Improvement of MiRS Sea Surface Temperature Retrievals Using a Machine Learning Approach

Shuyan Liu, Christopher Grassotti, Quanhua Liu, Yan Zhou, and Yong-Keun Lee

***Abstract*—We report on the development of a machine learning approach to improving sea surface temperature (SST) retrievals based on satellite-based microwave channel measurements at frequencies higher than 23 GHz. The approach uses a deep neural network (DNN) trained using Microwave Integrated Retrieval System physical retrievals as inputs and collocated European Centre for Medium-Range Weather Forecasts analyses for training and validation. The DNN was designed to characterize SST retrieval residual and then used to correct the original retrieval. Evaluation based on one year of independent data showed reduction in retrieval residual standard deviation from 3.22 to 1.80 K in January and 3.02 to 1.92 K in July and reduction in mean residual from 0.30 to 0.08 K in January and 0.61 to 0.22 K in July. Comparisons with multilinear regression and machine learning approaches that used measured brightness temperatures as inputs were significantly less effective in retrieving SST directly, although the DNN used brightness temperature also showed improvements. This indicates that physical retrieval provides valuable information useful in characterizing retrieval residual beyond that of the measured radiances. The DNN approach also effectively removed scan angle dependence of retrieval residuals—an important consideration with cross-track instruments. Sensitivity tests indicated that skill declines with time as time increases from training month, but that skill in the same month, one year later is nearly the same as that of the original training month. This suggests that it may be sufficient to pretrain a stratified model with monthly or seasonal dependence using one full annual cycle, which could then be used in subsequent years with continued good performance.**

Manuscript received December 10, 2021; revised February 2, 2022; accepted February 8, 2022. Date of publication February 14, 2022; date of current version February 24, 2022. This work was supported in part by the National Oceanic and Atmospheric Administration under Grant NA19NES4320002 to the Cooperative Institute for Satellite and Earth System Studies at the University of Maryland/Earth System Science Interdisciplinary Center, and Grant NA19OAR4320073 to the Cooperative Institute for Research in the Atmosphere at Colorado State University, and in part by the Joint Polar Satellite System. *(Corresponding author: Shuyan Liu.)*

Shuyan Liu is with the Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80523 USA, and also with the National Oceanic and Atmospheric Administration, National Environmental Satellite Data Information Service, Center for Satellite Applications and Research, College Park, MD 20740 USA (e-mail: shu-yan.liu@noaa.gov).

Christopher Grassotti and Yong-Keun Lee are with the Cooperative Institute for Satellite Earth System Studies, Earth System Science Interdisciplinary Center, College Park, MD 20740 USA, and also with the National Oceanic and Atmospheric Administration, National Environmental Satellite Data Information Service, Center for Satellite Applications and Research, College Park, MD 20740 USA (e-mail: christopher.grassotti@noaa.gov; yong-keun.lee@noaa.gov).

Quanhua Liu is with the National Oceanic and Atmospheric Administration, National Environmental Satellite Data Information Service, Center for Satellite Applications and Research, College Park, MD 20740 USA (e-mail: quanhua.liu@noaa.gov).

Yan Zhou is with the Cooperative Institute for Satellite Earth System Studies, Earth System Science Interdisciplinary Center, College Park, MD 20740 USA (e-mail: yanzhou@umd.edu).

Digital Object Identifier 10.1109/JSTARS.2022.3151002

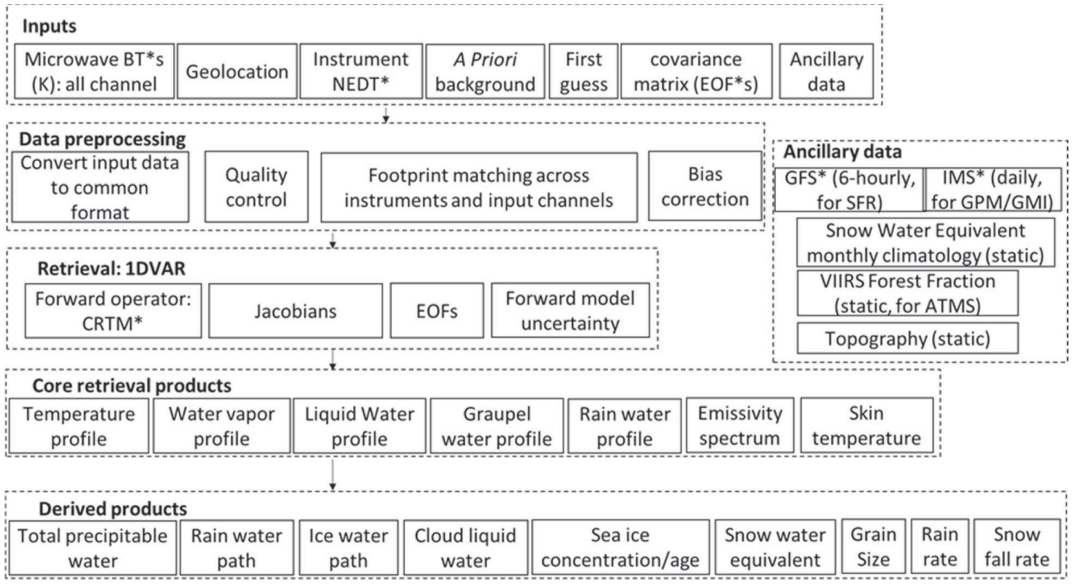
***Index Terms*—Machine learning, microwave integrated retrieval system (MiRS) retrieval, neural network, sea surface temperature (SST).**

I. Introduction

DUE to the longer wavelengths at which they operate, microwave (MW) sensors have inherent advantages in observing clouds and precipitation dynamics, monitoring land and sea surface properties, and estimating profiles of atmospheric temperature and water vapor under nearly all weather conditions. The Microwave Integrated Retrieval System (MiRS) [1] retrieves multiple geophysical parameters from MW observations and has been an official operational system at the National Oceanic and Atmospheric Administration (NOAA) since 2007. The MiRS uses a one-dimensional variational (1DVAR) algorithm to iteratively find the optimal solution based on minimizing a cost function comprised of two terms: departure of the retrieved state vectors from an *a priori* climatology background and departure of forward model simulated radiances from the satellite observed radiances [1], [2]. Empirical orthogonal functions (EOFs) are used to further reduce the degrees of freedom in the solution and stabilize the retrieval [3], [4]. The climatological background used in MiRS varies with latitude, longitude, season, and time of day.

MiRS uses the Community Radiative Transfer Model (CRTM) as the forward operator [5], [6]. Since CRTM computes radiances and the corresponding Jacobians under clear, cloudy, and precipitating sky conditions, MiRS performs retrievals in all these conditions. In addition to all sky conditions, MiRS also operates over all surface types, namely, ocean, land, ice, and snow. The parameters directly retrieved in MiRS include: temperature and water vapor vertical profiles, cloud and precipitation parameter vertical profiles, skin temperature, and emissivity spectrum. These parameters are retrieved simultaneously, and self-consistency among different parameters is maintained via both the background *a priori* constraints and the use of EOFs. The measurement constraint also ensures that the retrieved solution is consistent with the observed radiances. Once the core retrieval is completed, postprocessing algorithms are used to derive additional products, for example, total precipitable water (TPW) and cloud liquid water are vertically integrated from water vapor and cloud liquid water content profiles; snow water equivalent, surface precipitation rate, etc., are determined from

[1] [Online]. Available: https://www.star.nesdis.noaa.gov/mirs



*BT: brightness temperature
*NEDT: noise equivalent delta temperature
*EOF: Empirical Orthogonal Function, the basis functions used to represent the geophysical profile in the retrieval
*GFS: Global Forecast System
*IMS: Interactive Multisensor Snow and Ice Mapping System
*CRTM: Community Radiative Transfer Model

Fig. 1. Schematic of MiRS processing components and data flow showing MiRS core retrieval (produced by the 1DVAR) and derived products retrieval (produced by a postprocessing step). Core products are retrieved simultaneously as part of the state vector. Postprocessing products are derived through vertical integration or through predefined relationships (i.e., look-up tables or analytic functions) between state vector variables and the derived products.

other algorithms that use the core retrievals [7], [8] as input. A schematic of the MiRS processing components and data flow is shown in Fig. 1.

