

A Deep Learning Approach to Improve the Retrieval of Temperature and Humidity Profiles From a Ground-Based Microwave Radiometer

Xing Yan¹, Chen Liang, Yize Jiang¹, Nana Luo, Zhou Zang, and Zhanqing Li¹

Abstract—The ground-based microwave radiometer (MWR) retrieves atmospheric profiles with a high temporal resolution for temperature and humidity up to a height of 10 km. Such profiles are critical for understanding the evolution of climate systems. To improve the accuracy of profile retrieval in MWR, we developed a deep learning approach called batch normalization and robust neural network (BRNN). In contrast to the traditional backpropagation neural network (BPNN), which has previously been applied for MWR profile retrieval, BRNN reduces overfitting and has a greater capacity to describe nonlinear relationships between MWR measurements and atmospheric structure information. Validation of BRNN with the radiosonde demonstrates a good retrieval capability, showing a root-mean-square error of 1.70 K for temperature, 11.72% for relative humidity (RH), and 0.256 g/m³ for water vapor density. A detailed comparison with various inversion methods (BPNN, extreme gradient boosting, support vector machine, ridge regression, and random forest) has also been conducted in this research, using the same training and test data sets. From the comparison, we demonstrated that BRNN significantly improves retrieval accuracy, particularly for the retrieval of temperature and RH near the surface.

Index Terms—Deep learning, humidity, microwave radiometer (MWR), temperature.

I. INTRODUCTION

MONITORING the vertical profiles of atmospheric temperature and humidity is of great interest in understanding the evolution of climate systems [1]. Temperature and humidity structure determine atmospheric stability [2] and these meteorological quantities are also important parameters for numerical weather prediction models [3]. In addition to information about the temperature and humidity profiles, their temporal evolution is particularly valuable in atmospheric boundary layer studies [4].

Manuscript received June 20, 2019; revised November 7, 2019 and February 12, 2020; accepted April 10, 2020. This work was supported in part by the National Key Research and Development Plan of China under Grant 2017YFC1501702, in part by the National Natural Science Foundation of China under Grant 91544217, Grant 41801329, and Grant 91837204, and in part by the Fundamental Research Funds for the Central Universities. (Corresponding author: Zhanqing Li.)

Xing Yan, Chen Liang, Yize Jiang, Nana Luo, and Zhou Zang are with the State Key Laboratory of Remote Sensing Science, College of Global Change and Earth System Science, Beijing Normal University, Beijing 100875, China.

Zhanqing Li is with the Earth System Science Interdisciplinary Center, Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742 USA (e-mail: zli@atmos.umd.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2020.2987896

Currently, atmospheric vertical profiles of temperature and humidity can be obtained from the global radiosonde network [5]. Although the data acquired from this source are highly accurate, observations are limited by their cost and low temporal resolution (mainly once or twice a day), which is insufficient to capture the diurnal variation of the atmospheric structure [6]. Another way to obtain the profile information is by the use of passive satellite-based remote sensing [7]. However, the signal received by the satellite combines both surface and atmosphere signals; it is difficult to separate them and the surface signal could cause substantial errors when retrieving profiles near the ground [8]. Massaro *et al.* [6] indicated that passive satellite-based remote sensing could not provide sufficiently accurate profile data near ground level, so it was not recommended for use for boundary layer studies. To overcome this limitation, researchers have found that the ground-based microwave radiometer (MWR) is capable of providing valuable information on atmospheric profiles in the lower troposphere [9]–[11]. The advantages of atmospheric vertical profiles—such as temperature or humidity profiles—derived from MWR are its continuous measurements, which can provide high temporal resolution (1 min or less) data, and its ability to function in most normal weather conditions except on rainy days or other extreme conditions [6]. Thus, the MWR can capture a complete process in a weather event [12].

In the last decade, to improve retrieval accuracy, methods for MWR have rapidly developed: these include linear statistical inversion [13], the optimal estimation method [14], the one-dimensional variational (1D-VAR) retrieval method [15], and neural networks [12]. Among these methods, neural networks can offer the best performance for the solution of nonlinear relationships in the model, which is important for the retrieval of humidity profiles [16]. Measure *et al.* [17] first proposed a two-layer neural network for the inversion of radiometric measurements. Subsequently, Churnside *et al.* [18] extended this neural network to three layers (one output layer, one hidden layer, and one input layer) with the backpropagation algorithm. Even now, the three-layer backpropagation neural network (BPNN) is a very popular method for MWR, to retrieve atmospheric vertical profiles [3], [12], [19]. However, these BPNN models have only a single hidden layer and their capacity to model highly varying functions defining nonlinear structures is much less than using multiple hidden layers [20]. In addition, a traditional BPNN does not consider the distribution of the input changes of each layer during training,

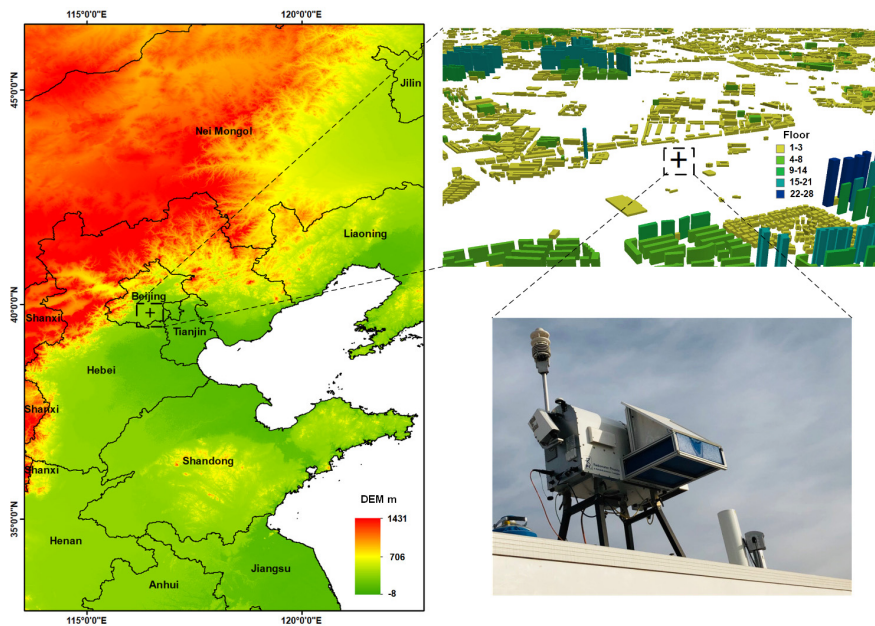


Fig. 1. Data measurement location.

as the parameters of the previous layers change, which slows down the training process [21]. This time-consuming problem is particularly serious when the amount of data is large. Last but not least, BPNNs for MWR usually employ a sigmoid [19] or tansig activation function [12], but Wang *et al.* [22] indicate that these common nonlinear activation functions can cause gradient vanishing and exploding problems in the training process.

