Review for: ***"Evaluation of the New York State Mesonet Profiler Network Data"*** (amt-2022-85) submitted to *Atmospheric Measurement Techniques* by *Bhupal Shrestha et al.*

General recommendation:
Major revision

Synopsis:
The manuscript presents the observations by a network of microwave radiometer, Doppler lidar and all-sky camera in New York state. For three of the 17 stations, comparisons with nearby radiosondes have been performed and are analysed in terms of profiles of temperature, humidity and wind as well as for different thermodynamic indices that were derived from the original profiles. Furthermore, the aerosol optical depth from the camera observations is evaluated using AERONET sun photometer observations.

General comments:
The presentation of this quite dense network of ground-based remote sensing observations will be very useful to many readers if some more information is given on the quality of the dataset, especially regarding the microwave radiometer (MWR) dataset.

The Doppler lidar observations seem to agree well with the radiosondes. Sections 4.1 and 4.2 are well presented and can be published like this.

However, I am concerned for the large biases in the MWR observations compared to the radiosondes (especially for temperature). This cannot be only explained by retrieval uncertainties, especially the 8 K bias for clear sky cases. The reason for these large biases must be further investigated before the manuscript is ready for publication.

Response: The overall extent of observed cold biases in the MWR temperature are consistent with previous studies by Cimini et al., 2011, Xu et al., 2015 and Cimini et al., 2015. These studies have attributed such cold biases to varieties of reasons such as neural network performance, calibration uncertainty, gas absorption model biases, radiosonde biases and radiosonde drift distances, though latter two might have less significant effect. The larger cold biases during cloudy days than precipitation days are consistent with the results by Cimini et al. (2011) whereas the larger cold biases during clear sky days than the cloudy days are consistent with the results by Xu et al. (2015).

Due to the near identical error profiles from three selected NYSM Profiler sites, those observed biases could be sensor specific, rather than poor neural network performance. The issue of cold biases has not been fully understood yet, and it is still under investigation by our team in consultation with the manufacturer, Radiometrics. However, we speculate that significant biases inherent to the brightness temperature could be a major factor for large temperature biases (For further details, please refer to responses to the first detailed question) and the resulting height dependent biases could be the result of rapidly fading weighting function. The Radiometrics has developed automatic calibration (Acal) technique that replaces liquid nitrogen calibration and its associated uncertainty. Our preliminary analysis of the application of Acal at Albany has shown quite an improvement in the retrievals and is therefore, under extensive investigation at multiple NYSM sites. We are also evaluating one-dimensional variation (1-DVAR) technique (Cimini eta al., 2011) to improve the MWR retrievals. While those techniques are under evaluation and consideration for long-term use, the radiosonde correction method discussed in the manuscript is shown as a temporary solution to minimize such observed biases.

It is well known from literature that the quality of MWR temperature profiles decreases significantly with height. Therefore, it might be useful to limit the evaluation of the MWR profiles to the lower troposphere (e.g. 3 km), as for wind profiles.

Response: We agree that it might be useful to limit the evaluation of the MWR profiles within 3 km similar to that of DL profiles. However, based on our data requests throughout the profile, we think that our data users would like to see MWR retrievals and their associated errors along the profile and hence, presented our analysis to the maximum range of 10 km. Please find the further comments at the end.

Some more detailed questions and comments to the retrievals and their application:

• How were the neural network retrievals trained that you use for the MWR profiles? You are presenting the retrievals in lines 205ff., but I am wondering whether a retrieval self-test (by applying to forward-modelled radiosonde profiles) would result in similar biases. Do you have any information about that? If the self-test produces also the observed biases then there is an inherent problem with the retrieval (e.g. the setup of the radiative transfer model, the gas absorption models used, or the neural network design). On the other hand, if the self-test is bias-free then the reasons for the bias must stem from the instrument (e.g. wrong channel bandpasses, faulty calibration, etc.). However, the second hypothesis is much less likely, because in this case, the performance would not show such similar bias/error patterns at the different locations.