The MiRS algorithm currently runs operationally on passive MW data from both conically scanning (i.e., DMSP F-17 and F-18 SSMIS, and GPM GMI), and cross-track scanning (S-NPP and NOAA-20 ATMS, NOAA-18, -19, MetopB, MetopC AMSUA-MHS) instruments and the retrieval products are used in operational weather analyses and forecasts routinely. Additionally, the products are used as inputs to other operational systems for various applications. For example, the algorithm to generate multisatellite blended layer precipitable water and blended TPW products ingests MiRS water vapor profiles and TPW [9]. The MiRS retrieved temperature and water vapor profiles are used as inputs to the Hurricane Intensity and Structure Algorithm, a tropical cyclone intensity estimation algorithm that was used operationally at the National Hurricane Center and developed at the Colorado State University/Cooperative Institute for Research in the Atmosphere [10]. The NOAA Climate Prediction Center Morphing Technique Algorithm [11], [12] uses MiRS precipitation rates are used as one of several satellite-based precipitation inputs.

In this study, we focus on the core retrieval variable skin temperature and investigate a means of improving its retrieval over ocean surfaces that, for the purposes of this discussion, we refer to by the commonly used term, sea surface temperature (SST). Minnett *et al.* [13] provide an excellent review of satellite-based estimates of SST over more than 50 years. Both infrared (IR) and passive MW measurements have been used, with IR-based estimates generally found to be more accurate than those from MW instruments. Validation studies typically have found root-mean-square errors with respect to reference data of IR and MW retrievals to be 0.3–0.5 and 0.6–0.7 K, respectively. The advantage of IR measurements lies in the much higher spatial resolution (∼1–5 km) relative to MW observations (∼25–50 km), but, as is the case with IR retrievals of many geophysical properties, reliable estimates are only possible in cloud-free scenes. MW SST retrievals, on the other hand, are possible in nearly all-weather conditions, with the exception of scenes containing precipitation.

To our knowledge, nearly all applications of satellite MW measurements to SST estimation have used conically scanning instruments (see, for example, [14] and [15]), which have inherent advantages relative to cross-track sensors. The advantages of conically scanning radiometers are primarily the presence of orthogonally polarized (vertical and horizontal) channels at most frequencies with all measurements made at a fixed angle (and spatial resolution), and the existence of lower frequency channels at 10 GHz and lower, which allows for higher sensitivity of radiance measurements to changes in SST [16], [17]. Conversely, cross-track scanning instruments in operation measure radiances with mixed vertical and horizontal polarization that changes with scan angle (as does the corresponding surface emissivity) and normally operate at frequencies starting at 23 GHz and higher where atmospheric absorption is higher. Additionally, the spatial resolution of cross-track measurements varies considerably from near nadir to the edge of scan. All these factors are challenges to estimating the radiometric signal of SST. Nevertheless, given the significant number of currently operational polar-orbiting satellites that carry cross-track

radiometers (e.g., S-NPP, the NOAA/NASA Joint Polar Satellite System (JPSS) series satellites, the EUMETSAT MetOp series) the potential increase in temporal and spatial coverage of reliable SST estimates in nearly all-weather conditions would be a benefit to the operational and research community. Cross-track instruments also feature wider swath widths than conically scanning radiometers, which offer the potential of more complete spatial coverage over time. In addition, since the MiRS algorithm utilizes all channel measurements (i.e., both surface sensitive and sounding channels) to simultaneously obtain a retrieval solution, and includes a physical accounting of atmospheric absorption, emission, and scattering via the CRTM forward model, there may be additional information related to SST yet to be extracted within the retrieval system.

Machine learning methods interpret data by building models based on large amount of input datasets without advanced knowledge and make predictions. Deep neural networks (DNNs) are one of the most widely used machine learning methods and are often used in supervised learning problems. In supervised learning, a training dataset is given in which each set of input variables (or predictors) is corresponding to an already known output. The purpose of neural network is to find relationship between the predictors and the outputs in the training dataset. When a new dataset is provided (testing dataset), predictions are made by applying the learned relationship on predictors from the testing dataset. Neural networks have been widely used in the retrieval of geophysical parameters based on remote sensing data and in other atmospheric science fields in recent years [18]–[23], and have been shown to be remarkably effective in learning predictive relationships between variables. Recently, Zhou and Grassotti [24] utilized a DNN to predict the radiometric biases in ATMS observations; when these bias corrections were applied within the MiRS system, significant improvements were seen in sounding products relative to the baseline operational system that used a static bias correction.

The purpose of this work was twofold: first, demonstrate a simple proof of concept for the potential of using a machine learning approach that can leverage the existing information content present in the MiRS geophysical retrievals to improve the accuracy of the SST based on 23 GHz and higher, which is very challenging for SST retrieval. Second, use MW data from a cross-track radiometer that has a wider swath and a better global coverage, rather than a conical scanning instrument, to determine the retrieval performance for this instrument design. This study explores the feasibility of using higher frequency channel measurements and cross-track instruments for SST estimation.

Additionally, perhaps more significantly in the context of the MiRS algorithm itself, given the importance of accurate specification of surface conditions to the entire 1DVAR retrieval, a refinement of the SST retrieval within MiRS, could be used to further improve the retrieval process within the algorithm by providing a stronger SST constraint on the variational inversion.

The rest of this article is structured as follows. Section II describes the datasets used and the DNN methodology employed for the SST improvement. Section III discusses the results from several retrieval experiments that were conducted. Finally, Section IV summarizes the work and points to possible future efforts that build on the results shown here.

TABLE I
ATMS INSTRUMENT CHARACTERISTICS

Chan No.	Center Freq. (GHz)	Band width (GHz)	Sensitivity (NEDT) (K)	Accuracy (K)	Beam width (degrees)
1	23.8	0.27	0.7	1.0	5.2
2	31.4	0.18	0.8	1.0	5.2
3	50.3	0.18	0.9	0.75	2.2
4	51.76	0.4	0.7	0.75	2.2
5	52.8	0.4	0.7	0.75	2.2
6	53.596±0.115	0.17	0.7	0.75	2.2
7	54.4	0.4	0.7	0.75	2.2
8	54.94	0.4	0.7	0.75	2.2
9	55.5	0.33	0.7	0.75	2.2
10	57.2903	0.33	0.75	0.75	2.2
11	57.2903±0.115	0.078	1.2	0.75	2.2
12	57.2903	0.036	1.2	0.75	2.2
13	57.2903±0.322	0.016	1.5	0.75	2.2
14	57.2903±0.322 ±0.010	0.008	2.4	0.75	2.2
15	57.2903±0.322 ±0.004	0.003	3.6	0.75	2.2
16	87-91(88.20)	2.0	0.5	1.0	2.2
17	164-167	3.0	0.6	1.0	1.1
18	183.31±7	2.0	0.8	1.0	1.1
19	183.31±4.5	2.0	0.8	1.0	1.1
20	183.31±3	1.0	0.8	1.0	1.1
21	183.31±1.8	1.0	0.8	1.0	1.1
22	183.31±1.0	0.5	0.9	2.0	1.1

II. DATA AND METHODS

A. NOAA-20 ATMS Data and European Centre for Medium-Range Weather Forecasts (ECMWF) Analyses

NOAA-20 is the second of NOAA's latest generation of the U.S. polar-orbiting satellites of the JPSS. The satellite was launched on November 18, 2017. The ATMS onboard NOAA-20 has 22 channels operating in cross-track scanning mode with frequencies ranging from 23 to 183 GHz, which allows ATMS to observe surface conditions under both clear and cloudy sky conditions. Table I provides ATMS channel specification information, including central frequency, polarization, noise equivalent differential temperature, and resolutions at nadir and swath edge. MiRS uses all of the 22 channels measured radiances as input to retrieve atmospheric and surface products simultaneously (see Fig. 1).

The ECMWF operational global analyses [2] are used as a reference dataset for training and validation. Analyses were available every 6 h at a resolution of 0.25° in latitude and longitude. For training and validation, ECMWF analyses were temporally and spatially interpolated to ATMS measurement locations.