In this article, we propose a deep learning approach called batch normalization and robust neural network (BRNN), to retrieve temperature and humidity profiles using data from a ground-based MWR. A comprehensive discussion of this method and a comparison with BPNN, extreme gradient boosting (XGBoost), support vector machine (SVM), ridge regression, and random forest (RF) are also presented.

II. DATA AND METHODS

A. MWR Data

In this article, the data were measured by an MWR located in the Beijing Nanjiao Meteorological Observatory (39.81° N, 116.48° E), China. The MWR used in this research was the Humidity And Temperature PROfiler (HATPRO; Radiometer Physics GmbH, Germany). The RPG-HATPRO is a 14-channel ground-based passive MWR, including seven water vapor absorption channels (K-band) between 22 and 30 GHz and seven oxygen absorption (V-band) channels between 51 and 59 GHz. The brightness temperatures measured by the RPG-HATPRO have a temporal resolution of 1 s [2], [5], [6], [12]. Furthermore, this RPG-HATPRO contains sensors to measure surface temperature, relative humidity (RH), and pressure. In the experimental period, we performed a liquid nitrogen calibration every six months; the details of the calibration process can be found in [23]. The overall brightness temperature accuracy in this RPG-HATPRO

is 0.5 K. As shown in Fig. 1, the RPG-HATPRO was installed on the roof of the laboratory cabin, with no high buildings nearby. There is a meteorological sensor onboard the instrument to collect the surface pressure, temperature, and RH data. The instrument also contains a zenith-pointing infrared radiometer to measure the cloud-based temperature. The radiosonde data were measured by an L-band GTS1 digital radiosonde at the same location; the radiosonde was launched twice a day, at 11:15 and 23:15 UTC during the research period. The collected MWR and radiosonde data from 2017 to 2018 were used in the training process and the data from January to May 2019 were applied for validation.

B. Preprocessing of the Data

First, the brightness temperatures from the MWR and the measurements from the radiosonde were checked by specific rules to assure the data quality. For the test and validation period from January to May 2019, only data collected in clear-sky conditions were considered sufficiently reliable. In order to avoid incorrect retrieval results, the data collected in cloudy, rainy, or other such uncertain conditions were removed. The radiosonde data were used to determine the weather conditions and estimate the cloud parameters. Theoretically, cloud formation usually occurs when the RH reaches 100%. However, in actual atmospheric conditions, clouds can form because of the existence of cloud condensation nuclei when RH reaches around 85% [24]. In the radiosonde data: 1) if the RH was greater than 84% from the ground to 600 m, the measured data were classified as rainy conditions, and if the RH was less than 84% near the surface but greater than 84% in the upper atmosphere, the data were classified as cloudy conditions [12] and other data were classified as clear sky conditions; and 2) the brightness temperature was in the 2.7–330 K range [6]. For the radiosonde

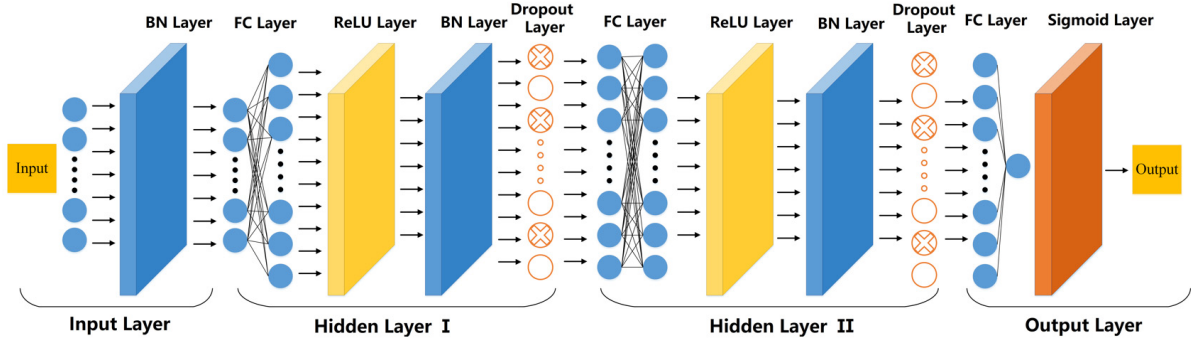


Fig. 2. Schematic of the BRNN.

measurements, it was checked that: 1) pressure was in the 1–1050 hPa range; 2) temperature was in the 210–330 K range; and 3) surface pressure was greater than 500 hPa [6]. In this research, 84% was selected as the threshold on RH for judging rain or cloud occurrence, following the study by Wang and Rossow [24]. The rain information provided by rain sensors onboard the MWR instrument was also used as a reference. A total of 679 quality-assured MWR matched with radiosonde sounding data were used as the training samples between 2017 and 2018, and 88 quality-assured test samples from 2019 were used for validation. Before the data were input to the BRNN for training or prediction, each input data item was normalized by the Z-score method [25]

$$\text{Normalized}(x_i) = \frac{x_i - \text{mean}(x)}{\text{std}(x)} \quad (1)$$

$$\text{std}(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n [x_i - \text{mean}(x)]^2}. \quad (2)$$

C. Batch Normalization and Robust Neural Network

The developed neural network, BRNN, consists of four layers (see Fig. 2): one input layer, two hidden layers, and one output layer. The input layer receives the collected data, which comprise 17 features. Among these, 14 features consist of the brightness temperature data from the RPG-HATPRO's 14 channels and three features are the surface pressure, temperature, and RH measured by the RPG-HATPRO. These features are then input to a batch normalization (BN) layer. The purpose of the BN layer is to standardize the distribution of the layer inputs x as the training progresses, which has the effect of stabilizing the learning process [21].

