179	96,760

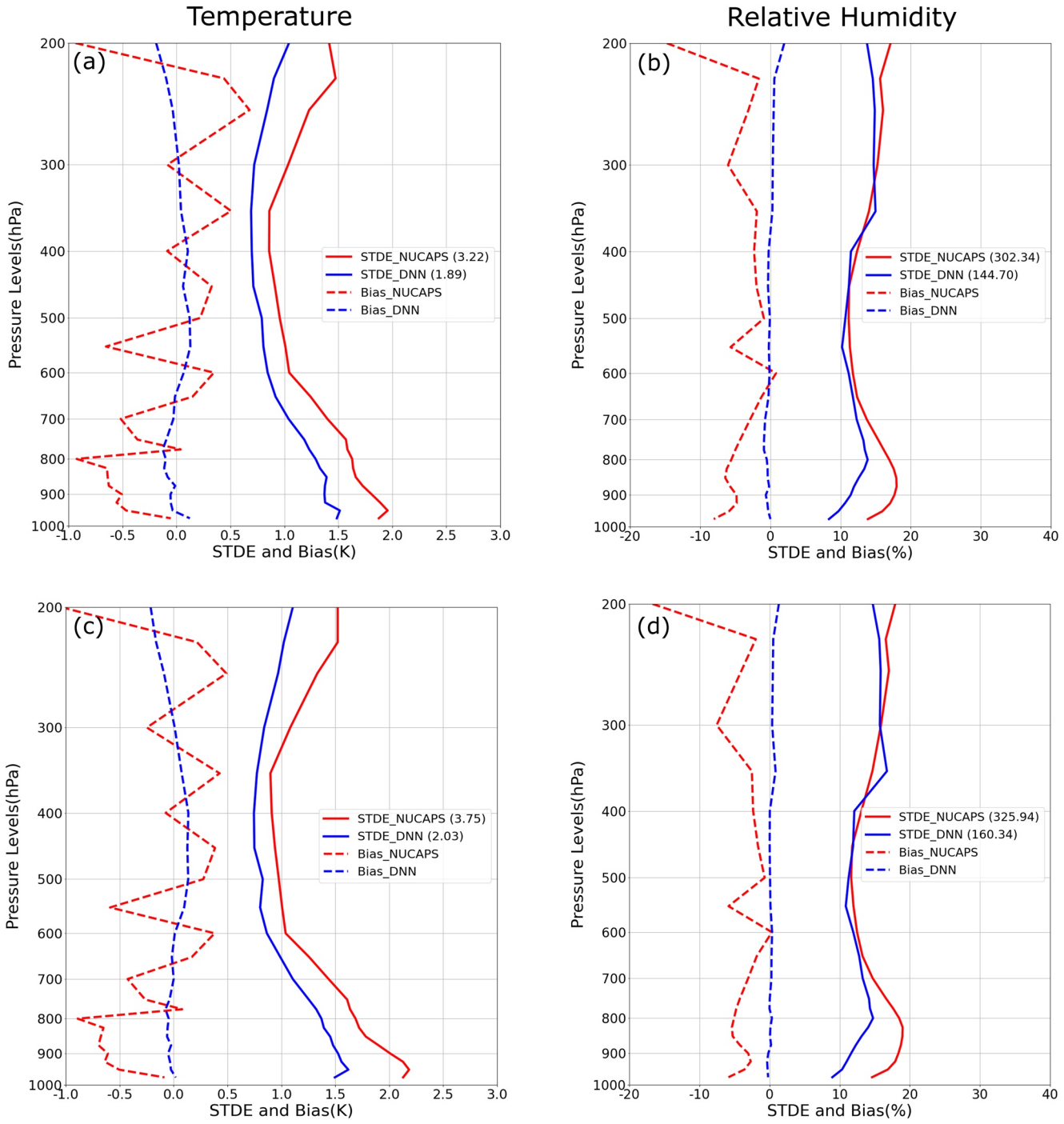

Figure 5. Vertical profiles of the STDE (solid lines) and mean bias (dashed lines) for (a), (c) temperature, and (b), (d) relative humidity of (red) NUCAPS and (blue) DNN model outputs versus ERA-5 from the 2017 data set under (a), (b) clear sky conditions, and (c), (d) partly cloudy conditions. Shown in the brackets are the values of mean variances (MV) of lower levels from 700 hPa to the surface for temperature and RH, respectively. NUCAPS, NOAA Unique Combined Atmospheric Processing System; RH, relative humidity; STDE, standard deviation of error.

to ensure no overfitting by the DNN model, as the differences in the results from 2017 to 2018 are not that large. These results reflect the general ability of the DNN model trained with 80% of the 2018 data. However, in applications, the DNN model should be ideally trained with a sample size as large as possible to increase its representativeness.

Table 5
Validation Sample Size for NUCAPS Collocated With RAOB at Each Pressure Level

Pressure level (hPa)	Sample size 2018	Sample size 2017	Sample size total
200–800	55	16	71
825–925	52	13	65
950	32	13	45

NUCAPS, NOAA Unique Combined Atmospheric Processing System; RAOB, radiosonde observations.

4. Validation Against RAOBs

The performance of the DNN-based algorithm is validated in the previous section using independent ERA-5 data. Since the ERA-5, a gridded reanalysis data set, is used as truth when training the DNN model, it is necessary to conduct further validation using independent observations. For this purpose, the collocated RAOBs collected at the NPROVS from April, May and June of 2017 to 2018 over CONUS are used to assess the performance of the DNN results. RAOBs have been widely used for satellite sounding evaluation (Nalli et al., 2018; Sun et al., 2010, 2017, 2018). To minimize the errors introduced from the mismatch of the observation times, the threshold for time difference is set as two hours between the launch time of the radiosonde and the satellite overpass. The average time difference of the filtered samples is 60.42 min, with a minimum value of 1 min and a maximum value of 119 min. The average spatial distance is 35.36 km, with a minimum value of 1.10 km and a maximum value of 115.46 km. The sample size for each level is listed in Table 5. As the number of collocated RAOBs is relatively small, samples under clear sky and partly cloudy conditions are put together into the statistics after applying their respective models.

