Revolutionizing Humidity Profile Retrieval with Multi-Angle Aware Networks for Ground-Based Microwave Radiometers

Yinshan Yang [a], Zhanqing Li [b], Jianping Guo [c], Yuying Wang [d], Hao Wu [e], Yi Shang [d], Ye Wang [a], Langfeng Zhu [e], Xing Yan [a*]

[a] Advanced Interdisciplinary Institute of Satellite Applications, Faculty of Geographical Science, Beijing Normal University, Beijing 100875 China;

[b] Department of Atmospheric and Oceanic Sciences, Earth System Science Interdisciplinary Center, College Park, MD 21029, USA;

[c] State Key Laboratory of Severe Weather, Chinese Academy of Meteorological Sciences, Beijing 100081 China;

[d] Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science & Technology, Nanjing 210044, China;

[e] China Meteorological Administration Key Laboratory of Atmospheric Sounding, School of Electrical Engineering, Chengdu University of Information Technology, Chengdu 610225, China.

* Correspondence to: Xing Yan (yanxing@bnu.edu.cn)

Supplementary Materials

Supporting Text 1 to 3

Text 1:

The MWR utilized in this research is the RPG-HATRPO-G4 (HATPRO; Radiometer Physics GmbH, Germany), a ground-based passive radiometer designed to measures atmospheric brightness temperature (BT) at 14 frequencies across two bands and at configurable eight angles. These measurements are subsequently processed to derive temperature and humidity profiles. The MWR is equipped with seven water vapor absorption channels (K-band) ranging from 22 to 32 GHz, and seven oxygen absorption channels (V-band) between 51.00 and 59.00 GHz [26]. The K-band channels are employed to detect water vapor density, as frequencies in the K-band are more sensitive to water vapor and cloud liquid water. Conversely, the V-band channels are utilized to retrieve temperature profiles, owing to their heightened sensitivity to thermal emissions from atmospheric molecular oxygen [13, 22]. V-band frequencies can be classified in two categories: opaque channels, with frequencies of 56.66 GHz and above, which are more informative about the atmospheric layer from the surface up to about 1 km; and transparent channels, ranging from 51 to 56 GHz, which provide better information about the atmosphere above 1 km. Among these, frequencies at 56.66 GHz, 57.30 GHz, and 58 GHz are particularly sensitive to near-surface temperatures due to their proximity to the oxygen absorption line [22, 27]. The MWR operates in two modes: zenith pointing (Z-mode) and elevation scanning (S-mode). In Z-mode, the MWR points vertically, while in S-mode, it scans the atmosphere at elevation angles between 5° and 90°. The MWR employs integration times of 20–60 s per angle (user-selectable), with a total scan duration of 2-6 minutes [37]. Measurements are taken at zenith (90° elevation) and seven oblique angles (30°, 19.2°, 14.4°, 11.4°, 8.4°, 6.6°, and 5.4°). In areas with unobstructed views, oblique scans down to 5° elevation can enhance profile accuracy in the lowest atmospheric layers (0-1 km or even 0-2 km) [14, 22, 29]. The MWR provides BTs measurements with a temporal resolution of 1 s and an accuracy of 0.5 K [2, 21]. Additionally, an onboard meteorological sensor records surface pressure, temperature, and RH, while a zenith-pointing infrared radiometer measures cloud-base temperature. A rainfall sensor is integrated to identify precipitation events (flag = 0 for no precipitation, flag = 1 for precipitation). During the experimental period, a liquid nitrogen calibration was conducted every six months; the details of the calibration process can be found in [19].

The high-resolution radiosonde data used in the present study were obtained from China Meteorological Administration, which provides basic atmospheric profiles, including temperature, pressure, humidity, wind speed, and wind direction. The radiosonde was launched twice daily at 07:15 and 19:15 LST during the research period. The original measurements have a temporal resolution of approximately 1 second, a sampling frequency of $50 \cdot min^{-1}$, an ascent speed of 400 $m \cdot min^{-1}$, corresponding to the vertical resolution of about 5-8 m. The reported data accuracy is 0.1 ℃ for temperature and 1% for RH [37, 38]. Radiosonde data were employed to assess weather conditions and estimate cloud parameters using Xu's method [56], which combines vertical gradients of air temperature and RH with altitude-dependent RH thresholds.

This experiment employed a Ka-band MWCR (HMB-KST) with a working frequency of 35 GHz and a sounding range of 20 km. The radar operated in a vertically pointing mode with a vertical resolution of 30 m and measured reflectivity (Ze), Doppler velocity (VD), Doppler spectrum width (WD), and linear depolarization ratio of cloud

(LDR). The MWCR reflectivity threshold was set at -30 dBZ. A solid-state transmitter enabled continuous observations [21, 39, 40, 56].