B. SST Retrieval Machine Learning Methods

Three machine learning methods for SST retrievals are developed and compared in this study. The first method is DNN based on MiRS retrievals (referred to as DNN-Retrieval hereafter). In this case, the training target is the original SST retrieval residual,

[2][Online]. Available: https://www/ecmwf.int

defined as the original MiRS SST retrieval minus the collocated ECMWF SST. The second method is a DNN based on ATMS observed brightness temperatures (referred to as DNN-TB hereafter), and the third method, also based on ATMS brightness temperatures, is a multivariate linear regression (referred to as MLReg-TB hereafter). Both DNN-TB and MLReg-TB use SST itself as a training target, again coming from the ECMWF analysis.

The DNN model is trained by using two neural network layers with 200 nodes in each layer. The activation function is the rectified linear unit and the optimizer is root-mean-square propagation (RMSprop) with mean squared error as the convergence metric. The learning rate for RMSprop is set at 0.001. The number of training epochs is another important factor for neural networks. Sufficient number of epochs can prevent underfitting, but can also lead to overfitting. This study chose a sufficiently large value as the maximum number of epochs, using the widely used early stopping method to terminate training before overfitting occurred and improve the generalization of the trained model. Possible overfitting is monitored by the validation data, which is a subset (50% in this study) of the input data. If, after the specified number of epochs (patience), there is no improvement using the validation dataset, the training stops. Through trial and error, the selected values of maximum epochs and patience are 1000 and 100, respectively.

The input variables of features of the three different methods varied with model. For the DNN-Retrieval model, the following 28 variables were used: cosine of satellite viewing angle, latitude, longitude, MiRS retrieved SST, MiRS retrieved emissivity for 22 channels, MiRS retrieved CLW, and MiRS retrieved TPW. The input layers for the DNN-TB and MLRegr-TB models are the same, and both include the following 25 variables: cosine of satellite viewing angle, latitude, longitude, and the observed ATMS brightness temperatures in channels 1–22. Only data over ocean surfaces were used. All the input variables were normalized by their respective mean and standard deviation calculated from the training dataset. Normalizing the data before training the model effectively improved the learning speed of the neural network to determine the optimal weights and biases. Fig. 2 shows the structure of the neural network used in this study; the input and output layers are illustrated for the case of the DNN-Retrieval model.

The training dataset is constructed based on 12 days in 2020, one day in each month, typically the first day. During the model training process, 50% of the data are randomly selected to build the model, and the remaining 50% of the data are used for validation. The total input dataset sample size is $\sim 21 \times 10^6$. The same model is applied to different days in 2020–2021 for independent predictions; the predictions dates are the 16th day of each month. The three tests of DNN-Retrieval, DNN-TB, and MLReg-TB follow this same data handling method. The prediction data has the sample size of $\sim 1.0 \times 10^6$.

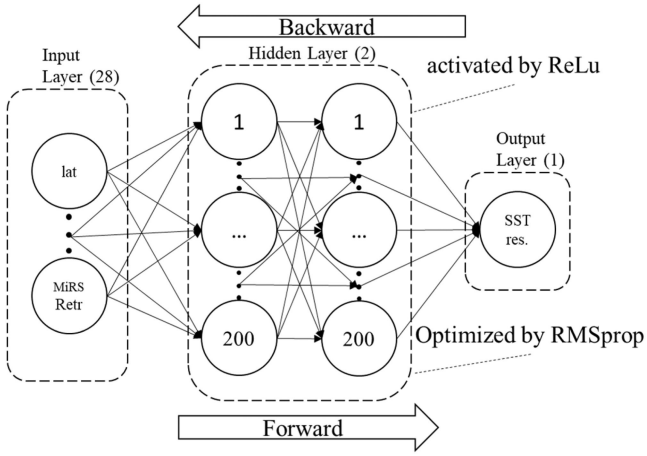


Fig. 2. DNN architecture used in this study. In this example, we use the model of the DNN-Retrieval experiment to illustrate the input and output layers. See discussion for details of the input and output layers of each experiment.

III. Experimental Results

A. Geospatial Distributions

The geospatial distributions for NOAA-20/ATMS SST predictions from DNN-Retrieval, DNN-TB, and MLReg-TB for 2021-01-16 and 2021-07-16, which represent northern hemisphere winter and summer, respectively, are shown in Figs. 3 and 4. MiRS operational SST retrievals for both days are also included as well as collocated ECMWF analyses. In the following discussion, we refer to the difference of the satellite-based estimate and the ECMWF analysis value as the residual. For both seasons, the MiRS operational SST retrieval shows certain retrieval limitations due to ATMS instrument frequency selection and the cross-track scanning method. For both descending (nighttime) and ascending (daytime), SSTs are biased low (relative to the ECMWF analysis) for high latitudes poleward of 60°N and °S, and show a strong scan dependence with warmer retrieved SSTs at the swath edge. The ascending orbit data also shows coastal artifacts due to the land/sea surface temperature and emissivity contrast, particularly the coastal regions of northwestern, northern, and northeastern Africa. Both DNN-Retrieval and DNN-TB correct the aforementioned problems and an improved global SST distribution with the tropical warm pool in the western Pacific Ocean and the eastern Indian Ocean are evident. The MLReg-TB shows improvement of the cold SST bias for high latitudes relative to MiRS operational, and it also eliminates the coastal artifacts seen for ascending orbit retrievals. But the scan position dependence and discontinuities still exist with swath edge retrievals colder than in the inner swath, which is contrary to the pattern seen in MiRS operational retrievals. Since 2021-01-16 is southern hemisphere summer and 2021-07-16 is northern hemisphere summer, the coastal artifacts seen in MiRS operational retrievals, which are tied to the strong land/ocean surface temperature contrast, are more pronounced near Australia and South Africa for 2021-01-16 and, as mentioned earlier, along the northern Africa coast for 2021-07-16.

B. Density Scatter Distributions and Histograms

Fig. 5 shows SST density scatterplots of ECMWF versus MiRS operational, DNN-Retrieval, DNN-TB, and MLReg-TB for the two study days. The MiRS operational retrievals show a higher degree of scatter when compared with the other three