Shown in Figure 6 are the profiles of statistical STDE and mean bias of the original NUCAPS observations and the enhanced results compared with corresponding collocated RAOBs. As the output of the DNN based data fusion model in this study uses the ERA-5 pressure levels, the NUCAPS and RAOB samples are both vertically interpolated to the ERA-5 levels listed in Table 4 for easier comparison. As shown in Figure 6, while the STDE and bias between NUCAPS and RAOBs appear to be larger than those between NUCAPS

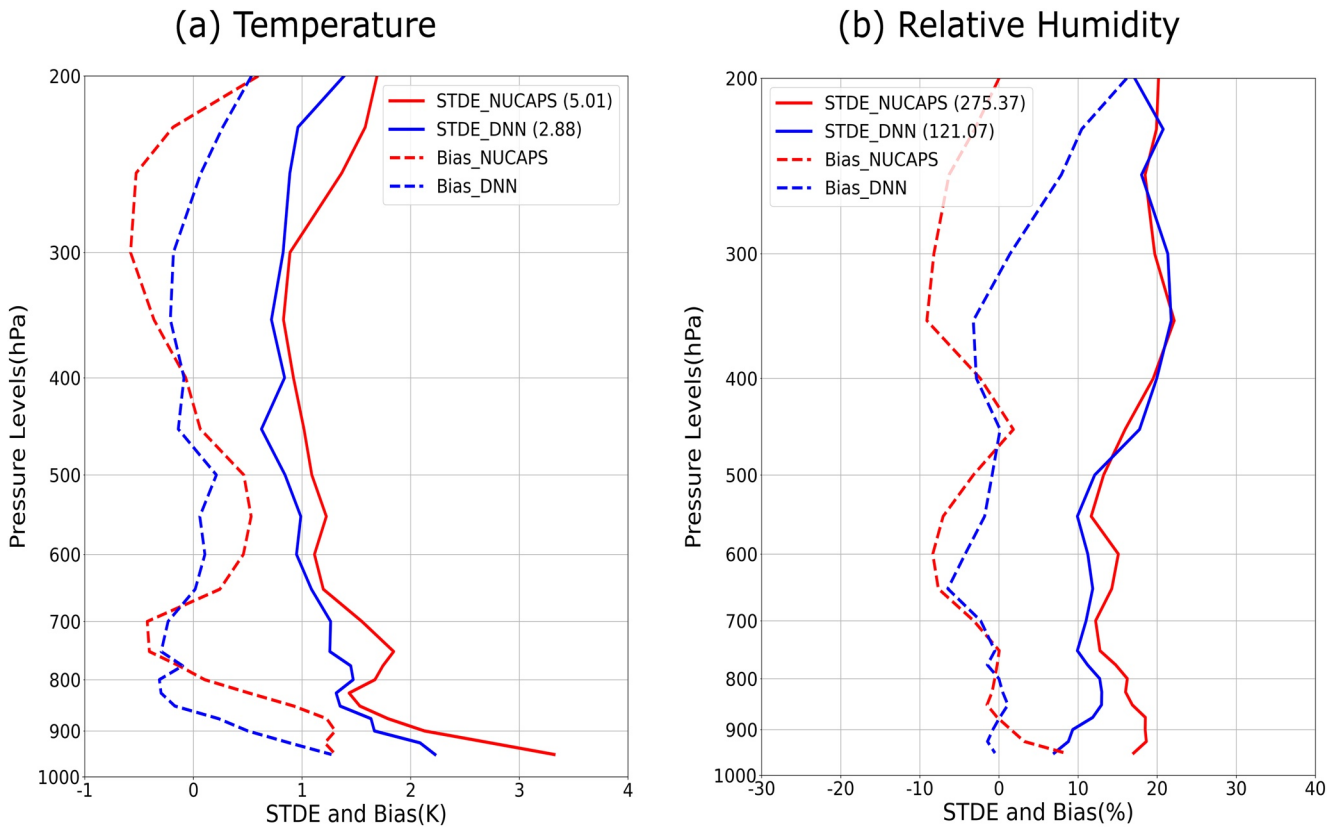

Figure 6. Vertical profiles of the STDE (solid lines) and mean bias (dashed lines) for (a) temperature, and (b) relative humidity of (red) NUCAPS and (blue) enhanced NUCAPS soundings against collocated RAOBs in both 2017 and 2018. Shown in the brackets are the values of mean variances (MV) of lower levels from 700 hPa to the surface for temperature and RH, respectively. NUCAPS, NOAA Unique Combined Atmospheric Processing System; RAOB, radiosonde observations; RH, relative humidity; STDE, standard deviation of error.

and ERA5, in the lower levels due to temporal differences, the enhanced retrievals from the DNN based fusion model are found to have smaller differences compared with the NUCAPS with RAOBs as reference, for both temperature and moisture. Smaller STDEs for temperature and relative humidity in the lower levels indicate improvement in the soundings made by fusing data with the DNN algorithm. Similar to the statistics compared against ERA-5, bigger improvements are recognized in the lower levels where the original STDE is relatively larger. The mean variance reductions of the lower levels are 42.6% for temperature and 56.0% for RH, respectively. These findings are consistent with those from the validations using ERA-5, although the magnitude of the improvement is smaller.

We noted similar findings for the upper levels in the validation against ERA-5 from 2017 although the performance is not as good as for the lower levels. One thing we noticed is the neutral performance at levels above 500 hPa for relative humidity. The main reason for that result is the 2 h threshold used for the collocating the NUCAPS and RAOBs. The large time difference degrades the statistics for the whole profile. However, since the fusion of data with DNN improves the lower levels more than high levels, the degradation appears to be less profound in the lower levels. To verify that conclusion, smaller time thresholds are used for NUCAPS/RAOB collocations, resulting in more substantial improvements in the lower levels, for both temperature and RH. And the smaller the time threshold is, the more substantial the improvement. For example, when 30 min is used as the threshold, the mean variance reductions are 57.5% for temperature and 69.7% for RH, much better than those shown in Figure 6. It should be pointed out that the sondes with a 30-min collocation window are actually the S-NPP synchronized dedicated sondes. They are primarily the GRUAN processed RS92 and RS41 with standard operational processing. Those two sondes are better in accuracy, especially for humidity data, than the RS92 with standard operational processing that are included for validation shown in Figure 6. Nevertheless, since the sample size is only 17, such results are not considered statistically significant and thus not shown in this study.

5. Application Demonstration

5.1. Sounding Cases

In this section, we use two sounding profiles, one in clear sky, and the other in partly cloudy conditions, to demonstrate the performance of fused data when depicting the vertical structures of the lower atmosphere. Along with the original NUCAPS profile and the output from the DNN-based fusion model, the collocated RAOB profile from NPROVS is also included for reference. Figure 7a shows the profile under clear-sky conditions on April 26, 2018, with the NUCAPS observation time at 1742 UTC, centered at (−76.84°E, 39.03°N), and the collocated RAOB sounding launch at 1633 UTC located at (−76.87°E, 39.05°N). The temperature profile from NUCAPS shows a relatively deep boundary layer from the surface to around 800 hPa, capped by a reasonably deep entrainment layer reaching ~650 hPa. The RAOB, on the other hand, shows a much shallower boundary layer topped at 900 hPa, with the capped entrainment layer reaching about 750 hPa. This leads to a significant cold bias from the NUCAPS between the levels of 850 hPa and 650 hPa. The output from the data fused model shows a much improved profile, depicting the depth of the boundary layer and the entrainment layer much better than the NUCAPS. In terms of moisture, compared to the smooth profile in the lower levels with NUCAPS, the two dry peaks around 750 hPa and 650 hPa from the RAOB are better characterized by the model.