Response: The neural networks are trained using the historical radiosonde data from a site that is close to MWR site in terms of climatology and elevations. The three selected NYSM Profiler site MWR are very close to NWS radiosonde site and hence, those MWRs at Buffalo, Albany and Stony Brook are trained using NWS radiosonde data at Buffalo, Albany, and Upton.

We did the self-test for one month of the data at Albany and found out significant cold biases in V-band brightness temperature that is consistent with Ware et al., 2013, which could be due to instrumental and/or calibration issues. It is also reported that absorption models have some biases in each V-band channel (Hewison 2007). Similarly, significant V-band brightness temperature biases has also been reported immediately following liquid nitrogen calibration (Lohnert et al., 2012, Illingworth et al., 2019). Therefore, we agree that there is an inherent problem with the retrieval, and we are actively investigating it. These potential issues are mentioned in the paragraph of line 401-422.

As mentioned above, our preliminary analysis of Acal technique developed by Radiometrics has shown improvements in the MWR temperature retrievals. Therefore, we are extensively evaluating Acal technique along with 1-DVAR technique to reduce the MWR retrievals biases for the operational use, mentioned in the summary and conclusions line 689-693.

• Do you apply separate retrievals for 90° and 20° elevation observations? Or do you have any combined retrieval which uses brightness temperatures at both angles as input? How do the zenith and off-zenith observations compare e.g. during clear sky? I suppose that there is a stronger bias in the off-zenith observations as the optical path through the atmosphere is much longer and there is less contribution to the signal from the higher levels due to attenuation (esp. for temperature profiles).

Response: There are two separate retrievals for zenith (90°) and two off-zenith (20° facing North and South) observations. Since extensive study was carried out by Xu et al., 2014 showing the off-zenith retrievals better than zenith retrievals not only during precipitation but also during non-precipitation days, we followed their analysis and only used average of two off-zenith retrievals in this study. Based on our limited study between off-zenith and zenith retrievals, we agree with Xu et al., 2014 that off-zenith retrievals are better in precipitation days while those retrievals are somewhat comparable to zenith retrievals during non-precipitation days, though the off-zenith retrievals have reduced lower V-band weighting functions as compared to zenith retrievals (the zenith weighting function of 2km is reduced to ~0.8 km for off-zenith).

• The problem due to drifting radiosondes is rather minor and should not affect the retrieval bias (only the RMSE) as the sondes might randomly drift to warmer/colder/drier/moister regions. (see also p.19, l. 405-406).

Response: We totally agree that problem due to drifting radiosondes is minor and significant contribution in the retrieval biases could be due to error inherent to the MWR, discussed above. The sentence is reworded accordingly in line 417.

• Concerning the calibration of the instruments: How often were the radiometers calibrated using liquid nitrogen? Do you see jumps in the performance before and after calibrations?

Response: As recommended by the manufacturer Radiometrics, liquid nitrogen calibration is done once every 6 months. We do see some differences in the performances before and after the calibrations within 2-5%. More importantly, due to some biases observed immediately after liquid nitrogen calibration as also reported by Lohnert et al., 2012, the Radiometrics developed automatic calibration (Acal) that will eventually replace liquid nitrogen calibration is under evaluation by the NYSM team for operational use. As mentioned earlier, our preliminary results showed improvement in retrievals using Acal.

• I am a bit concerned by the bias correction. I am sure that on average you are improving the profiles, but how do you know that e.g. in severe weather situations (which are often at the edge of the statistical distribution), the bias correction performs well?

Response: The radiosonde-based correction method significantly improves the profiles (precisely temperature) for all three different types of weather conditions. The differences in the performances of corrected and uncorrected MWR retrievals and corresponding indices in the event of severe thunderstorm are demonstrated in the case study in Section 4.7 and Fig. 13, and the corrected data clearly outperforms the uncorrected data. Overall, the corrected data helps in meeting the threshold criteria set by the NWS. Further evaluation and statistical analysis of MWR retrievals (both corrected and uncorrected) in nowcasting severe weather is a part of ongoing study.