In BRNN, each of the two hidden layers includes one fully connected layer, one rectified linear unit (ReLU) layer, one BN layer, and one dropout layer. In the fully connected layer, the number of neurons is 256. We used the ReLU as the activation function in this study because it can overcome the problems of saturation and vanishing gradients [26] and is much faster than the traditional activation functions used in the BPNN, such as the sigmoid activation function [21]. To prevent overfitting, we introduced a dropout layer to each hidden layer. The dropout method has been demonstrated to significantly reduce overfitting and improve the performance

of neural networks [27]. The feed-forward operation in this hidden layer of BRNN can be described as

$$\tilde{y}^m = r^m \times y^m \quad (3)$$

$$z_i^{m+1} = W_i^{m+1} \tilde{y}^m + b_i^{m+1} \quad (4)$$

$$P_i^{m+1} = \text{BN}\{f(z_i^{m+1})\} \quad (5)$$

where m is the index of the hidden layer; z^m is the vector of inputs into layer m ; y^m is the vector of outputs from layer m (y^0 is the vector of outputs from the input layer); W^m and b^m are the weights and biases at layer m , respectively; P_i^{m+1} is the output from hidden layer $m+1$ at the neuron node i ; r^m is a vector of independent Bernoulli random variables, each of which has probability p of being 1 and probability $1-p$ of being 0; \tilde{y}^m is the thinned vector of y^m ; and $\text{BN}\{\}$ is the BN function, the details of which can be found in [21]. The $f(\cdot)$ function is the ReLU activation function

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x < 0. \end{cases} \quad (6)$$

In the output layer, the input data will first be processed by a fully connected layer and a sigmoid layer. We introduce the sigmoid layer to scale the output to a reasonable range in both training and prediction processes. For example, the normal range of the RH is 0%–100%.

The final output result of the BRNN is the 47-feature vertical profiles of either the temperature or humidity corresponding to different heights. The vertical resolution in this study is every 100 m between the heights of 0 and 1 km and every 250 m between the heights of 1 and 10 km.

D. Other Retrieval Techniques

In addition to our proposed BRNN method, described earlier, we used several other machine learning models for the comparison.

- 1) XGBoost is a gradient boosting-based integrated learning algorithm proposed by Chen and Guestrin [28]. XGBoost can automatically use central processor (CPU) multithreading to carry out parallel computation and is an efficient algorithm. Pan [29] developed an XGBoost algorithm to forecast hourly PM_{2.5} (particulate matter with an aerodynamic diameter less than 2.5 μm) concentrations in Tianjin by analyzing air quality monitoring

data. Zhai and Chen [30] also applied XGBoost in their stacked ensemble model to predict and analyze daily average $\text{PM}_{2.5}$ concentrations in Beijing and concluded that their model was highly interpretable and generalizable in forecasting atmospheric pollution. However, the applications of the XGBoost algorithm for the retrieval of meteorological profiles using MWR are still few in number.

- 2) The BPNN is a multilayer feedforward neural network that targets the minimum squared error of prediction and adjusts the weight and threshold of the network to approximate the expected value [31]. Solheim *et al.* [32] analyzed various retrieval methods by using synthetic data and found that the neural network method outperformed other methods for retrieving temperature and water vapor profiles from radiometric data. Even now, BPNN is one of the most commonly used methods for atmospheric temperature and humidity profile retrieval from ground-based MWRs. In this study, the structure of the BPNN used is the same as in [12] in clear sky conditions.
- 3) SVM was first proposed by Vladimir [33]. It has many unique advantages in solving small sample, nonlinear, and high-dimensional pattern recognition problems, and can be applied to function fitting and other machine learning problems [34]–[36]. The SVM method is based on the Vapnik–Chervonenkis dimension and the structural risk minimization principle of statistical learning theory. The basic idea of SVM is to find the optimal hyperplane between the positive samples and negative samples [37]. The SVM method has also been applied to atmospheric environment research [37]. Zhang *et al.* [39] used SVM to set up a dynamic model that improved the accuracy of forecasting $\text{PM}_{2.5}$ concentrations 1 h in advance. However, SVM methods have rarely been used for meteorological parameter profile inversion based on MWR.
- 4) RF is a powerful machine learning method that constructs ensembles of unpruned classification or regression trees generated by the selection of random features in tree induction and bootstrap samples of training data. It performs well when the number of variables is much larger than the number of observed values and can handle high-dimensional data and complex interactive structures without feature selection. It has high training speed and easy parallelization of calculation, and can return highly correlated and important characteristic variables. The RF algorithm has often been applied to atmospheric environment research in recent years [40], [41].
- 5) Ridge regression is a biased estimation regression method dedicated to collinear data analysis, particularly for ill-conditioned data, for which it is stronger than the least squares method. Compared with the least squares method, ridge regression abandons the unbiased requirement from the perspective of reducing the accuracy and loss of partial information, and still obtains reliable and realistic estimates [42].

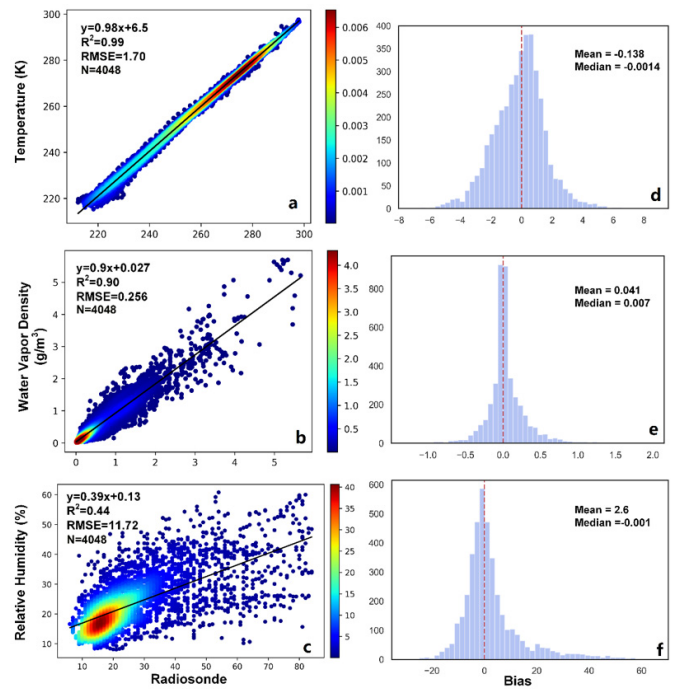


Fig. 3. (a)–(c) BRNN retrievals for temperature, WVD, and RH, respectively, as a function of radiosonde data from all 47 atmospheric vertical layers. The black solid line is the fit line from linear regression. The regression equations and coefficients of determination (R^2) are given, as well as the number of data points ($N = 4048$) and the RMSEs. Colored areas show the density of data points. (d)–(f) Histograms of the BRNN bias (BRNN retrieval minus radiosonde). The red dashed line is the zero line.