Another case is shown in Figure 7b under partly cloudy conditions observed on April 18, 2018, with NUCAPS at 0710 UTC, centered at (−76.58°E, 39.17°N), and the collocated RAOB with a launch time of 0728 UTC, located at (−76.87°E, 39.05°N). From the RAOB temperature profile, a strong inversion can be seen from 850 to 750 hPa. However, the NUCAPS does not show this inversion at all. The output from the data fused model not only reproduces the inversion, but also successfully corrects the warm bias under the bottom of the inversion at 850 hPa. For the moisture profile, the model output shows its improvement by reproducing the dry layer between 800 hPa and the surface, in contrast to the relatively large wet bias shown by the NUCAPS soundings. These two cases clearly show that the reduced accuracy from the NUCAPS is not limited to the boundary layer. The DNN regression model is able to correct the boundary layer, the entrainment layer (including the temperature inversion), and above. This enhancement will provide much needed confidence to forecasters to use the enhanced sounding products.

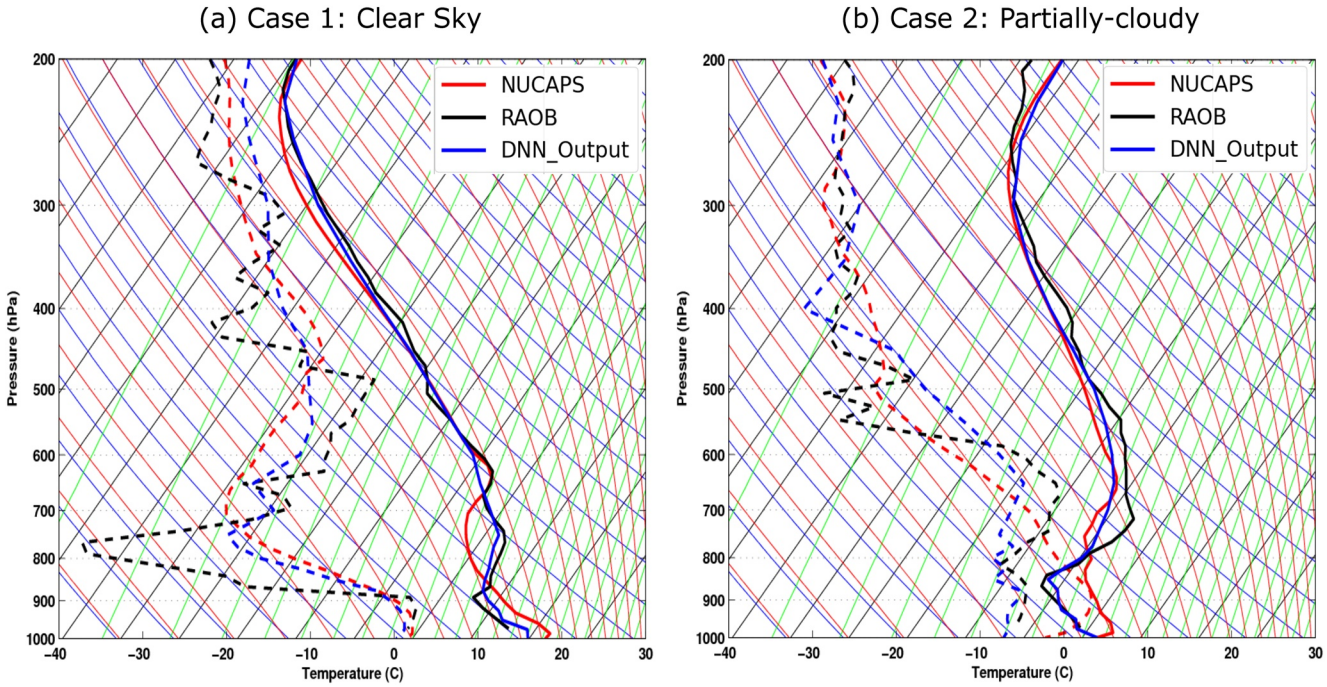

Figure 7. Temperature (solid lines) and relative humidity (dashed lines) profiles plotted in skew-t diagrams for (a) clear sky case with NUCAPS at 1742 UTC, FOR centered at (−76.84°E, 39.03°N), RAOB launched at 1633UTC from (−76.87°E, 39.05°N), and output of the DNN data fusing model based on the NUCAPS sounding on April 26, 2018, and (b) partly cloudy case with NUCAPS sounding at 0710 UTC, FOR centered at (−76.58°E, 39.17°N), and RAOB at 0727UTC from (−76.87°E, 39.05°N) on April 18, 2018. DNN, deep neural network; NUCAPS, NOAA Unique Combined Atmospheric Processing System; RAOB, radiosonde observations.

5.2. Nowcasting a High Impact Weather Event

To demonstrate the applications of the enhanced soundings on high impact weather (HIW) events, typical convective cases are presented. On June 18, 2017, several severe storm events occurred in the eastern United States, including damaging hail and strong winds. According to the Storm Prediction Center (SPC) storm reports, the atmospheric conditions were warm and moist in the lower levels, conducive to convective initiation, leading to the storms that evening (https://www.spc.noaa.gov/exper/archive/events/). During that day, the eastern United States was behind a warm front and in advance of a strong cold front moving eastward. The surface dewpoint in that region was consistently high (upper 70°s–80°s°F), creating a relatively moist boundary layer before the storms occurred. According to the mesoscale analysis from the SPC severe weather events archive (https://www.spc.noaa.gov/exper/archive/event.php?date=20170618), the lower atmosphere on the east coast was stable at 1200 UTC with relatively low surface-based convective available potential energy (SBCAPE) and high convective inhibition (CIN) values; as the day progressed the CIN continued to erode while SBCAPE grew larger. Until 1800 UTC, the SBCAPE along the East Coast increased to more than 2,000 J·kg^{-1} and kept at high values to around 2200 UTC. As seen from Figure 8a, the larger SBCAPE distribution along the coast is well represented by the ERA-5 hourly analysis at 1800 UTC. This distribution is consistent with the mesoscale analysis from SPC (not shown, but available on the SPC website mentioned above), indicating that ERA5 does well in reproducing the convective environment for this case.