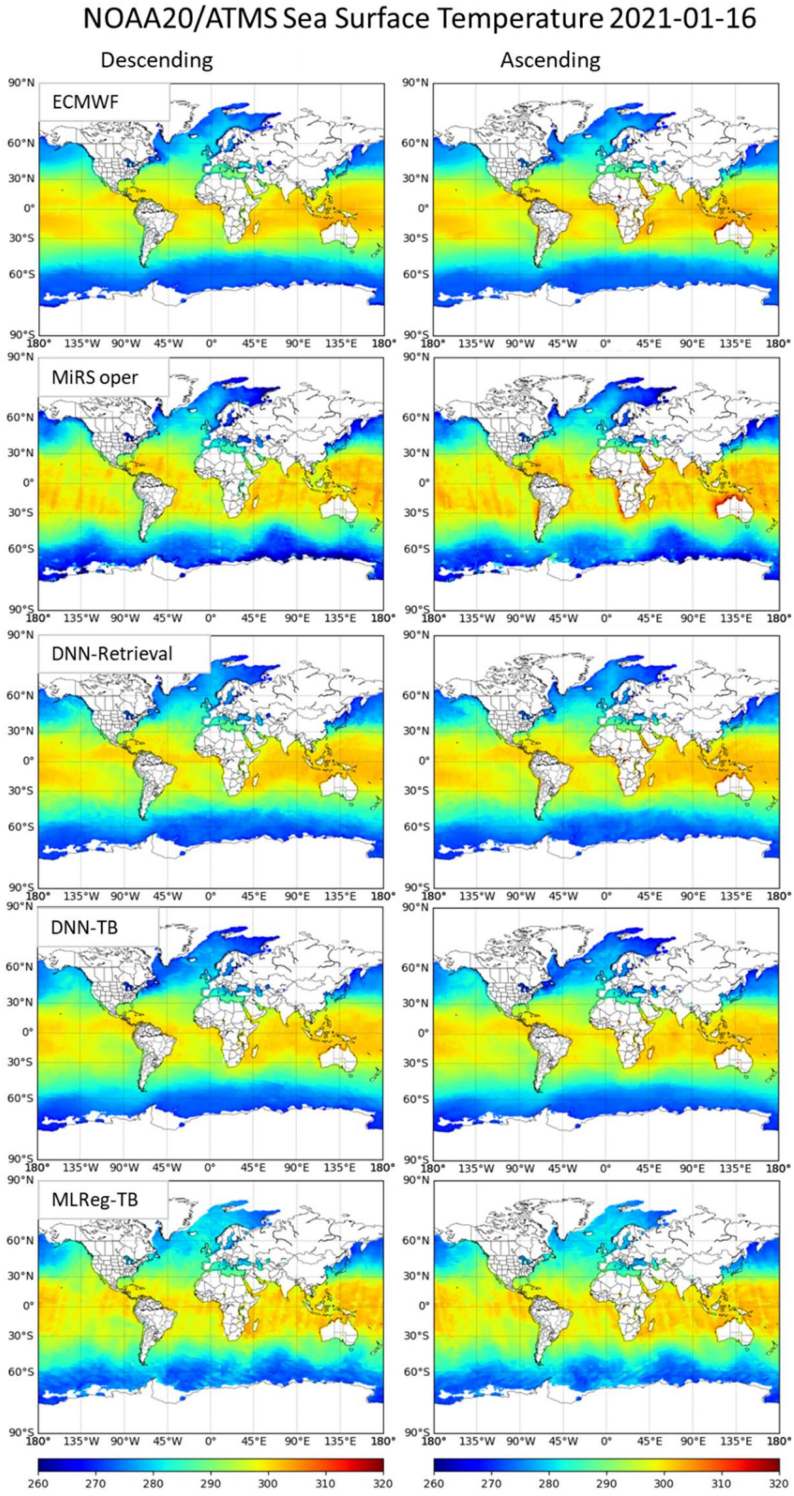


Fig. 3. Global SST maps valid 2021-01-16 corresponding to: ECMWF analyses (top row), MiRS operational retrievals (second row), DNN-Retrieval (third row), DNN-TB (fourth row), MLReg-TB (bottom row) experiments for descending (left column) and ascending (right column), model training using 12 days in 2020. See discussion for description of each experiment.

algorithms, with cold biases for low SST values. Despite the underestimation, the overall mean residuals are 0.3/0.6 K and standard deviations are 3.22/3.02 K for the two days. The DNN-Retrieval algorithm corrected the residuals seen in the operational retrievals and shows the best performance overall, with mean residuals of –0.08/–0.22 K and a residual standard deviation of 1.80/1.92 K for January/July. The DNN-TB results show some improvement over the operational retrievals but have higher errors than the DNN-Retrieval retrievals, with a residual mean and standard deviation of –0.41/–0.42 and 2.15/2.27 K for

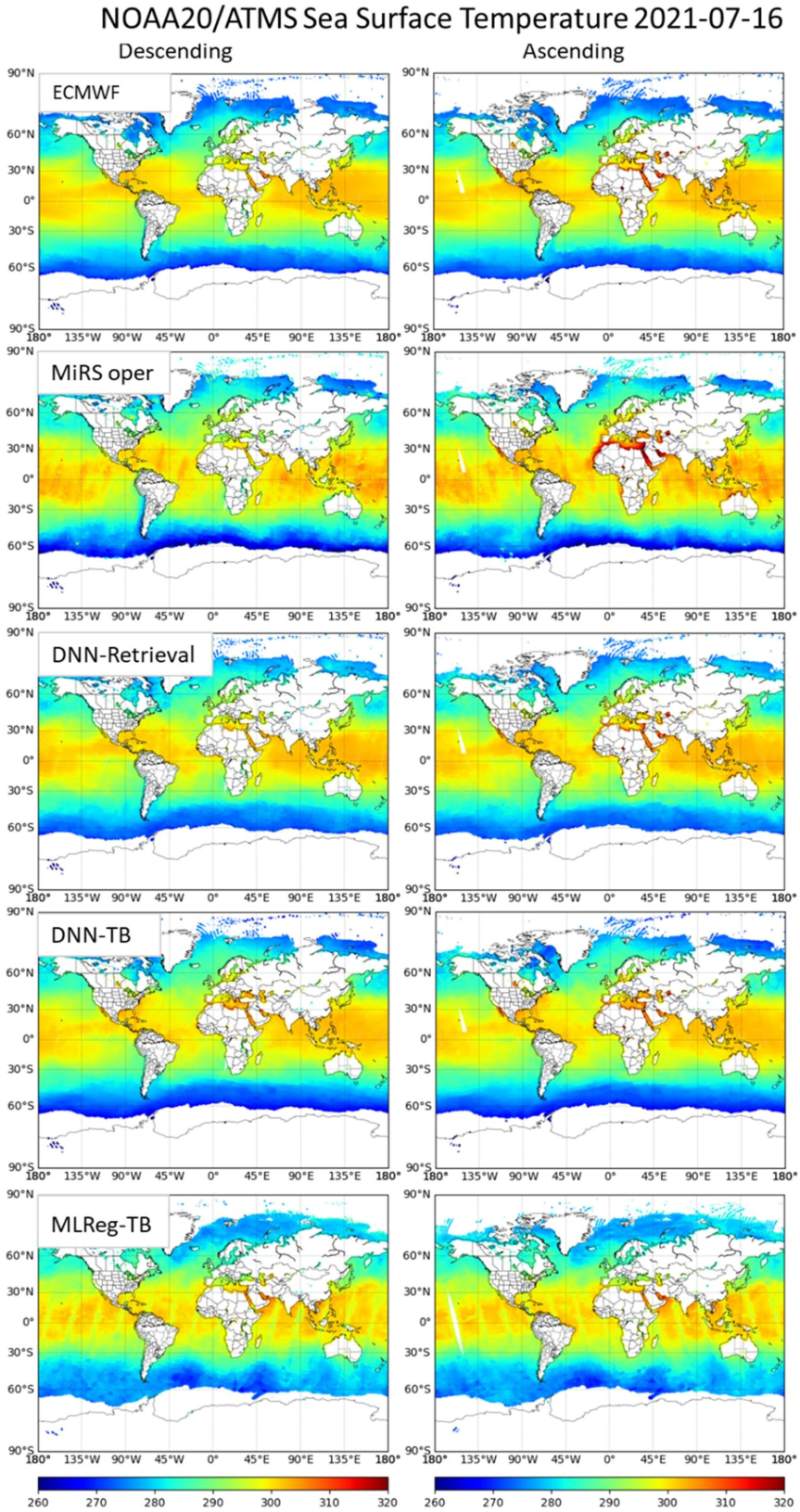


Fig. 4. Global SST maps valid 2021-07-16 corresponding to: ECMWF analyses (top row), MiRS operational retrievals (second row), DNN-Retrieval (third row), DNN-TB (fourth row), MLReg-TB (bottom row) experiments for descending (left column) and ascending (right column), model training using 12 days in 2020.