III. RESULTS

A. Validation of BRNN With Radiosonde

To test the performance of the neural network proposed in this study, results from BRNN and from the radiosonde were compared. Fig. 3(a)–(c) shows the BRNN temperature, water vapor density (WVD), and RH as a function of the radiosonde measurements from all 47 atmospheric vertical layers up to 10 km. High kernel density values with red color show where most of the data lie. As shown in Fig. 3(a), the linear regression relation between the BRNN temperature and radiosonde temperature has a slope of 0.98 and a y-intercept of 6.5, with a coefficient of determination (R^2) of 0.99 and a root-mean-square error (RMSE) of 1.70. For WVD [see Fig. 3(b)], the R^2 is 0.90 and the slope is 0.9 with an RMSE of 0.26. From Fig. 3(a) and (b), we observe that the temperature and WVD from BRNN agree well with the radiosonde measurements. In contrast to the temperature and WVD validation, the result of RH is more scattered. Fig. 3(c) shows that the R^2 is 0.44 and the RMSE is 11.72. The frequency distribution of the bias (radiosonde minus BRNN retrieval) is plotted in Fig. 3(d)–(f). Candlish *et al.* [11] indicate that, if the data set is sufficiently large and the errors are random, the frequency distributions of the bias should be normal and centered on 0. Our results are consistent with this phenomenon: all biases of BRNN retrievals have a noticeable normal distribution with a median close to 0 (the biases for temperature, WVD, and RH are -0.0014 , 0.007 , and -0.001 , respectively). Fig. 3 shows that the results for RH are not

TABLE I
COMPARISON OF THE PERFORMANCE OF DIFFERENT METHODS

	Method	RMSE	R^2	StDev	MAE
Temperature (K)	BRNN	1.70	0.99	1.69	1.29
	BPNN	2.14	0.99	2.13	1.67
	SVM	2.06	0.99	2.06	1.57
	XGB	2.31	0.99	2.30	1.72
	Ridge	3.33	0.98	3.32	2.56
	RF	3.85	0.97	3.79	3.04
RH (%)	BRNN	11.72	0.44	11.44	0.074
	BPNN	13.79	0.25	13.23	0.095
	SVM	13.17	0.25	13.15	0.098
	XGB	14.07	0.14	14.06	0.099
	Ridge	14.64	0.09	14.54	0.103
	RF	14.55	0.09	14.55	0.104
WVD (g/m ³)	BRNN	0.256	0.90	0.25	0.16
	BPNN	0.288	0.88	0.28	0.20
	SVM	0.287	0.88	0.28	0.17
	XGB	0.285	0.88	0.29	0.18
	Ridge	0.679	0.65	0.68	0.57
	RF	0.439	0.78	0.42	0.28

as good as those for temperature and WVD. Miaci and Angelis [46] indicated that humidity profiles have expected errors of approximately 5%–25% depending on altitude, with the larger error at high altitudes due to the radiometer's spatial resolution.

B. Comparison of BRNN With Other Retrieval Techniques

Table I illustrates the retrieval performance of the six methods for temperature, RH, and WVD retrieval. For temperature retrieval, the R^2 for all of the methods is above 0.96, which shows that the temperature obtained by all six methods is highly correlated with the radiosonde data. Although the methods have little difference in temperature retrieval, in terms of R^2 , only the BRNN method can make the RMSE less than 2 K. In the results of RH, the R^2 for the six methods is obviously decreased along with the R^2 of the temperature. Among these methods, BRNN achieves the best accuracy in terms of RMSE (11.72), standard deviation (StDev) (11.44), and mean average error (MAE) (0.074), in comparison with the radiosonde. It is noticeable that the SVM and BPNN methods also perform quite well for RH retrieval, with RMSE of 13.17 and 13.79. For WVD, BRNN also achieves good retrieval ability, with an RMSE of 0.256 and an R^2 of 0.90. In addition to BRNN, the XGBoost method performs better than the remaining methods in terms of RMSE (0.285), compared with SVM (0.287) and BPNN (0.288). The value of R^2 (0.88) is the same as for the SVM, BPNN, and XGBoost methods, indicating that these three methods are comparable for WVD retrieval. However, as shown in Fig. 4, the BPNN and SVM can output negative values for WVD. Because BPNN and SVM do not limit the WVD range in the training and prediction processes, it is not surprising that anomalous results are output. Nevertheless, this limitation has been improved in BRNN by introducing a sigmoid layer to

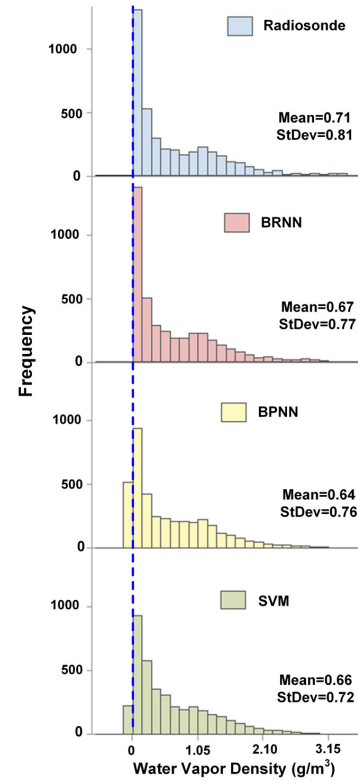


Fig. 4. Histograms of the WVD obtained by radiosonde, BRNN, BPNN, and SVM. The blue dashed line is the zero line.

the output layer. From Fig. 4, we observe that the frequency of the BRNN-based result agrees well with the radiosonde result and the BRNN produces no extreme values. As shown in Table I, BRNN and XGBoost achieve similar performance for temperature and WVD retrieval, even though there are fundamental differences between these two algorithms. XGBoost is a tree-based model, which is an interpretation-focused method, whereas neural network-based BRNN allows for the encoding of more structural information and is more suitable for large amounts of training data. According to Table I, the performance of the ridge regression and RF methods—for all three parameters—seems unsatisfactory compared with the other methods, indicating that these two methods may not be suitable for MWR retrieval of temperature and humidity profiles. Cule and Iorio [47] indicated that ridge regression can work well when the multicollinearity problem prevails. However, the nonlinear relationships found between variables in MWR-based humidity retrieval [16] make this unsuitable for ridge regression.