The NUCAPS overpass was around 1827 UTC. Calculated from the temperature and moisture profiles from NUCAPS soundings, the SBCAPE for overpasses from 1827 to 1839 UTC is shown in Figure 8b. The NUCAPS shows zero or low values for the majority of FORs in the preconvection area. Compared to the ERA5 SBCAPE plot in Figure 8a, NUCAPS did not capture the large SBCAPE values in this area. The SBCAPE calculated from the output of the DNN model is shown in Figure 8c. Compared with NUCAPS, the data fused DNN model in this study depicts the high SBCAPE values prior to any convection much better. Similar phenomenon is also seen from the spatial distribution of LI, as shown by Figure 9. Areas where the LI

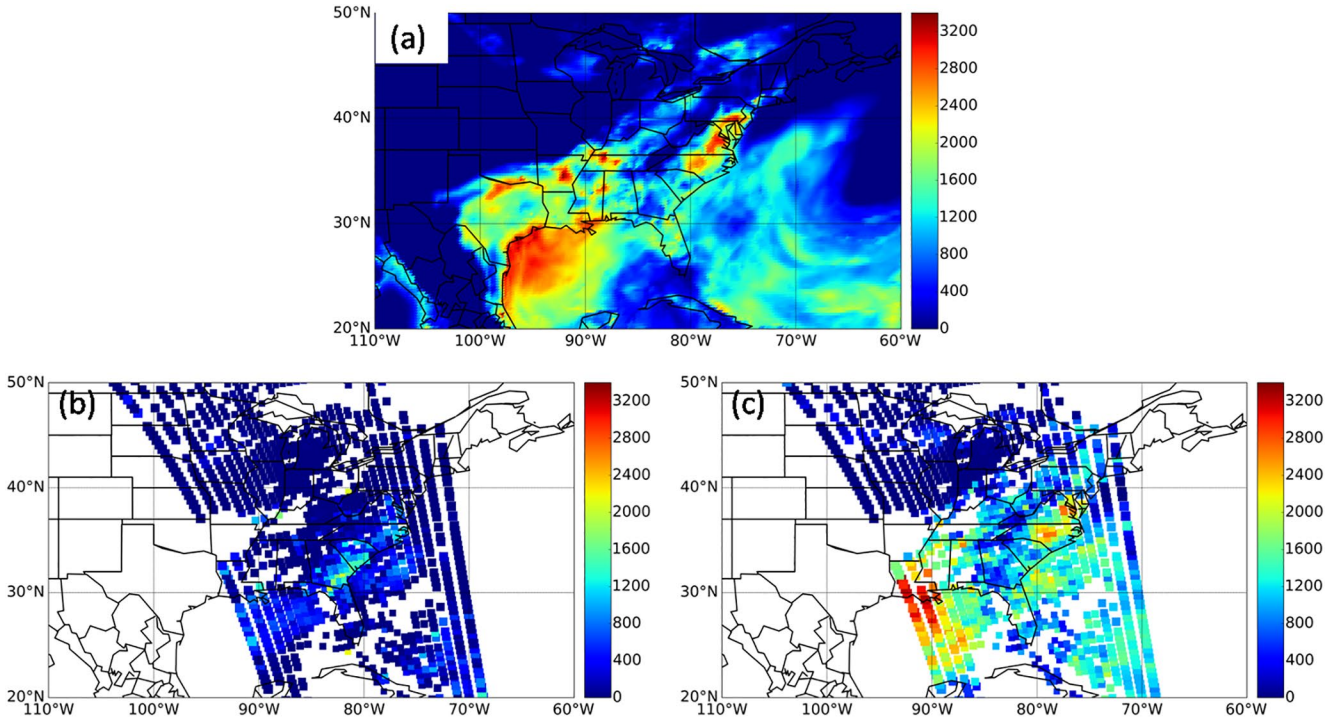

Figure 8. Surface-based convective atmospheric potential energy (SBCAPE) ($J \cdot kg^{-1}$) from (a) ERA5 Reanalysis at 1800 UTC, and calculations from (b) NUCAPS, and (c) DNN model outputs for NUCAPS overpasses from 1827 to 1839 UTC, June 18, 2017. DNN, deep neural network; ERA5, ECMWF re-analysis version 5; NUCAPS, NOAA Unique Combined Atmospheric Processing System.