This study also utilized the European Centre for Medium-Range Weather Forecasts Reanalysis Version 5 (ERA5) data (https://www.ecmwf.int/) from ECMWF [41]. The meteorological elements obtained from ERA5 reanalysis data include RH from 1000 to 150 hPa across 25 levels, total cloud cover, and low cloud cover. Hourly data with spatial resolution is 0.25° × 0.25° from 2018 to 2024, corresponding to the time periods of the long-term comprehensive field experiments, is used in this study. Clear-sky samples were selected to mitigate the impact of clouds, and ERA5 RH profiles were employed to fill temporal gaps in radiosonde data.

Text 2:

Measurements from the MWR, MWCR, and radiosonde were validated using specific quality control rules. Outliers were removed through cubic spline interpolation, and atmospheric profiles from different instruments were unified to vertical resolutions of 30, 60, 150, and 300 m within altitude ranges of 0–1500, 1500–3690, 3690–6090, and 6090–9990 m, respectively. This resulted in 115 layers from the surface to 10 km, aligning with the vertical resolution of the MWR.

The brightness temperatures from the MWR were monitored to ensure they fell within the range of 270 to 330 K [45]. To remove the non-hydrometeor targets(e.g., atmospheric pollutants, dust, insects) at low altitudes and retain the hydrometeor echo as much as possible from the MWCR, according to the of the Ze-LDR frequency distribution of cloud and precipitation samples and suspension samples [46] In addition, determine the cloud boundary by using morphological expansion and erosion methods, so as remove radial interference confusion and random noise. For the radiosonde measurements, it was checked that: pressure was in the range of 1 to 1050 hPa range; air temperature was in the range of 210 to 330 K; and surface pressure was greater than 500 hPa [45].

Additionally, given that the detection range of the MWR extends to 10 kilometers, it takes approximately 20 minutes for the balloons to reach this altitude. The balloons are launched at 07:15 and 19:15 LST, respectively.

Therefore, MWR and MWCR select the average data from the minute-level profiles between 07:15-07:35 and 19:15-19:35 LST to align with the profiles from the radiosonde. The comprehensive field experiment data from these time periods are time-matched with the ERA5 reanalysis data at 08:00 and 20:00 LST, following the radiosonde time-matching protocol.

Before the data were input to machine learning methods for training or prediction, each input data item was normalized by the Z -score method [26].

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

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

Text 3:

In this article, the performance of each model is evaluated using three commonly employed metrics: Coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). For RH profiles, the relative error envelope (EE) is employed to determine the accuracy of RH profiles retrieval at specific altitudes. Results within this interval are considered to fall within the acceptable error range, whereas those outside the range typically indicate overestimation or underestimation, respectively [47]. In this article, we set the EE for RH of 0-100% as ($\pm$10%).

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{1}^{n}\left|y_i - \hat{y}_i\right| \tag{5}$$

where y_i, and $\bar{y}$ represent the true value and mean value at the (i)th; $\hat{y}_i$ is the corresponding retrieval value; and n denotes the total number of samples.

Figures S1 to S4

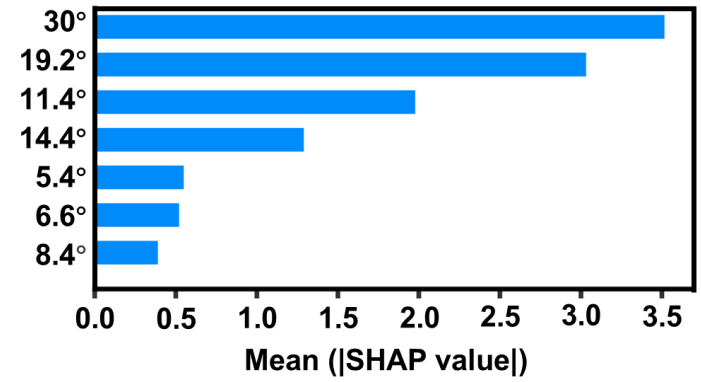


Fig. S1. Mean SHAP values of each elevation scanning angles form LGBM

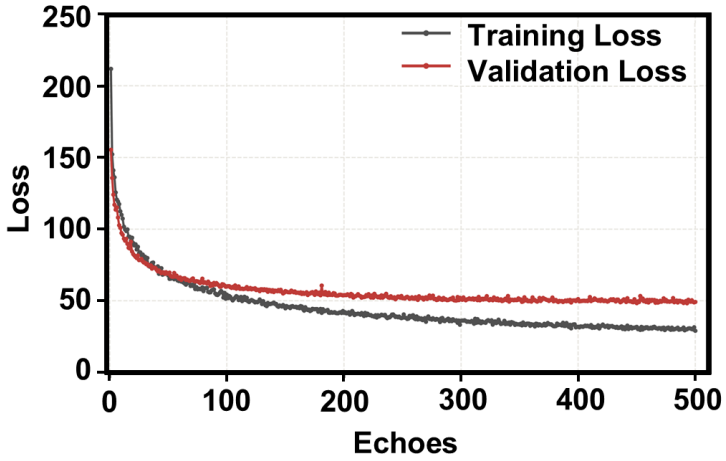


Fig. S2. AngleNet method Training loss (black dotted line) and validation loss (red dotted line) variation with echoes.