C. Bias and RMSE Variation With Height

The temperature profile retrieval bias of the various techniques is presented in Fig. 5. The red dots represent the mean value of the bias and the short blue lines inside the box represent the median. The blue shadow means ± 1 -K bias of temperature. The left and right boundaries of the box contain the data from the first quartile to the third quartile. The whiskers (blue dotted lines) extend from each quartile to the minimum or maximum of the bias. As shown

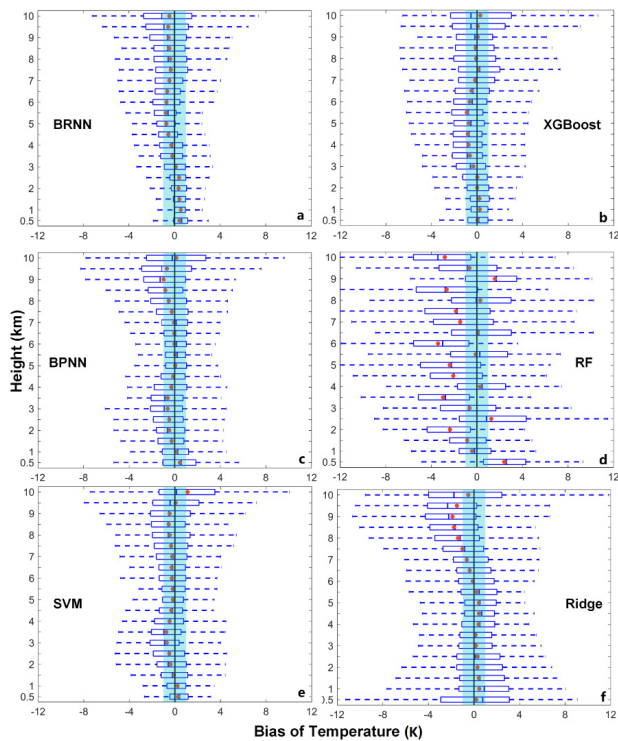


Fig. 5. Temperature retrieval bias (The retrieval minus radiosonde). The means, medians, and 66% ($1 - \sigma$) intervals are shown as red dots, horizontal lines within the boxes, and the boxes themselves, respectively. The blue shadow means ± 1 -K bias of temperature. (a)–(f) Retrieval bias of the BRNN, XGBoost, BPNN, RF, SVM, and Ridge.

in Fig. 5(a), most of the BRNN-based temperature bias is within ± 1 K below 3 km, indicating that BRNN achieves a high accuracy for retrieving temperature profiles near the surface. At heights between 3 and 10 km, the length of the box for the BRNN method is shorter than for the other methods [see Fig. 5(b)–(f)], showing that the bias of derived temperature is more concentrated near the mean and median. In addition, the mean and the median of the temperature bias by the BRNN method are very close, showing that the temperature bias of BRNN is more uniform and centralized. For the fluctuation trend up to a height of 10 km, the mean temperature bias obtained by BRNN is positive near the surface, decreases with height, and becomes negative above 3.5 km; this is similar to the trends of [2] and [43]. Cimini *et al.* [16] also found that the retrieval temperature in the upper atmosphere has a negative bias of 1–2 K. At the middle layer (5.5 km), the negative mean bias reaches its maximum and then gradually decreases to 0. In Fig. 5, the bias of RF shows a large scatter over all height layers. This is because for RF, the range of predictions is bound by the highest and lowest values in the training data. This becomes problematic in situations where the training and prediction inputs differ in their range and/or distributions. This so-called covariate shift is difficult for most models to handle, but it is especially so for RF [48].

Fig. 6 shows the bias of RH for the six methods; this is the same as Fig. 5 but the blue shadows show the bias range of $\pm 8\%$. In all of the six methods, the mean bias is usually close to the median in the lower levels, but drastically deviates

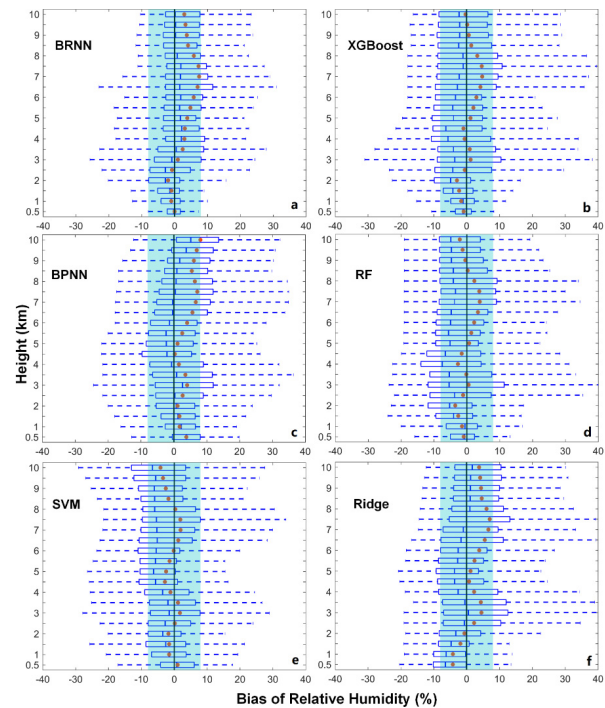


Fig. 6. Same as Fig. 5 but for RH retrieval bias. The blue shadow means a bias of $\pm 8\%$ of the RH. (a)–(f) Retrieval bias of the BRNN, XGBoost, BPNN, RF, SVM, and Ridge.

from the median above 2 km. The RH bias of the BRNN method remains within $\pm 8\%$ below 3 km, and its interquartile range is almost the smallest of all six methods at each height. As for SVM, BPNN, and XGBoost, their bias also remains in the range of $\pm 8\%$ near the surface, but the interquartile range is much larger than BRNNs and the maximum bias exceeds 30%.