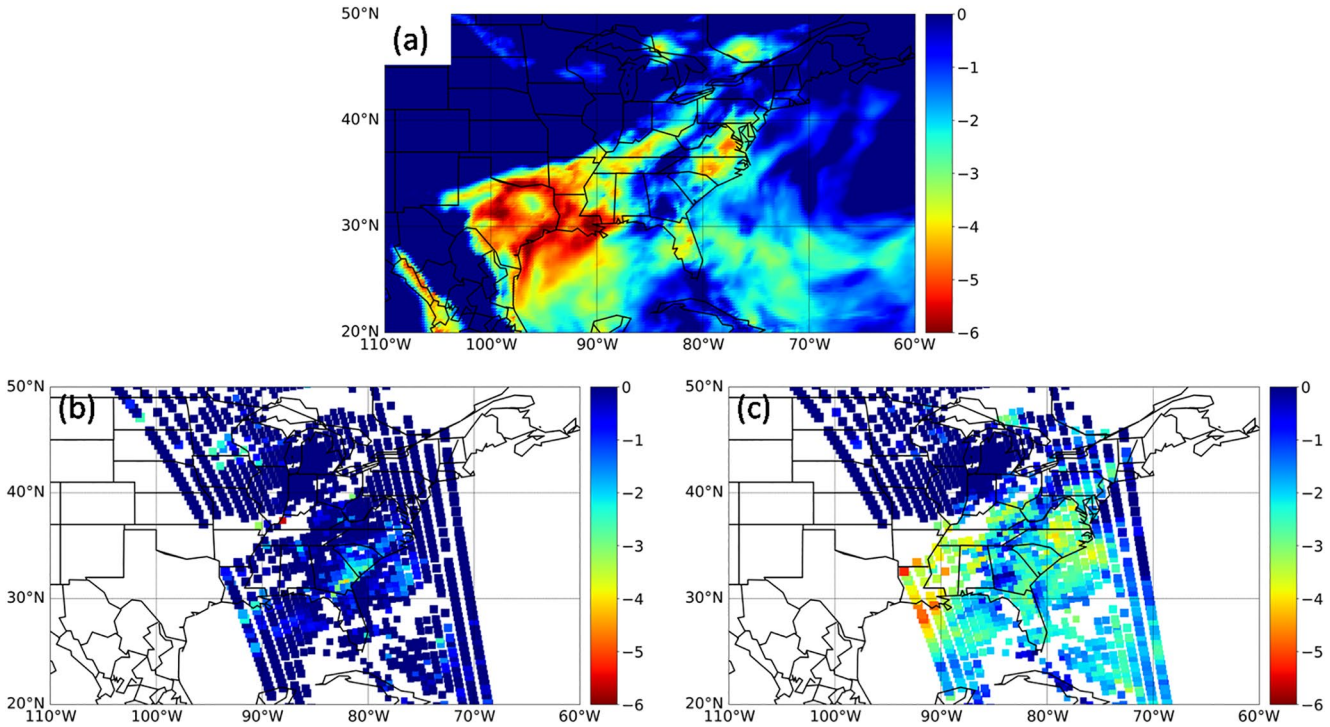

Figure 9. Lifted Index (LI) (K) from (a) ERA5 Reanalysis at 1800 UTC, and calculations from (b) NUCAPS, and (c) DNN model outputs for NUCAPS overpasses from 1827 to 1839 UTC, June 18, 2017. DNN, deep neural network; ERA5, ECMWF re-analysis version 5; NUCAPS, NOAA Unique Combined Atmospheric Processing System.

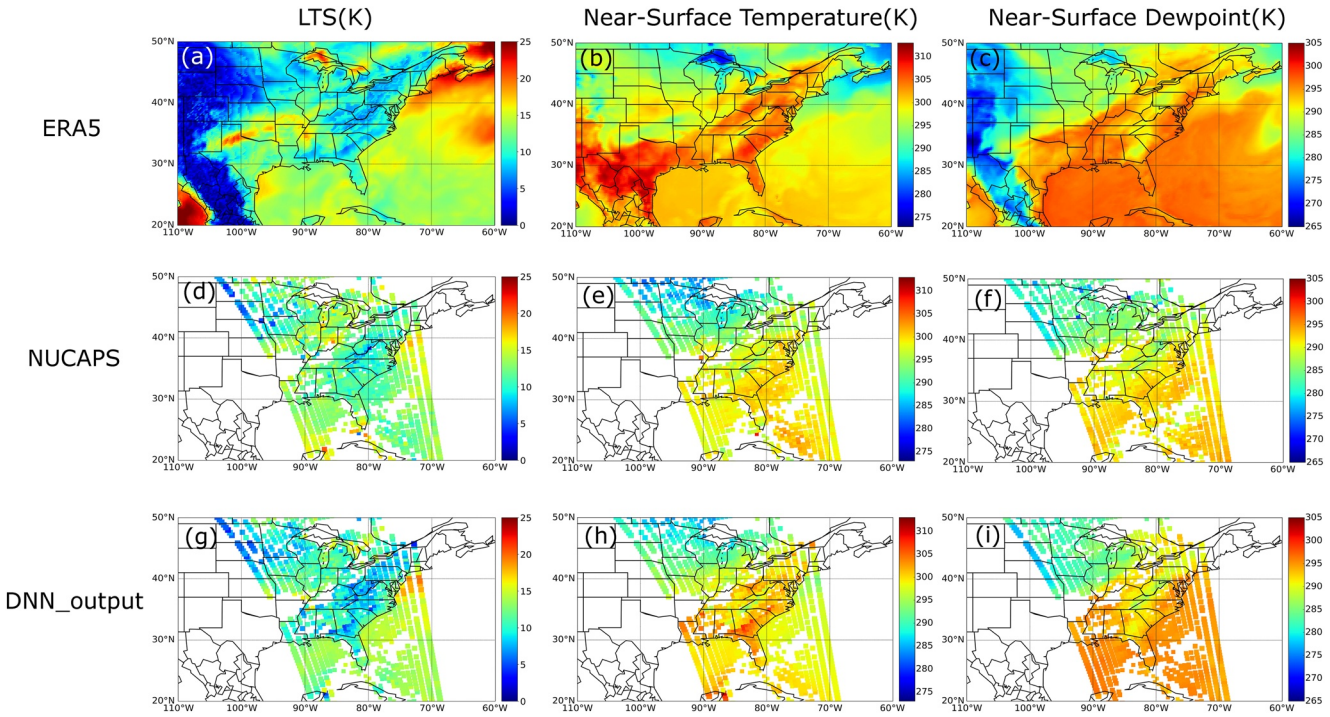

Figure 10. Comparison of derived fields of (a–c) ERA5 at 1800 UTC, and estimations from (d–f) NUCAPS, and (g–i) DNN model outputs for NUCAPS overpasses from 1827 to 1839 UTC on June 18, 2017. The left column is for LTS (K); the middle column is for Temperature (K) at the lowest level above the surface; and the right column is for Dewpoint (K) at the lowest level above the surface. DNN, deep neural network; LTS, lower troposphere stability; NUCAPS, NOAA Unique Combined Atmospheric Processing System.

value is less than −2 K are generally considered unstable and thunderstorms may occur. Compared to the ERA5 plot in Figure 9a, the NUCAPS soundings tend out to be too stable by showing zero or positive values for most FORs while failing to capture the lower LI values along the eastern coast and near Louisiana. The DNN output shown in Figure 9c succeeded in reproducing the spatial distribution of LI prior to convections, like what is done for CAPE. The enhanced soundings will provide additional reliable satellite based observations to help forecasters with severe storm nowcasting (situational awareness) and forecasting.

As for the profiling of low level atmosphere, we take lower troposphere stability (LTS), defined as the potential temperature difference between 700 hPa and surface (Klein & Hartmann, 1993), as a description of the thermodynamic characteristics of the lower atmosphere. As is shown in Figure 10, NUCAPS in this case overestimates the LTS over land along the eastern coast and underestimates the LTS off the coast. The LTS distribution depicted by DNN output, as shown in Figure 10g is more consistent with ERA5 (Figure 10a) when compared with that from NUCAPS (Figure 10d). This comparison shows to a certain extent the ability of the DNN model for enhancing the accuracy of lower atmosphere thermodynamics. Also shown in Figure 10 are the near-surface temperature and dewpoint fields represented by the values from the lowest pressure levels above the surface for each point. NUCAPS in this case underestimates both temperature and moisture at the surface level when compared with ERA5, thus affecting the depiction of convective indicators like SBCAPE, while the data fusion with DNN model corrects the deviation to a certain extent.

6. Summary and Future Work

This study applied a DNN-based data fusion technique to develop a retrieval model capable of enhancing the lower levels of operational NUCAPS sounding products through combining NUCAPS, GOES-16 ABI and RTMA. Three months of data from April to June 2018, including NUCAPS, ABI radiances, RTMA and ERA-5 within the CONUS region are collocated and split into two data sets: 80% for training and the remaining 20% for validation. The STDE, mean bias and mean variances are analyzed to evaluate the performance of the fusion model with the validation data set. Results show that the temperature and moisture

profiles of enhanced soundings are improved by more than 30% in STDE at low levels under both clear-sky and partly cloudy conditions. The improvement, in the mean variance reductions of lower levels, is 68.1% for temperature and 65.9% for relative humidity in clear skies, and 58.1% for temperature and 65.5% for relative humidity in partly cloudy conditions. The extension from clear sky to partly cloudy conditions significantly increases the data yield by four times.