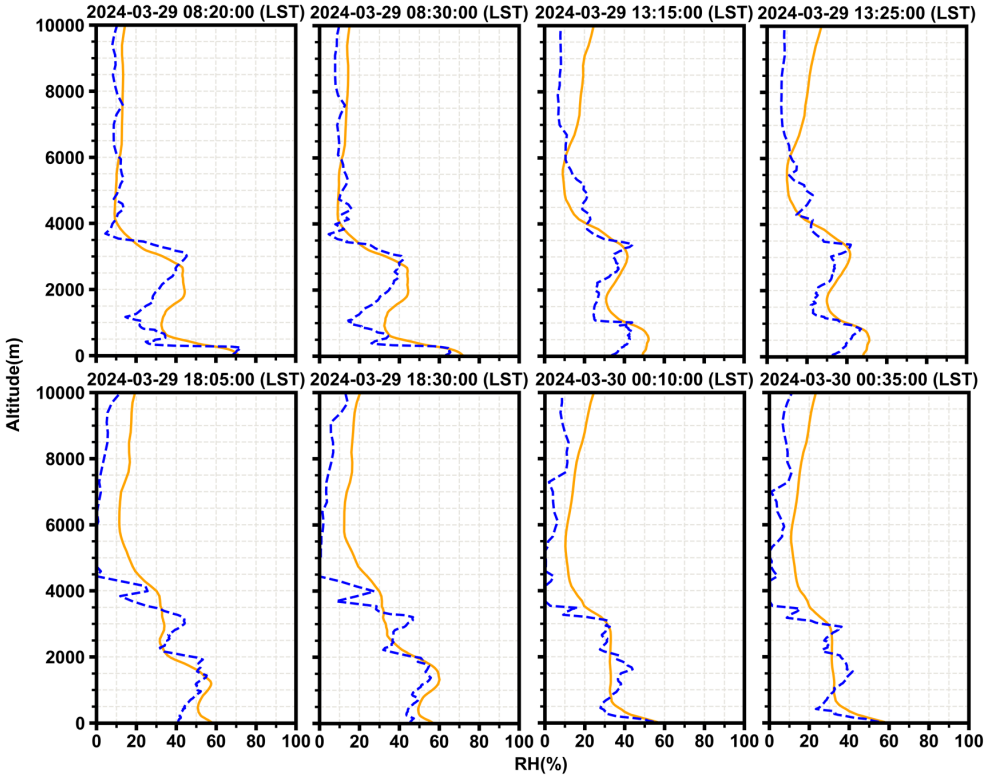
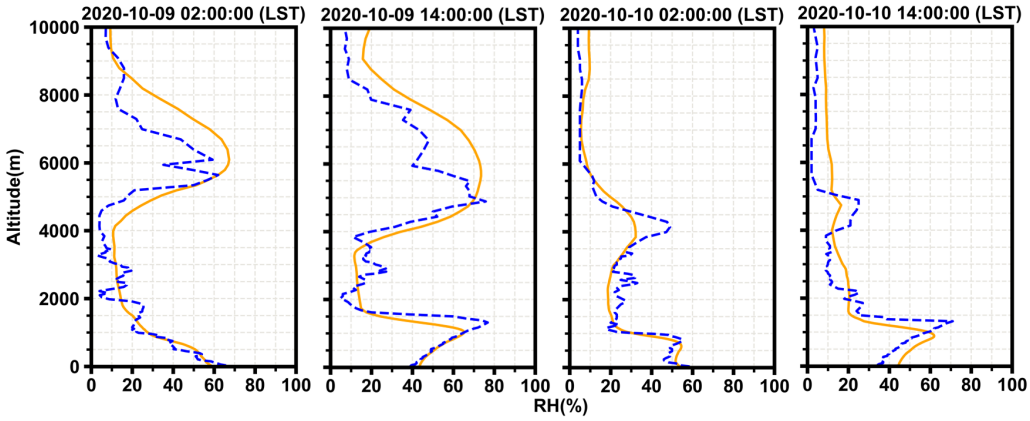


Fig. S3 Additional independent verification of comparison between the RH profiles generated from the AngleNet (blue dotted lines), and Radiosonde (orange solid lines) in clear-sky conditions at Shanghai.