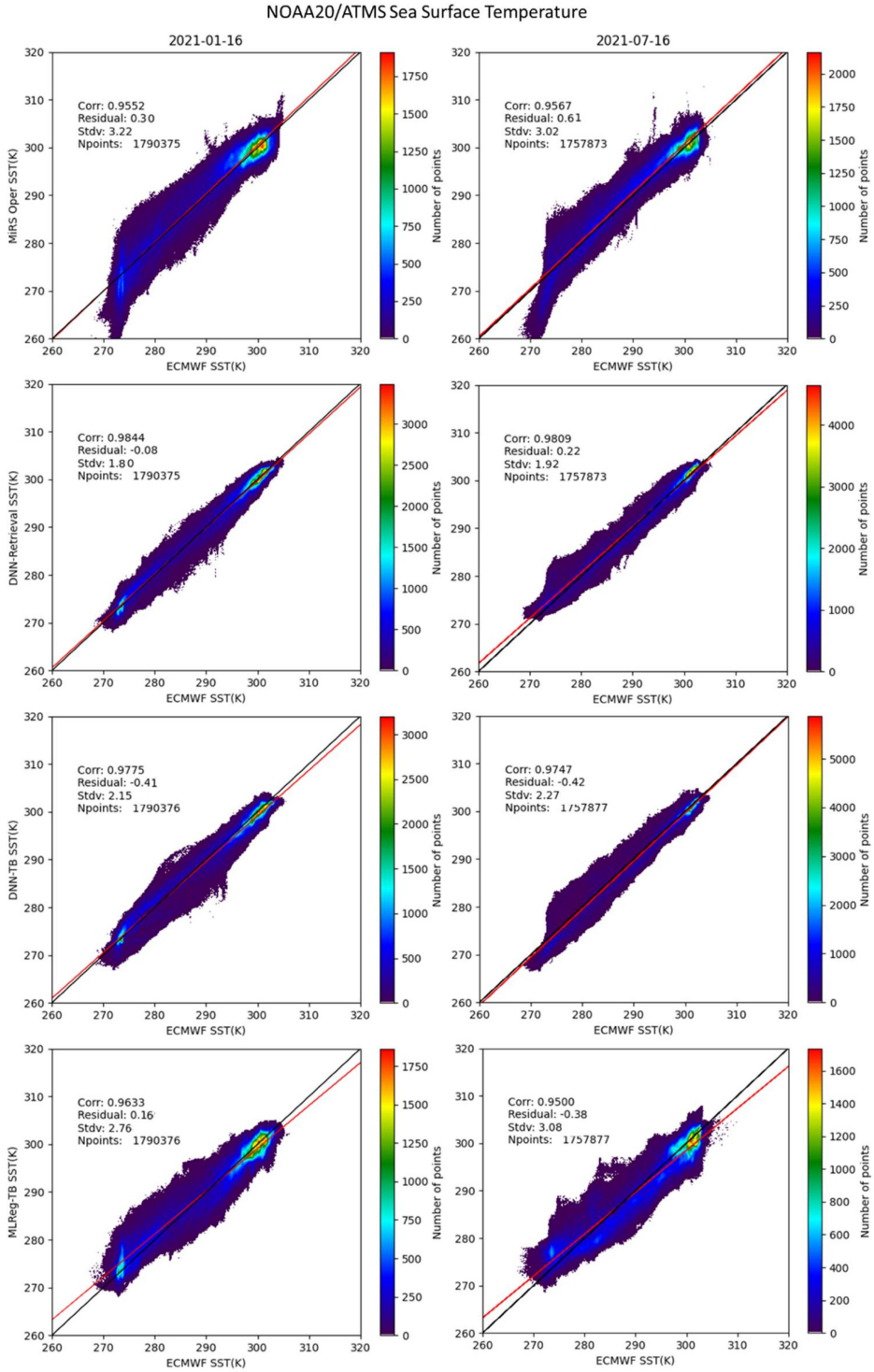


Fig. 5. Density scatterplots comparing ECMWF analysis SST with retrieved SST from: MiRS operational algorithm (top row), DNN-Retrieval (second row), DNN-TB (third row), and MLReg-TB (fourth row), model training using 12 days in 2020. Results from 2021-01-16 (left column) and 2021-07-16 (right column) are shown.

the two days, respectively. The MLReg-TB results also show large scatter with a warm difference of 0.16 K for 2021-01-16 and cold differences of –0.18 K for 2021-07-16, whereas the standard deviations are 2.76 and 3.01 K, respectively.

Fig. 6 contains histograms of SST retrieval residuals for each of the four retrieval experiments on both test days with the bin size of 0.5 K ranging from –10 to 10 K. The histograms show that the distribution of the residuals is considerably narrower and more sharply peaked near zero for both the DNN-Retrieval and DNN-TB experiments, and that the DNN-Retrieval distribution is slightly narrower. DNN-Retrieval peaks at –0.5 K for January and 0 K for July, both days with frequency over 14%. DNN-TB

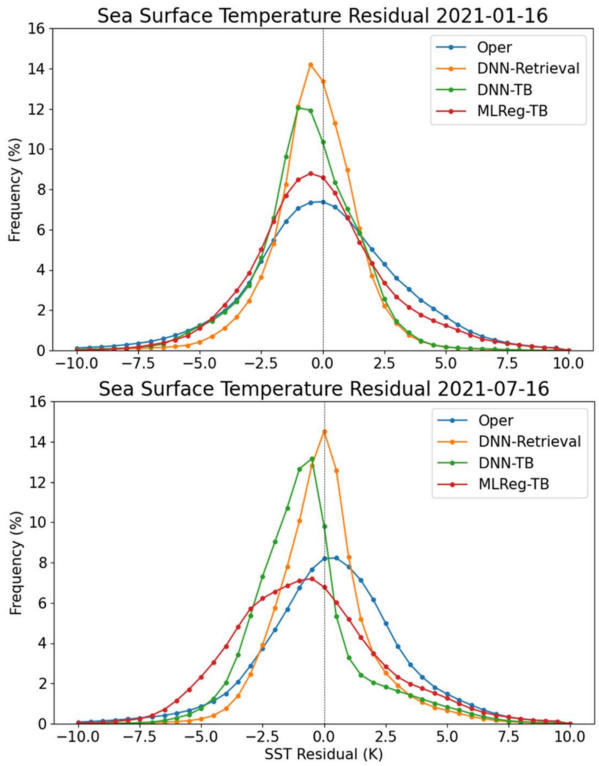


Fig. 6. Frequencies of retrieved SST residual for MiRS operational algorithm, DNN-Retrieval, DNN-TB, and MLReg-TB experiments on 2021-01-16 and 2021-07-16, model training using 12 days in 2020.

peaks at –1.0 K (12.04%) and –0.5 K (13.15%) for January and July. The MLReg-TB curve is much flatter with lower maximum values than both of the DNN experiments. For January 16, frequencies from –1.0 to 0.0 K are all over 8% with the maximum value of 8.79% at –0.5 K. For July 16, the maximum value for MLReg-TB is 7.2% at –0.5 K and the second largest value is 7.1% at –1.0 K, all the other frequencies are less than 7%. MiRS operational has its highest frequency of 7.38% at 0.0 K for January. For July, the highest frequency of MiRS operational is 8.22% at 0.5 K with 8.2% at 0 K.

C. Performance Time Series

Time series of SST residuals and difference standard deviations for each of the prediction days from 2020-01-16 to 2021-07-16 are shown in Fig. 7. MiRS operational retrievals are warmer than ECMWF across the whole year with slightly smaller residuals in April. In contrast, DNN-TB retrievals are consistently colder than ECMWF with the only exception in February. In terms of standard deviation, the DNN-Retrieval approach showed consistently smaller values than the other three algorithms, with the maximum standard deviation value of 2.16 K in February 2021. The DNN-Retrieval approach results in significantly improved SST retrieval performance over the entire 21-month period.

D. Scan Angel Dependence

As noted, the MiRS operational and MLReg-TB retrievals showed strong scan dependence as shown in the geospatial distribution maps contained in Figs. 2 and 3. To quantify this behavior, Fig. 8 shows the mean residuals and residual standard deviation as a function of satellite zenith angle from –65° to 65°. The operational and MLReg-TB retrievals exhibit strong scan position dependence with the largest residual located at larger angles (i.e., the swath edge). In addition, the MLReg-TB retrievals show seasonal dependence with January and July having an opposite mean residual sign at the scan edges. The operational bias variations are similar between the two days, with both showing an increased warmer difference near the scan edges. The DNN-Retrieval SST retrievals show significantly improved performance relative to the operational retrievals, and contain only slight angle dependence, whereas the DNN-TB residuals are nearly constant across all angles, indicating little angle dependence. Mean residuals and residual standard deviations for January and July show that both DNN-Retrieval and DNN-TB have little to no scan dependence, whereas DNN-Retrieval algorithm exhibits the lowest standard deviation for all angles.

E. Impact of Intra-Annual Variability

The preceding analysis has indicated that the DNN-Retrieval exhibits the best SST performance among the four algorithms studied. The training of the model was based on 12 days from the year 2020 in order to capture the annual climatology. In order to assess the impact of targeted training optimized to capture the intra-annual variability, we used the same DNN-Retrieval model design but trained two models using one single day of global data from 2020-01-01 and 2020-07-01. We then applied the two models to independent days from the same corresponding months in 2021 to assess the performance. Fig. 9 compares the SST residual and difference standard deviation over the period January 2020 to September 2021 (one day per month, for a total of 21 days) for three algorithms: MiRS operational, and predictions from the model trained on 2020-01-01 and 2020-07-01, respectively. First, as noted above, the DNN-Retrieval performance is always better than operational. Second, it is seen that the performance of the DNN-Retrieval is optimum (smaller residual mean and standard deviation) in the portion of the year close to the corresponding training date, and degrades slightly as time increases away from the training dates. In summary, better performance is achieved when prediction and training are within the same season. This suggests that an operational implementation that stratifies training monthly or seasonally will have the best overall performance.