Relative impacts from ABI and RTMA are evaluated through sensitivity studies, where certain predictors are excluded in the OSEs. Results show both the ABI and RTMA have substantial impacts on reducing the uncertainties in the lower levels. The ABI is found to have lower level mean variance reductions of 7.2% for temperature and 10.2% for RH in clear skies. In partly cloudy skies, the improvements are slightly less profound, but still comparable, with 7.1% for temperature and 8.6% for RH. The RTMA has a slightly smaller impact with mean variance reductions of 5.5% for temperature and 3.9% for RH in clear skies, and 10.9% for temperature and 8.4% for RH in partially cloudy skies. However, the DNN itself is found to have the largest contribution in reducing the lower level mean variances: 55.4% for temperature and 51.7% for RH in clear skies, and 40.1% for temperature and 48.4% for RH in partially cloudy skies. In addition, the ABI contributes to most levels from the lower to upper troposphere, while RTMA only provides contributions to the lower levels. Among the seven ABI spectral bands used in this study, the CO_2 channel and the window channels provide the major contribution to temperature and low-level moisture, while water vapor channels provide a larger contribution to moisture above 800 hPa.

With the training data set from 2018, the collocated data set from 2017 is used for further independent validation of the DNN model. Similar but less profound improvements are found, indicating the DNN model is not overfitting to the 2018 training data set. However, for applications, it is ideal to use a training data set as large as possible to increase its representativeness. The RAOBs collected at NPROVS for 2017 and 2018 are also used to validate the performance of the developed DNN model, as additional independent validation. Results show that the performance of the DNN model for the lower levels is stable and sound. The evaluation is affected significantly by the time difference between NUCAPS and RAOB. Smaller time differences do lead to larger improvements in the DNN model, but with a smaller sample size, which is not statistically significant.

Two individual profiles are highlighted to demonstrate the significant improvement in the enhanced soundings for depicting the low level atmospheric structures when compared with RAOB. Applying the enhanced soundings on HIW event nowcasting is also demonstrated through a convection case in eastern US on June 18, 2017. The enhanced soundings captured the moist boundary layer and surface-based CAPE high values better than the original NUCAPS in the preconvection environment. The enhanced soundings from combining data with the DNN based fusion model is expected to increase forecasters' confidence in severe storm nowcasting/forecasting.

Future work includes refining the DNN-based data fusion model to include RTMA wind observations. Since the temperature and moisture vertical structures in the low levels are also associated with atmospheric movement, ABI tendency information (temporal change) will also be tested. Also, more data covering the entire year will be included in the training process to enhance the representativeness of the model for real-time applications over different seasons. Furthermore, we will update our model with data from the latest NUCAPS v3.0 and extend the applications and analysis to products from NOAA-20, the new JPSS satellite (Zhou et al., 2016). In addition, it is important to help users understand the performance in different surface types, which may result from different PBL types and RTMA accuracy.

Compared with the traditional 1DVAR approach, this DNN technique offers two significant advantages. First, it is capable of fusing multiple sources of observations with no need of forward operator. This is critically important for special information such as sub-FOR homogeneity that has no forward operator. Another important advantage is the high computational efficiency. Once the training is done, which could be time consuming, the application is much faster. There are also several limitations of this DNN technique. The main limitation is that it highly depends on the training data set. If the training data set is complete and is well representative of all possible cases, it can be applicable to all different situations. It is therefore important to include more data covering the entire multiple years into training to further enhance the model's representativeness. The second limitation is that machine learning based technique might fail to generate

physically meaningful results, that is, oversaturation or unrealistic lapse rate. Therefore, additional quality control is needed to remove such results.

One limitation of this work is that the enhanced NUCAPS soundings still have the same low spatial resolution of 45 km and temporal resolution of 12 h. This may limit the application of the enhanced NUCAPS in short lived small scale weather events. It is therefore important to investigate how to fuse different sources of data, especially from GEO and LEO, while maintaining the high vertical resolution from LEO and high spatial and temporal resolutions from GEO. Smith et al. (2020) have shown promising potentials of such data fusion techniques for NWP applications, especially precipitations and tornado occurrence of LSS.

Combining all available information from geostationary and polar orbit satellites as well as NWP, through data assimilation (Smith et al., 2020), is an effective approach for nowcasting and forecasting products. While this study provides another approach on combining information based on statistical model for nowcasting applications. It is worth noting that this enhanced NUCAPS product can be produced at near global coverage for NRT applications, for example, through combining NUCAPS from Direct Broadcast data using Community Satellite Processing Package and ABI, Advanced Himawari Imager, Advanced Geosynchronous Radiation Imager, and Advanced Meteorological Imager.

Data Availability Statement

The NUCAPS environmental data records (EDRs) and GOES-16 ABI cloud mask products are downloaded from the Comprehensive Large Array-data Stewardship System of NOAA at https://www.avl.class.noaa.gov/saa/products/welcome. GOES-16 ABI radiance data are provided by the University of Wisconsin-Madison SSEC Data Center (https://www.ssec.wisc.edu/datacenter). RTMA is downloaded from National Digital Guidance Database (NDGD) at https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/national-digital-guidance-database-ndgd. ERA-5 data are obtained from ECMWF at https://www.ecmwf.int/en/forecasts/datasets/browse-reanalysis-datasets.The authors would like to thank the following colleagues and GRUAN for their contributions to the radiosonde data collection and management effort: A. Reale, L. Borg, B. Demoz, D. Holdridge and J. Mather.