127

128 **Fig. S4.** Additional independent verification of comparison between the RH profiles generated from the AngleNet

129 (blue dotted lines), and Radiosonde (orange solid lines) in clear-sky conditions at Nanjing.

Tables S1 to S7

Table S1 Observational Time Periods at the Three Cities.

Site	Observational time period	Instruments	Lon	Lat	Sample numbers
Beijing	2018.04-2019.10	Radiosonde; MWR	39.81	116.48	4097(276)
Nanjing	2020.09-2022.01	Radiosonde	31.93	118.90	3027(205)
	2022.06-2022.08	MWR; MMCR	32.18	118.72	
Shanghai	2020.04-2020.06	Radiosonde MWR;	31.39	121.45	1543(95)
	2023.11-2024.06	Radiosonde	31.39	121.45	
		MWR; MMCR	31.33	121.51	

In the last column of Table 1, the numbers outside the brackets represent the total sample numbers of profiles from MWR and MWCR matched with radiosonde and ERA5 times. At the same time, those inside parentheses indicate the sample numbers where MWR and MWCR are matched solely with radiosonde times.

Table. S2 Comparison of "multi-angle aware" input different oblique elevation angles

input angles	R^2	RMSE (%)	MAE (%)	Within EE (± (10%)
30°	0.70	10.78	7.57	72.74%
19.2°	0.70	10.44	7.46	74.18%
30°+19.2°	**0.71**	**10.39**	**7.44**	**74.84%**
30°+19.2°+11.4°	0.70	10.41	7.56	73.64%

Table S3. Summary of our model's input variables.

Variable	Class	Name	Detailed Description
Dependent variables		RH	RH profiles from radiosonde and ERA5
Independent variables	Continuous variables	C1-C14	BTs from Channels 1 to 14 of the MWR
		A1-A8	Scanning elevation angle of the MWR
		Surface pressure	Surface meteorological data from MWR as global feature
		Surface Temperature	
		Surface RH	
	Categorical variables	Location	
		Month	
		Hour	

Table. S4 Comparison of 5-fold cross-validation results for the AngleNet.

Round	**R^2**	**RMSE**	**MAE**	**Within EE (± (10%)**
1	0.69	12.01	10.43	72.85%
2	0.70	11.12	9.44	72.66%
3	0.71	10.40	7.45	74.51%
4	0.68	12.59	11.43	72.75%
5	**0.71**	**10.39**	**7.44**	**74.84%**
Best	**0.71**	**10.39**	**7.44**	**74.84%**

Table S5 Hyperparameters of the model settings.

Short Name	Configuration	Detailed Description
batch_size	64	The number of data samples captured in one training step
d_model	64	The number of hidden-layer nodes in each layer
n_tf_head	4	The number of attention heads used in the multi-head attention encoder layer
n_tf_layer	2	The number of transformer encoder layers
p_tf_drop	0.1	The percentage of neurons temporarily removed from the transformer encoder layers
n_brnn_layer	2	The number of BRNN layers
p_brnn_drop	0.1	The percentage of neurons temporarily removed from the BRNN layers
lr	0.0001	Learning rate
max_epochs	500	The maximum number of epochs to train for the model

Table S6 Additional independent validation case in Nanjing

Time (LST)	**R^2**	**RMSE**	**MAE**	**Within EE (± (10%)**
2020/11/09 02:00	0.73	8.97	6.27	84.35%
2020/11/09 14:00	0.71	11.63	7.86	73.04%
2020/11/10 02:00	0.87	5.72	4.45	91.30%
2020/11/10 14:00	0.76	9.67	7.87	88.70%

R^2 coefficients of determination; RMSE: root-mean-square error; MAE: mean absolute error.

Table S7 Additional independent validation case in Shanghai

Time (LST)	**R^2**	**RMSE**	**MAE**	**Within EE (± (10%)**
2024/03/29 08:20	0.65	9.82	7.82	61.74%
2024/03/29 08:30	0.66	9.07	7.31	73.91%
2024/03/29 13:15	0.47	8.31	7.36	74.78%
2024/03/29 13:25	0.49	8.18	6.99	76.52%
2024/03/29 18:05	0.75	9.24	7.97	61.74%
2024/03/29 18:30	0.77	8.80	7.33	73.91%
2024/03/30 00:10	0.71	7.77	6.46	80.01%
2024/03/30 00:35	0.65	8.34	6.99	74.78%

R^2 coefficients of determination; RMSE: root-mean-square error; MAE: mean absolute error.