This is further confirmed in Fig. 10, which is the density scatterplot of ECMWF SST versus DNN-Retrieval SST for 2021-01-16 and 2021-07-16 using the model trained on data from the same month, but one year earlier, i.e., 2020-01-01 and 2020-07-01.

Comparing with DNN-Retrieval predictions based on the 12-day model (see Fig. 5), both mean residuals and residual standard deviations are improved. The mean residuals are improved from –0.08 to 0.03 K for January, and from 0.22 to –0.15 K for July. Similarly, the residual standard deviation is improved from 1.80 to 1.67 K and 1.92 to 1.46 K for January and July, respectively.

Finally, Fig. 7 also contains results from training a DNN-Retrieval model with stratified training in which a separate DNN

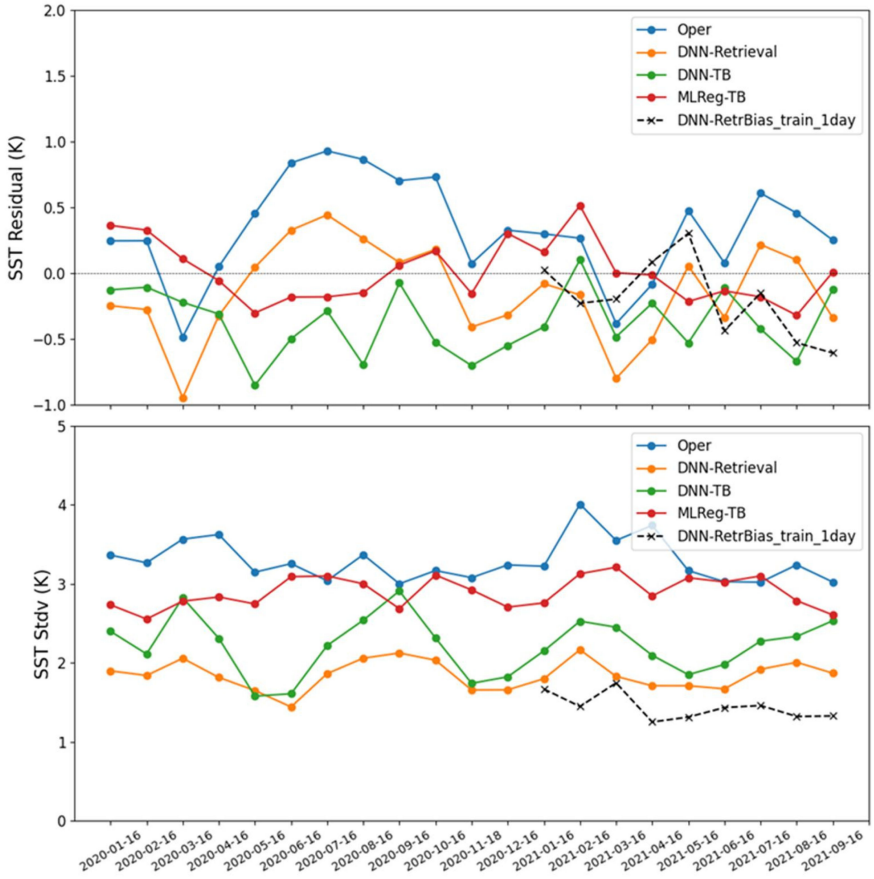


Fig. 7. Time series of SST retrieval residual and difference standard deviation with respect to ECMWF analyses for the 19 prediction days in 2020 to 2021, based on model training using 12 days in 2020. Results are shown for: MiRS operational algorithm, DNN-Retrieval, DNN-TB, and MLReg-TB. See discussion for description of each experiment.

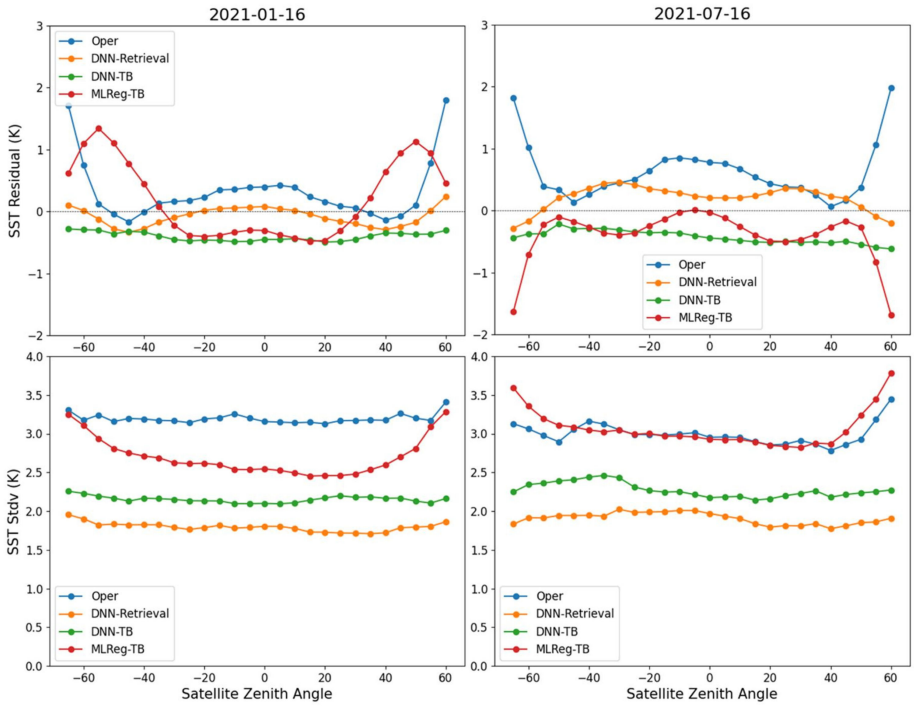


Fig. 8. Satellite zenith angle dependence of SST retrieval bias and difference standard deviation with respect to ECMWF analyses. Results are shown for: MiRS operational algorithm, DNN-Retrieval, DNN-TB, and MLReg-TB, model training using 12 days in 2020.

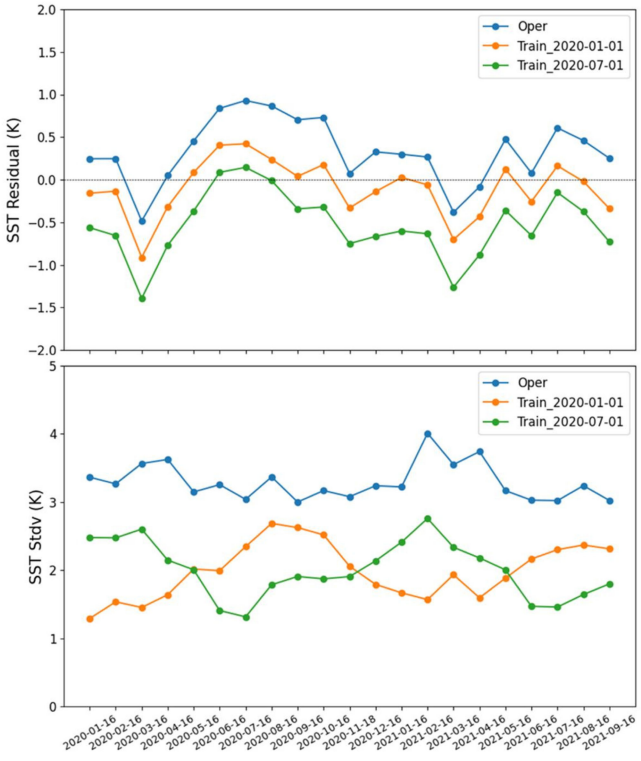


Fig. 9. Time series of SST retrieval performance with respect to ECMWF analyses from January 2020 to September 2021. Results are shown for: MiRS operational algorithm and DNN-Retrieval trained using data from 2020-01-01 and 2020-07-01, respectively.