Fig. 7 shows the bias of the derived WVD; this is the same as Fig. 5 but the blue shadows show the bias range of ± 0.15 g/m³. From Fig. 7, we observe that the bias of ridge regression and RF vary much more than the bias of the other four methods. The deviations of the bias in BRNN, XGBoost, BPNN, and SVM appear mainly in the lower atmosphere and gradually decrease above 5 km. This is consistent with [9] and [43], which found that, in water vapor profile retrieval, there were mainly positive deviations in the lower atmosphere, which tended to 0 at the top of the atmosphere.

In Fig. 8, we plot the RMSE of the temperature, RH, and WVD profiles resulting from BRNN, XGBoost, BPNN, and SVM at different heights, using the radiosonde observation as the reference. Because the RF and ridge regression retrieval performance seem poor, as mentioned earlier, we do not compare their results. For the temperature, using the BRNN method, the RMSE below 2 km is around 1 K and remains less than 2 K below 8 km; this is apparently less than the RMSE of XGBoost, BPNN, and SVM. The XGBoost method has a better accuracy than SVM and BPNN below 2 km, but has a large RMSE in the upper atmosphere (above 3.5 km). From the figure, the temperature retrieval results in BRNN are comparable to the studies of [5] and [6] in the lower

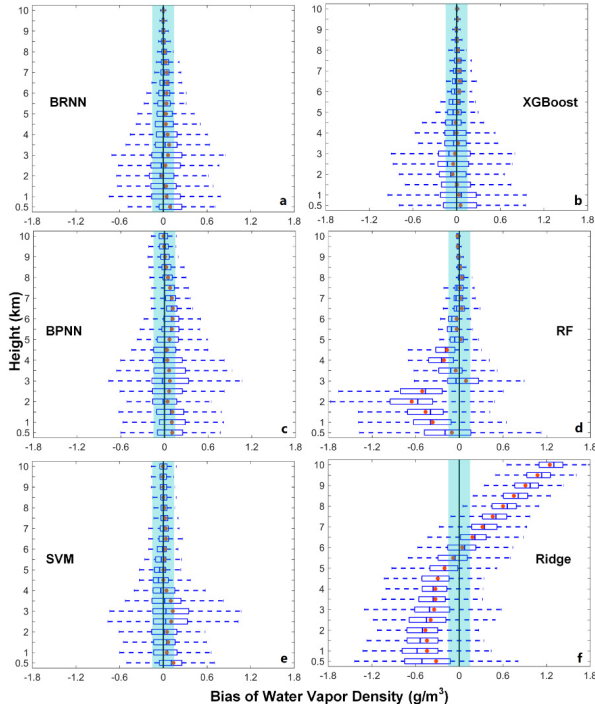


Fig. 7. Same as Figs. 5 and 6 but for WVD retrieval bias. The blue shadow means a bias of $\pm 0.15 \text{ g/m}^3$ of the WVD. (a)–(f) Retrieval bias of the BRNN, XGBoost, BPNN, RF, SVM, and Ridge.

atmosphere (below 4 km) and are even slightly better than those above 4 km: the RMSE of BRNN remains less than 1.5 K at 5.25 km and 3.5 K at 10 km. For RH, the RMSE of BRNN can be controlled at about 5% below 1 km. The XGBoost method also performs well in the lower atmosphere, but its RMSE is significantly larger than the others from 3 to 7 km. The distribution of RMSE for all four methods from surface to 10 km is similar to the results of Che *et al.* [12]: the RH RMSE tends to increase with height and the maximum deviation occurs in the middle atmosphere. In terms of WVD, the situation is exactly the opposite: the RMSE of the lower layers is larger and that of the higher layers is smaller. The superiority of the BRNN method is that its RMSE value does not exceed 0.4 g/m^3 from the surface to 10 km; in particular, its RMSE is about 0.35 g/m^3 from the surface up to 3.5 km and then decreases to below 0.1 g/m^3 at 8 km. Among the remaining three methods, the performance shows little variation, except for the SVM method, which has the largest error in the layer near the ground. Therefore, from the overall characteristics of the RMSE of the three variables, the temperature RMSE is mostly concentrated in the upper layer, the high RMSE of RH mainly occurs in the middle layer, and the high RMSE of WVD mostly appears in the lower layer; this is in accordance with the situation found by Cimini *et al.* [9] and Sanchez *et al.* [43].

D. Case Analysis

To illustrate and compare the outcomes of BRNN and other retrieval techniques, we use two cases as examples. The radiosonde measurements are used as standard and compared with these methods. A comparison of the derived

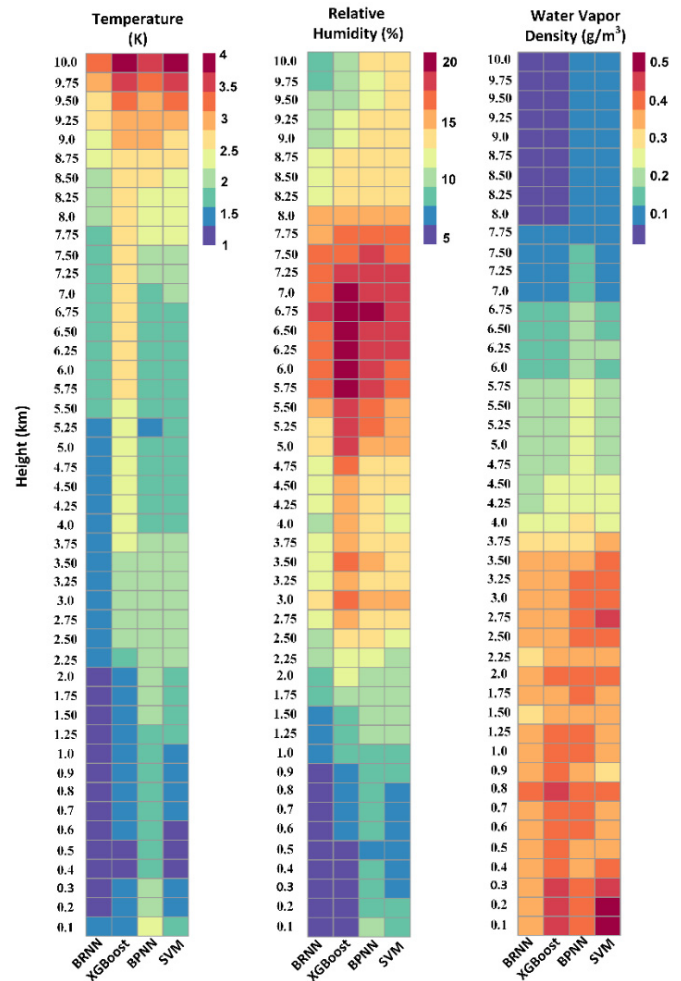


Fig. 8. Profile retrieval RMSEs for (Left) temperature, (Center) RH, and (Right) WVD, with respect to radiosonde, for BRNN, XGBoost, BPNN, and SVM techniques.