Acknowledgments
The authors would also thank L. Avila for English proofreading. This work is partly supported by SSEC internal funding and the NOAA Quantitative Observing System Assessment Program (QOSAP) nowcasting OSSE project. Z. Ma is supported by scholarship from University of Chinese Academy of Sciences (UCAS). 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

Bengio, Y. (2009). Learning deep architectures for AI. *FNT in Machine Learning*, *2*(1), 1–127. https://doi.org/10.1561/2200000006

Bloch, C., Knuteson, R. O., Gambacorta, A., Nalli, N. R., Gartzke, J., & Zhou, L. (2019). Near-real-time surface-based CAPE from merged hyperspectral IR satellite sounder and surface meteorological station data. *Journal of Applied Meteorology and Climatology*, *58*(8), 1613–1632. https://doi.org/10.1175/JAMC-D-18-0155.1

Bodeker, G. E., Bojinski, S., Cimini, D., Dirksen, R. J., Haeffelin, M., Hannigan, J. W., et al. (2016). Reference upper-air observations for climate: From concept to reality. *Bulletin of the American Meteorological Society*, *97*(1), 123–135. https://doi.org/10.1175/bams-d-14-00072.1

Boukabara, S.-A., Krasnopolsky, V., Stewart, J. Q., Maddy, E. S., Shahroudi, N., & Hoffman, R. N. (2019). Leveraging modern artificial intelligence for remote sensing and NWP: Benefits and challenges. *Bulletin of the American Meteorological Society*, *100*(12), ES473–ES491. https://doi.org/10.1175/BAMS-D-18-0324.1

Clevert, D.-A., Unterthiner, T., & Hochreiter, S. (2016). *Fast and accurate deep network learning by exponential linear units (ELUs)*. Retrieved from http://arxiv.org/abs/1511.07289

Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2011). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio Speech and Language Processing*, *20*(1), 30–42.

De Pondeca, M. S. F. V., Manikin, G. S., DiMego, G., Benjamin, S. G., Parrish, D. F., Purser, R. J., et al. (2011). The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Weather and Forecasting*, *26*(5), 593–612. https://doi.org/10.1175/WAF-D-10-05037.1

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., & Vömel, H. (2014). Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. *Atmospheric Measurement Techniques*, *7*(12), 4463–4490. https://doi.org/10.5194/amt-7-4463-2014

Feltz, M. L., Borg, L., Knuteson, R. O., Tobin, D., Revercomb, H., & Gambacorta, A. (2017). Assessment of NOAA NUCAPS upper air temperature profiles using COSMIC GPS radio occultation and ARM radiosondes. *Journal of Geophysical Research: Atmospheres*, *122*(17), 9130–9153. https://doi.org/10.1002/2017JD026504

Gambacorta, A. (2013). *The NOAA Unique CrIS/ATMS Processing System (NUCAPS): Algorithm theoretical basis documentation* (p. 1). NOAA Center for Weather and Climate Prediction (NCWCP), Version.

Gambacorta, A., Barnet, C., & Goldberg, M. (2015). *Status of the NOAA Unique CrIS/ATMS Processing System (NUCAPS): Algorithm development and lessons learned from recent field campaigns*. In Proceeding of the ITSC.

Gambacorta, A., Barnet, C., Wolf, W., Goldberg, M., King, T., Nalli, N., et al. (2012). *The NOAA unique CrIS/ATMS processing system (NUCAPS): First light retrieval results*. In Proceedings of the ITWG meeting. Toulouse: ITWG.

Goldberg, M. D., Kilcoyne, H., Cikanek, H., & Mehta, A. (2013). Joint polar satellite system: The United States next generation civilian polar-orbiting environmental satellite system. *Journal of Geophysical Research: Atmosphere*, *118*(24), 13–463. https://doi.org/10.1002/2013jd020389

Hannon, S. E., Strow, L. L., & McMillan, W. W. (1996). *Atmospheric infrared fast transmittance models: A comparison of two approaches* In Optical spectroscopic techniques and instrumentation for atmospheric and Space research II (Vol. 2830, pp. 94–105). International Society for Optics and Photonics.

Hayden, C. M. (1988). GOES-VAS simultaneous temperature-moisture retrieval algorithm. *Journal of Applied Meteorology*, *27*(6), 705–733. https://doi.org/10.1175/1520-0450(1988)027<0705:gvstmr>2.0.co;2

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification*. In 2015 IEEE International Conference on Computer Vision (ICCV) (pp. 1026–1034). Santiago, CL: IEEE. https://doi.org/10.1109/ICCV.2015.123

Heidinger, A., & Straka, W., III. (2013). Algorithm theoretical basis document: ABI cloud mask. NOAA/NESDIS. *Center for Satellite Applications and Research Tech*. Rep.

Hinton, G. E. (2006). Reducing the dimensionality of data with neural networks. *Science*, *313*(5786), 504–507. https://doi.org/10.1126/science.1127647

Hoffmann, L., Günther, G., Li, D., Stein, O., Wu, X., Griessbach, S., et al. (2019). From ERA-Interim to ERA5: The considerable impact of ECMWF's next-generation reanalysis on Lagrangian transport simulations. *Atmospheric Chemistry and Physics*, *19*(5), 3097–3124. https://doi.org/10.5194/acp-19-3097-2019

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization*. Retrieved from http://arxiv.org/abs/1412.6980

Klein, S. A., & Hartmann, D. L. (1993). The seasonal cycle of low stratiform clouds. *Journal of Climate*, *6*(8), 1587–1606. https://doi.org/10.1175/1520-0442(1993)006<1587:tscols>2.0.co;2

Li, J., Li, Z., & Schmit, T. J. (2020). *ABI legacy atmospheric profiles and derived products from the GOES-R Series*. In The GOES-R Series (pp. 63–77). Elsevier. https://doi.org/10.1016/b978-0-12-814327-8.00007-x

Liu, J. N., Hu, Y., You, J. J., & Chan, P. W. (2014). Deep neural network based feature representation for weather forecasting. In Proceedings on the International Conference on Artificial Intelligence (ICAI) (p. 1). The steering committee of the world congress in Computer Science. Computer Engineering and Applied Computing (WorldComp).

Liu, Q., Wolf, W., Reale, T., & Sharma, A., & NOAA JPSS Program Office. (2014). *NUCAPS: NOAA unique combined atmospheric processing system environmental data record (EDR) products*.

Liu, Y., Chen, X., Peng, H., & Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, *36*, 191–207. https://doi.org/10.1016/j.inffus.2016.12.001

Menzel, W. P., Schmit, T. J., Zhang, P., & Li, J. (2018). Satellite-based atmospheric infrared sounder development and applications. *Bulletin of the American Meteorological Society*, *99*(3), 583–603. https://doi.org/10.1175/bams-d-16-0293.1

Michael, B., & Kristin, C. (2017). *GOES-R and JPSS proving ground demonstration at the 2017 summer experiment—experimental warning program (EWP)*. NOAA Hazardous Weather Testbed (HWT).

Nagle, F. W., & Holz, R. E. (2009). Computationally efficient methods of collocating satellite, aircraft, and ground observations. *Journal of Atmospheric and Oceanic Technology*, *26*(8), 1585–1595. https://doi.org/10.1175/2008jtecha1189.1

Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve restricted boltzmann machines*. In Proceedings of the 27th International Conference on Machine Learning (pp. 807–814). ICML-10.