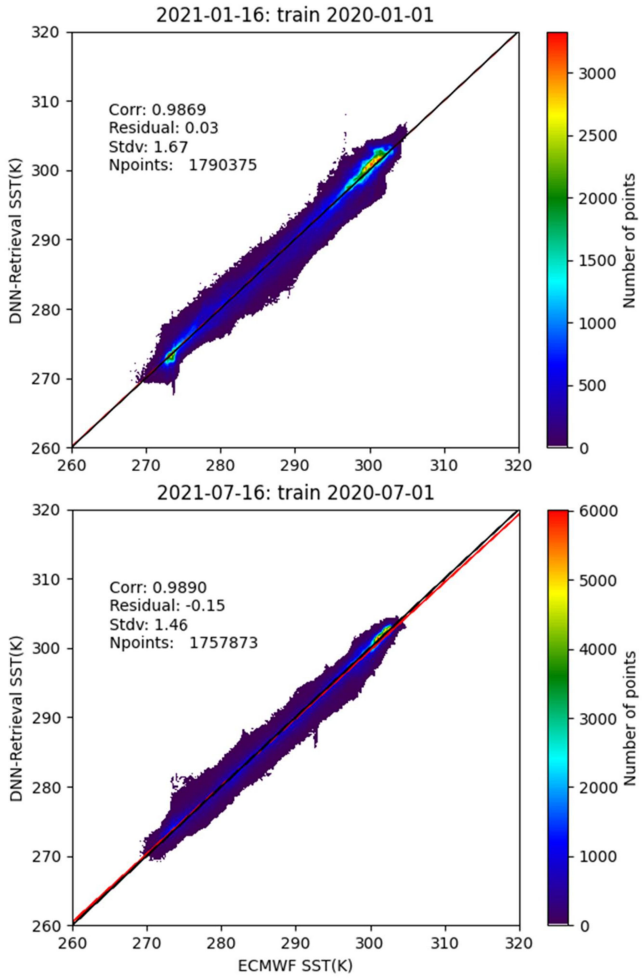


Fig. 10. Density scatterplots of ECMWF analysis SST versus retrieved SST on independent test days of 2021-01-16 and 2021-07-16, in which two DNN-Retrieval models were trained on data from January and July 2020, respectively.

was developed for each month using only that month's data (but from one year earlier) for training. It can be seen that the stratified DNN mean residual and residual standard deviation are consistently lower in each month than the single DNN-Retrieval model trained on the 12 months of data. In practice, a moving window for training with overlapping time periods would likely minimize any temporal discontinuities in retrievals as one moves from one month to the next.

IV. Summary and Conclusion

This article described a method for characterizing and correcting the residual in MiRS retrieved SST. The method used a DNN trained and validated with collocated MiRS NOAA-20 ATMS retrievals and ECMWF operational global analyses. For comparison, two other predictive models were also trained, which used observed ATMS brightness temperatures directly, a DNN and a multilinear regression model. Training data were taken from the year 2020, and independent validation experiments were run using independent 2020 to 2021 data. Regardless of the training and validation time period chosen, the DNN-corrected MiRS retrievals showed the best performance with a significant reduction of residual and standard deviation relative to the ECMWF analysis reference, and a significant reduction of scan angle-dependent artifacts. It appears that inclusion of the geophysical information already extracted from the brightness temperatures by the MiRS 1DVAR retrieval provides valuable information that can be exploited by the DNN to produce a much improved SST estimate. Sensitivity tests indicated that, while training the model using one month's data results in a DNN prediction that loses skill within two to three months from the training month, the same model retains a high level of skill when applied to data exactly one year later. This suggests that for operational purposes, it may be sufficient to train a static model using (approximately) monthly data for one full annual cycle.

The method developed in this study is based on MW frequencies greater than or equal to 23 GHz and has higher mean residual and residual standard deviation than 10 GHz dominated methods, which is to be expected. The results presented here should be seen as a preliminary proof of concept in that testing on additional time periods and more careful validation procedures will be required. Incorporation of alternate reference data for training and/or validation would help to further gauge the sensitivity of the method to training data and to better quantify the accuracy of the retrievals. In particular, the use of *in situ* buoy measurements [25], or gridded SST analyses such as those from the Canadian Meteorological Centre [26], [27], or the NOAA daily Optimum Interpolation Sea Surface Temperature [28], [29] would be highly desirable. Finally, in the context of the MiRS retrieval itself, the approach described here could be used to

enhance the overall MiRS retrievals, beyond simply improving the SST. Since the 1DVAR is a simultaneous physical retrieval, using a more accurate SST (the DNN correction process is computationally rapid) as input to a second 1DVAR retrieval with an improved surface temperature constraint could yield improvements to other components of the geophysical state vector. Planning for this future application is currently underway.

Acknowledgment

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce. The authors would like to thank the editor for their effort and two anonymous reviewers for their constructive comments.

References

[1] Q. Liu and F. Weng, "One-dimensional retrieval algorithm of temperature, water vapor, and cloud water profiles from advanced microwave sounding unit (AMSU)," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 5, pp. 1087–1095, May 2005.

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

[3] S. A. Boukabara *et al.*, "A physical approach for a simultaneous retrieval of sounding, surface, hydrometer, and cryospheric parameters from SNPP/ATMS," *J. Geophys. Res. Atmos.*, vol. 118, pp. 12600–12619, 2013, doi: 10.1002/2013JD020448.

[4] S. A. Boukabara, K. Garrett, and C. Grassotti, "Dynamic inversion of global surface microwave emissivity using a 1DVAR approach," *Remote Sens.*, vol. 10, no. 5, pp. 679–696, 2018, doi: 10.3390/rs10050679.

[5] Y. Han *et al.*, "Community radiative transfer model (CRTM) - Version 1," Nat. Ocean. Atmos. Admin., Camp Springs, MD, USA, NOAA Tech. Rep. 122, 2006, pp. 1–33.

[6] S. Ding *et al.*, "Validation of the community radiative transfer model," *J. Quant. Spectrosc. Radiat. Transf.*, vol. 112, no. 6, pp. 1050–1064, 2011, doi: 10.1016/j.jqsrt.2010.11.009.

[7] C. Grassotti *et al.*, "Precipitation estimation from the microwave integrated retrieval system (MiRS)," in *Satellite Precipitation Measurement* (Advances in Global Change Research Book), vol. 67, V. Levizzani, C. Kidd, D. B. Kirschbaum, C. D. Kummerow, K. Nakamura, and F. J. Turk, Eds. Cham, Switzerland: Springer, 2020, pp. 153–168, doi: 10.1007/978-3-030-24568-9_9.

[8] S. Liu *et al.*, "The NOAA microwave integrated retrieval system (MiRS): Validation of precipitation from multiple polar-orbiting satellites," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3019–3031, Jun. 2020, doi: 10.1109/JSTARS.2020.3000348.

[9] J. M. Forsythe, S. Q. Kidder, K. K. Fuell, A. LeRoy, G. J. Jedlovec, and A. S. Jones, "A multisensor, blended, layered water vapor product for weather analysis and forecasting," *J. Oper. Meteorol.*, vol. 3, no. 5, pp. 41–58, 2015, doi: 10.15191/nwajom.2015.0305.

[10] G. Chirokova, M. DeMaria, R. DeMaria, J. Dostalek, and J. Beven, "Use of JPSS ATMS-MiRS retrievals to improve tropical cyclone intensity forecasting," in *Proc. 20th Conf. Satell. Meteorol. Oceanogr.*, Phoenix, AZ, USA, 2015, p. P157. [Online]. Available: https://ams.confex.com/ams/95Annual/webprogram/Paper263652.html

[11] R. J. Joyce and P. Xie, "Kalman filter based CMORPH," *J. Hydrometeorol.*, vol. 12, no. 6, pp. 1547–1563, 2011 doi: 10.1175/JHM-D-11-022.1.

[12] R. J. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie, "CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution," *J. Hydrometeorol.*, vol. 5, no. 3, pp. 487–503, 2004, doi: 10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2.

[13] P. J. Minnett *et al.*, "Half a century of satellite remote sensing of sea-surface temperature," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111366, doi: 10.1016/j.rse.2019.111366.