temperature profiles is presented in Fig. 9. On January 3, 2019 (23:15 UTC), BRNN shows only a small difference with the radiosonde data below the height of 5 km. In particular, between the heights of 3 and 5 km, BPNN, SVM, and XGBoost all underestimate the temperature but BRNN fits it well. On May 4, 2019 (11:15 UTC), the four methods differ slightly from the radiosonde measurements from the ground to a height of 8 km. However, above 8 km, the difference increases noticeably. Overall, from these two cases, we observe that the deviation of BRNN is less than the deviations of the other three methods near the surface.

Fig. 10 shows the results of WVD profiles by four different methods. The difference between the derived WVD profiles and the radiosonde is larger than that for temperature. In contrast to temperature retrieval, the agreement between BRNN and radiosonde WVD is less near the surface than in the upper atmosphere. Although the deviation is increased in all methods, the deviation for BRNN is clearly lower than for BPNN, XGB, and SVM.

Fig. 11 shows the derived RH profiles on April 2, 2019, and March 14, 2019. In these two cases, the profiles based on BRNN follow the RH vertical structure measured by the radiosonde closely. Although the RH profiles derived by

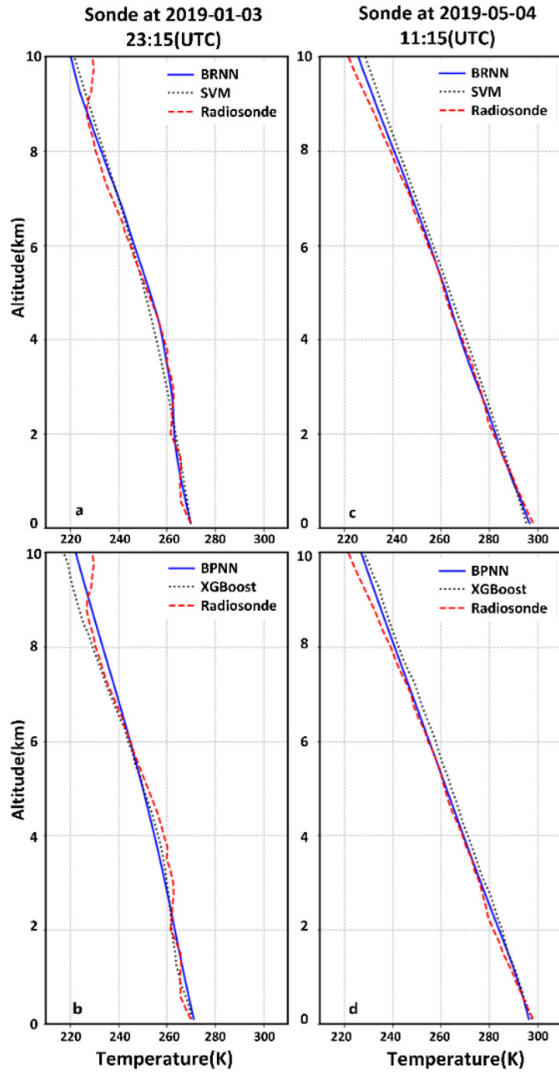


Fig. 9. Comparison between the temperature profiles generated using the BRNN, SVM, BPNN, and XGBoost retrieval methods and the radiosonde at (a) and (b) 23:15 UTC on January 3, 2019, and (c) and (d) 11:15 UTC on May 4, 2019.

BPNN and SVM still approximately follow the patterns of the radiosonde data, the deviation is clearly larger than BRNN's. Of the four methods, the XGBoost data are in the worst agreement with the radiosonde data, showing little ability to retrieve the RH vertical structure.

IV. DISCUSSION

NNs with a single hidden layer have been developed to retrieve temperature and humidity profiles in MWR for more than 20 years [18]. Although some improvements have been made for this type of NN to achieve a higher accuracy [12], its capacity to describe nonlinear relationships is much less than an NN using multiple hidden layers [20]. However, nonlinear relationships are very important in the humidity retrievals and Cimini *et al.* [16] indicate that MWR-based humidity could be improved using nonlinear minimization. As shown in [43] and [12], the correlation between the MWR-derived humidity and the radiosonde data is much less than the correlation of temperature. Thus, it is important to use a