Nalli, N. R., Barnet, C. D., Reale, A., Tobin, D., Gambacorta, A., Maddy, E. S., et al. (2013). Validation of satellite sounder environmental data records: Application to the cross-track infrared microwave sounder suite. *Journal of Geophysical Research: Atmosphere*, *118*(24), 13–628. https://doi.org/10.1002/2013jd020436

Nalli, N. R., Gambacorta, A., Liu, Q., Barnet, C. D., Tan, C., Iturbide-Sanchez, F., et al. (2018). Validation of atmospheric profile retrievals from the SNPP NOAA-unique combined atmospheric processing system. Part 1: Temperature and moisture. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(1), 180–190. https://doi.org/10.1109/TGRS.2017.2744558

Pedamonti, D. (2018). *Comparison of non-linear activation functions for deep neural networks on MNIST classification task*. Retrieved from http://arxiv.org/abs/1804.02763

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Reale, T., Sun, B., Tilley, F. H., & Pettey, M. (2012). The NOAA Products Validation System (NPROVS). *Journal of Atmospheric and Oceanic Technology*, *29*(5), 629–645. https://doi.org/10.1175/JTECH-D-11-00072.1

Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., & Lebair, W. J. (2017). A closer look at the ABI on the GOES-R series. *Bulletin of the American Meteorological Society*, *98*(4), 681–698. https://doi.org/10.1175/BAMS-D-15-00230.1

Schmit, T. J., Gunshor, M. M., Menzel, W. P., Gurka, J. J., Li, J., & Bachmeier, A. S. (2005). Introducing the next-generation advanced baseline imager on GOES-R. *Bulletin of the American Meteorological Society*, *86*(8), 1079–1096. https://doi.org/10.1175/bams-86-8-1079

Schmit, T. J., Li, J., Lee, S. J., Li, Z., Dworak, R., Lee, Y. K., et al. (2019). Legacy atmospheric profiles and derived products from GOES-16: Validation and applications. *Earth and Space Science*, *6*(9), 1730–1748. https://doi.org/10.1029/2019ea000729

Schmit, T. J., Li, J., Li, J., Feltz, W. F., Gurka, J. J., Goldberg, M. D., & Schrab, K. J. (2008). The GOES-R advanced baseline imager and the continuation of current sounder products. *Journal of Applied Meteorology and Climatology*, *47*(10), 2696–2711. https://doi.org/10.1175/2008jamc1858.1

Seltzer, M. L., Yu, D., & Wang, Y. (2013). *An investigation of deep neural networks for noise robust speech recognition*. In 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 7398–7402). IEEE.

Service (C3S). (2017). *ERA5: Fifth generation of ECMWF atmospheric reanalyzes of the global climate*.

Smith, A., Atkinson, N., Bell, W., & Doherty, A. (2015). An initial assessment of observations from the Suomi-NPP satellite: Data from the cross-track infrared sounder (CrIS). *Atmospheric Science Letters*, *16*(3), 260–266. https://doi.org/10.1002/asl2.551

Smith Sr, W., Weisz, E., & Revercomb, H. (2015). The retrieval of atmospheric profiles from satellite radiances for NWP data assimilation. *Proceding. 20th International TOVS study Conference* (Vol. 4). Lake Geneva, WI: International TOVS Working Group.

Smith, W. L., Weisz, E., Kireev, S. V., Zhou, D. K., Li, Z., & Borbas, E. E. (2012). Dual-regression retrieval algorithm for real-time processing of satellite ultraspectral radiances. *Journal of Applied Meteorology and Climatology*, *51*(8), 1455–1476. https://doi.org/10.1175/jamc-d-11-0173.1

Smith, W. L., Zhang, Q., Shao, M., & Weisz, E. (2020). Improved severe weather forecasts using LEO and GEO satellite soundings. *Journal of Atmospheric and Oceanic Technology*, *37*(7), 1203–1218. https://doi.org/10.1175/jtech-d-19-0158.1

Space Studies Board, Division on Engineering and Physical Sciences, National Academies of Sciences Engineering, & Medicine. (2019). *Thriving on our changing planet: A decadal strategy for Earth observation from space: An overview for decision makers and the public*. Washington, DC: National Academies Press. https://doi.org/10.17226/25437

Strow, L. L., Hannon, S. E., De Souza-Machado, S., Motteler, H. E., & Tobin, D. (2003). An overview of the AIRS radiative transfer model. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 303–313. https://doi.org/10.1109/tgrs.2002.808244

Sun, B., Reale, A., Pettey, M., Smith, R., Nalli, N., & Zhou, L. (2018). *Leveraging the strengths of dedicated, Gruan and conventional radiosondes for satellite hyperspectral geophysical sounding assessment*. In IGARSS 2018-2018 IEEE International geoscience and remote Sensing symposium (pp. 7743–7745). IEEE.

Sun, B., Reale, A., Seidel, D. J., & Hunt, D. C. (2010). Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics. *Journal of Geophysical Research*, *115*(D23), D23104. https://doi.org/10.1029/2010JD014457

Sun, B., Reale, A., Tilley, F. H., Pettey, M. E., Nalli, N. R., & Barnet, C. D. (2017). Assessment of NUCAPS S-NPP CrIS/ATMS sounding products using reference and conventional radiosonde observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(6), 2499–2509. https://doi.org/10.1109/jstars.2017.2670504

Tao, Y., Gao, X., Hsu, K., Sorooshian, S., & Ihler, A. (2016). A deep neural network modeling framework to reduce bias in satellite precipitation products. *Journal of Hydrometeorology*, *17*(3), 931–945. https://doi.org/10.1175/JHM-D-15-0075.1

Tao, Y., Hsu, K., Ihler, A., Gao, X., & Sorooshian, S. (2018). A two-stage deep neural network framework for precipitation estimation from bispectral satellite information. *Journal of Hydrometeorology*, *19*(2), 393–408. https://doi.org/10.1175/JHM-D-17-0077.1

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over north America. *Hydrology and Earth System Sciences*, *24*(5), 2527–2544. https://doi.org/10.5194/hess-24-2527-2020

Weisz, E., Smith, W. L., & Smith, N. (2013). Advances in simultaneous atmospheric profile and cloud parameter regression based retrieval from high-spectral resolution radiance measurements. *Journal of Geophysical Research: Atmospheres*, *118*(12), 6433–6443. https://doi.org/10.1002/jgrd.50521

Zhou, L., Divakarla, M., & Liu, X. (2016). An overview of the joint polar satellite system (JPSS) science data product calibration and validation. *Remote Sensing*, *8*(2), 139. https://doi.org/10.3390/rs8020139

Zhou, Y., & Grassotti, C. (2020). Development of a machine learning-based radiometric bias correction for NOAA's microwave integrated retrieval system (MiRS). *Remote Sensing*, *12*(19), 3160. https://doi.org/10.3390/rs12193160