[14] K. Hosoda, "A review of satellite-based microwave observations of sea surface temperatures," *J. Oceanogr.*, vol. 66, no. 4, pp. 439–473, 2010, doi: 10.1007/s10872-010-0039-3.

[15] D. B. Chelton and F. J. Wentz, "Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research," *Bull. Amer. Meteorol. Soc.*, vol. 86, no. 8, pp. 1097–1116, 2005, doi: 10.1175/BAMS-86-8-1097.

[16] C. Prigent *et al.*, "Analysis of the potential and limitations of microwave radiometry for the retrieval of sea surface temperature: Definition of MICROWAT, a new mission concept," *J. Geophys. Res. Oceans*, vol. 118, pp. 3074–3086, 2013, doi: 10.1002/jgrc.20222.

[17] C. L. Gentemann, T. Meissner, and F. J. Wentz, "Accuracy of satellite sea surface temperatures at 7 and 11 GHz," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1009–1018, Mar. 2010.

[18] W. J. Blackwell and F. W. Chen, *Neural Networks in Atmospheric Remote Sensing*. Norwood, MA, USA: Artech House, 2009.

[19] R. K. Gangwar, A. K. Mathur, B. S. Gohil, and S. Basu, "Neural network based retrieval of atmospheric temperature profile using AMSU-A observations," *Int. J. Atmos. Sci.*, vol. 2014, 2014, Art. no. 763060, doi: 10.1155/2014/763060.

[20] V. M. Krasnopolsky, "Neural network emulations for complex multidimensional geophysical mappings: Applications of neural network techniques to atmospheric and oceanic satellite retrievals and numerical modeling," *Rev. Geophys.*, vol. 45, 2007, Art. no. RG3009, doi: 10.1029/2006RG000200.

[21] V. M. Krasnopolsky, M. S. Fox-Rabinovitz, and A. A. Belochitski, "Decadal climate simulations using accurate and fast neural network emulation of full, longwave and shortwave, radiation," *Monthly Weather Rev.*, vol. 136, no. 10, pp. 3683–3695, 2008, doi: 10.1175/2008MWR2385.1.

[22] Y. Lee, D. Han, M.-H. Ahn, J. Im, and S. J. Lee, "Retrieval of total precipitable water from Himawari-8 AHI data: A comparison of random forest, extreme gradient boosting, and deep neural network," *Remote Sens.*, vol. 11, no. 15, 2019, Art. no. 1741, doi: 10.3390/rs11151741.

[23] A. Manzato, "Hail in Northeast Italy: A neural network ensemble forecast using sounding-derived indices," *Weather Forecasting*, vol. 28, no. 1, pp. 3–28, 2013, doi: 10.1175/WAF-D-12-00034.1.

[24] Y. Zhou and C. Grassotti, "Development of a machine learning-based radiometric bias correction for NOAA's microwave integrated retrieval system (MiRS)," *Remote Sens.*, vol. 12, 2020, Art. no. 3160, doi: 10.3390/rs12193160.

[25] S. L. Castro, G. A Wick, and W. J. Emery, "Evaluation of the relative performance of sea surface temperature measurements from different types of drifting and moored buoys using satellite-derived reference products," *J. Geophys. Res. Atmos.*, vol. 117, 2012, Art. no. C02029, doi: 10.1029/2011jc007472.

[26] Canadian Meteorological Centre, "GHRSST level 4 CMC0.1deg global foundation sea surface temperature analysis (GDS version 2)," NOAA National Centers for Environmental Information, Dataset, 2017. Accessed: Jul. 16, 2021. [Online]. Available: https://www.ncei.noaa.gov/archive/accession/GHRSST-CMC0.1deg-CMC-L4-GLOB

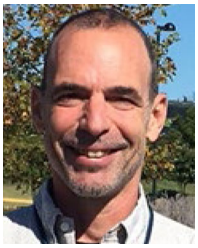
[27] B. Brasnett and D. Surcel-Colan, "Assimilating retrievals of sea surface temperature from VIIRS and AMSR2," *J. Atmos. Ocean. Technol.*, vol. 33, pp. 361–375, 2016, doi: 10.1175/jtech-d-15-0093.1.

[28] R. W. Reynolds, T. M. Smith, C. Liu, D. B. Chelton, K. S. Casey, and M. G. Schlax, "Daily high-resolution-blended analyses for sea surface temperature," *J. Climate*, vol. 20, pp. 5473–5496, 2007, doi: 10.1175/2007JCLI1824.1.

[29] B. Huang *et al.*, "Improvements of the daily optimum interpolation sea surface temperature (DOISST) version 2.1," *J. Climate*, vol. 34, pp. 2923–2939, 2021, doi: 10.1175/JCLI-D-20-0166.1.

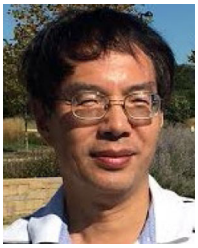
Shuyan Liu received the B.S. and M.S. degrees in atmospheric physics and environment from the Nanjing Institute of Meteorology, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in meteorology from the Nanjing University of Information Science and Technology, Nanjing, China, in 2006.

From 2007 to 2011, she was a Climate Model Developer with Illinois State Water Survey Division, University of Illinois at Urbana-Champaign, Champaign, IL, USA. From 2011 to 2015, she was with the Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA. Since 2015, she has been with the Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO, USA, and with the NOAA Center for Satellite Application and Research, National Environmental Satellite, Data, and Information Service, College Park, MD, USA.



Christopher Grassotti received the B.S. degree in earth and space science from the State University of New York at Stony Brook, Stony Brook, NY, USA, in 1982, the M.S. degree in meteorology from the University of Wisconsin-Madison, Madison, WI, USA, in 1986, and the M.S. degree in viticulture and enology from L'Institut Agro Montpellier, Montpellier, France, in 2007.

From 1986 to 1991 and again from 1993 to 2005, he was a Research Associate and Senior Research Associate with the Atmospheric and Environmental Research, Inc. From 1991 to 1993, he was with the Atmospheric Environmental Service, Environment Canada, Dorval, QC, Canada. Since 2008, he has been with the National Oceanic and Atmospheric Administration, NOAA Center for Satellite Application and Research, National Environmental Satellite, Data, and Information Service, College Park, MD, USA.



Quanhua Liu received the B.S. degree in physics 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 Sciences, Beijing, China, in 1984, and the Ph.D. degree in meteorology and remote sensing from the University of Kiel, Kiel, Germany, in 1991.

He is currently a Physical Scientist with the NOAA/NESDIS and is leading the soundings team at NOAA/STAR, College Park, MD, USA. The sounding system utilized satellite-based microwave observations and infrared hyperspectral measurements to acquire vertical profiles of atmospheric temperature, water vapor, ozone, CO, CH4, CO_2, and others chemical species. He studied the infrared hyperspectral sensor and the community radiative transfer model (CRTM). The CRTM has been operationally supporting satellite radiance assimilation for weather forecasting. The CRTM also supports JPSS/NPP and GOES-R missions for instrument calibration, validation, long-term trend monitoring, and satellite retrieved products.

Yan Zhou received the B.S. degree from Zhejiang University, Hangzhou, China, in 2006, the M.S. degree from the University of Georgia, Athens, GA, USA, in 2008, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2014, all in the major of atmospheric science.

From 2015 to 2018, she worked for the CICS-MD Observing System Simulation Experiment Project with the Earth System Science Interdisciplinary Center (ESSIC), University of Maryland. Since 2019, she has been with the ESSIC, in cooperation with NOAA/STAR for the Microwave Integrated Retrieval System Project.

Yong-Keun Lee received the B.S. and the M.S. degrees in atmospheric science from Seoul National University, Seoul, South Korea, in 1994 and 1996, respectively, and the Ph.D. degree in atmospheric science from Texas A&M University, College Station, TX, USA, in 2006.

From 2006 to 2018, he was with the Space Science and Engineering Center, University of Wisconsin. Since 2018, he has been with the Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA, and the National Oceanic and Atmospheric Administration, NOAA Center for Satellite Application and Research, National Environmental Satellites, Data, and Information Service, College Park, MD, USA.