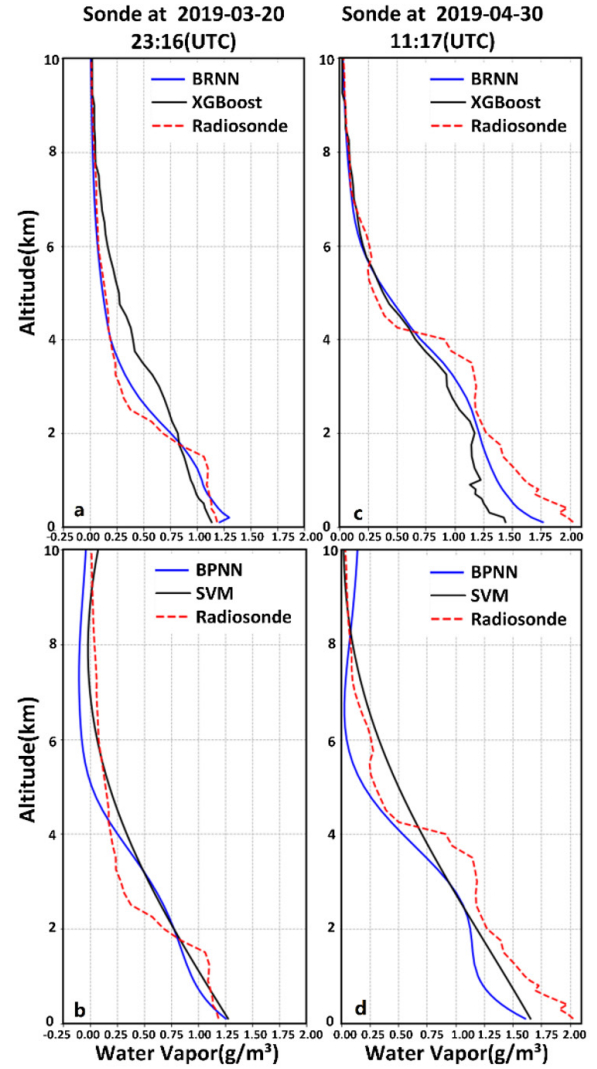


Fig. 10. Comparison between the WVD profiles generated using the BRNN, SVM, BPNN, and XGBoost retrieval methods and the radiosonde at (a) and (b) 23:16 UTC on March 20, 2019, and (c) and (d) 11:17 UTC on April 30, 2019.

better NN structure to improve the retrieval accuracy. In this research, to extend the NN for nonlinear relationship modeling, we introduced two hidden layers in the proposed BRNN. However, traditional BPNNs for MWR have only considered the individual normalization of input data, such as scaling each input data item to range from 0 to 1 [18]. Unfortunately, neglecting the variation of distribution between the input data slows down the training process and makes it difficult for BPNN to train models with saturating nonlinearities [21]. Therefore, we also normalized the data between the inputs in BRNN by BN technology, which fixes the mean and variance of the inputs. As shown in Table I, BRNN achieves a better performance than the traditional BPNN. In particular, for RH, the R^2 increases from 0.25 to 0.44 and the RMSE decreases from 13.79% to 11.72%. Another feature of BRNN is that we introduced a sigmoid layer into the output layer. As shown in Fig. 4, this can prevent abnormal results appearing in the final output. Comparing with BPNN and SVM, no negative values are output for WVD by BRNN. Generally, in machine

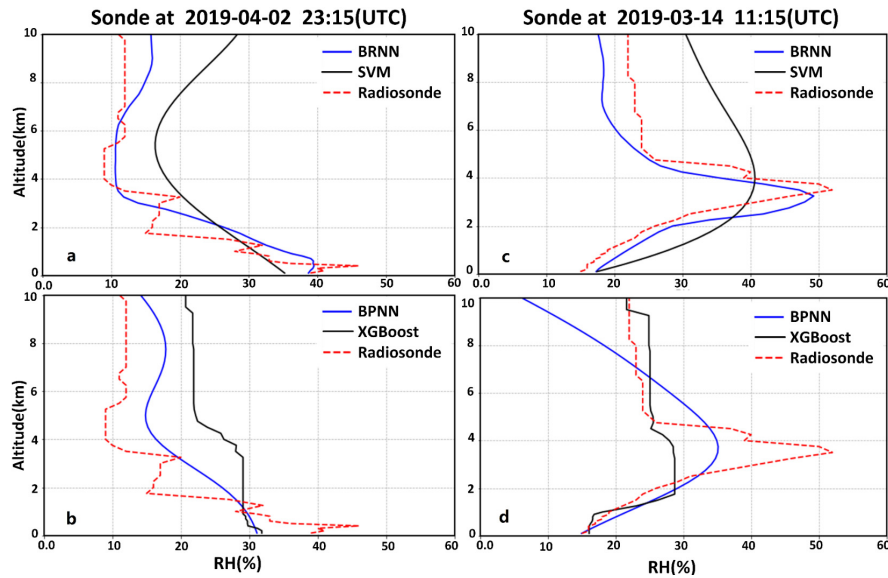


Fig. 11. Comparison between the RH profiles generated using the BRNN, SVM, BPNN, and XGBoost retrieval methods and the radiosonde at (a) and (b) 23:15 UTC on April 2, 2019, and (c) and (d) 11:15 UTC on March 14, 2019.

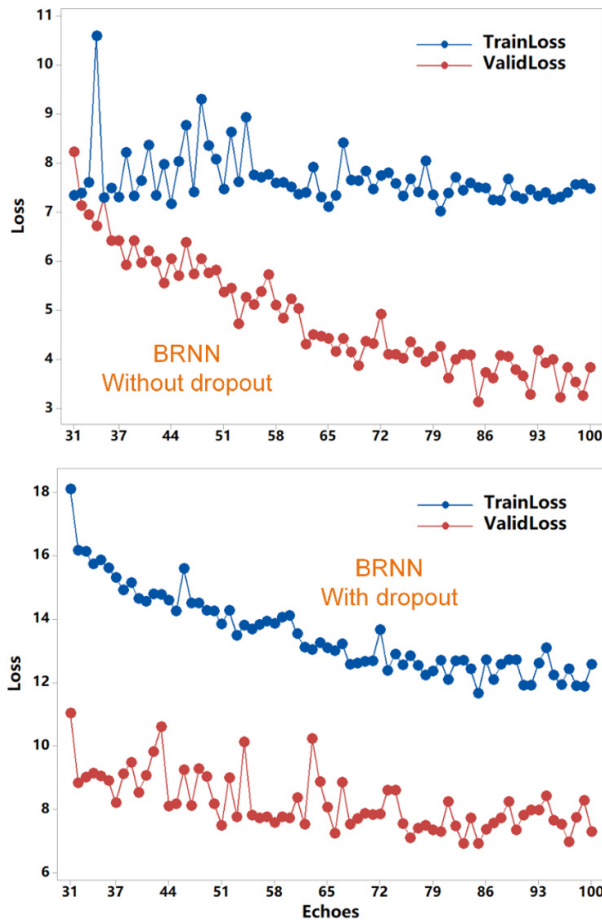


Fig. 12. Training loss (blue dotted line) and validation loss (red dotted line) variation with echoes. (Top) BRNN method without dropout. (Bottom) BRNN method with dropout.

learning models, overfitting is a serious problem which cannot be neglected [27]. To reduce the overfitting problem, we introduced a dropout layer to each hidden layer in the

BRNN. Fig. 12 shows a test for BRNN with or without dropout layer by calculating the training loss and valid loss variation with echoes. In deep neural networks, the loss function is used to measure the difference between the predicted