It is also to be noted that radiosonde correction method discussed in the manuscript is a temporary solution and will be replaced in near future by Acal and/or one dimensional variational (1-DVAR) technique for long-term use that has shown significant improvement in MWR data (Cimini et al., 2011). Since our studies have shown limited clear sky radiosonde data can help to reduce biases in MWR retrievals, we believe that this information provides other researchers/users with a method to replicate this simple linear correction method until a better method is available and implemented.

• For the bias correction, you are separating your dataset in clear sky/cloudy/rainy cases (Fig. 10a,b,c). What is the variability of these biases as a function of height for each of the three weather classes? Are there any seasonal differences? What about very cold winter days or hot summer conditions? I am pretty sure that the bias is quite variable, and the mean bias of 0°C / 0 gm-3 in Fig. 11 (a) and (d) does not show the range or the uncertainty of possible biases. Also, I would like to know more about the physical reason of the different biases depending on the weather.

Response: We have updated our error profile plots with error bars representing one standard deviation (Fig. 3, 6, 8 and 10) to show the variability of the biases as a function of height. We have observed seasonal differences in MWR data that are relatively larger during winter than summer, but those seasonal differences were not as pronounced as those seen during different weather conditions. Therefore, we decided to present the classification of the MWR data based on three weather conditions only. Since the radiosonde correction method is a temporary solution (and not to make plots crowded), we did not include error bars in Fig. 11 as compared to other previous plots. We believe it is quite evident with differences in MBE, MAE and RMSE before and after correction that radiosonde-based correction method can be a useful application.

We are not quite clear on why there are different biases depending on three different weather conditions. However, we speculate that better accuracies of temperature during precipitation and cloudy days than clear sky days could be due to the temperature profiles trending towards moist adiabat and reduced temperature inversion layers. As shown in Fig. 7, the MWR temperature profiles fail to detect elevated temperature inversion layers giving rise to a marked increase in cold biases that are more prominent on clear sky days. This physical reasoning is added in line 458-460. The better accuracies of water vapor density during clear sky days could be due to lower variability of the moisture than cloudy/precipitation days.

• Usually, a bias correction is done before applying the retrieval, i.e. a correction of the brightness temperatures (as it is done e.g. in Lohnert and Maier, 2012), but I guess that this is beyond the scope of this paper.

Response: Yes, this is beyond the scope of this paper.

Detailed comments:

• p.1, l.20-21: Why do you only mention biases here?

Response: Not any specific reason. Just do not want to overwhelm reader with too many error numbers in the abstract.

• p.3, l.81: The RMSE of <= 7 K is a very high number. Other publications show quite low numbers.

Response: The RSME up to 7 K is the maximum value observed along the profile reported in those publications cited, though few have reported much lower RMSE than 7 K. Since the performance of the MWR depends upon weather conditions as well as seasons, those publications show marked variations in errors, which could be due to different selected case study period.

• p.6, l.140-141: The vertical resolution is determined by the retrieval algorithm. The “true” vertical resolution is very coarse and decreasing with height (see e.g. Crewell and Lohnert, 2007). For these passive observations, often the degrees of freedom are determined, which is between 2 and 5 for the whole profile (depending on the choice of frequencies and angles).

Response: We agree that the vertical resolution is determined by the retrieval algorithm, which depends on inverse method and atmospheric conditions and degrades with height, as also reported by Hewison, 2007 and Cimini et al., 2015. To be more precise, we have changed “measurement” to “the retrieved profile” in line 142.

• p.6, table 1: Are you sure that the manufacturer gives a relative humidity accuracy of 2%?

Response: Based on the Radiometrics manual, temperature accuracy is up to 2°C and relative humidity accuracy is up to 2% at the surface, that tend to decrease along the height. This note is added in the table for clarity.

• p.16, fig. 6 a-c: These scatterplots are not very useful, as the performance of the MWR retrievals depends strongly on the height above ground (and not on the temperature itself). You could try to make a color plot Temperature bias vs. height for bins of 1°C vs. 100m (or the vertical resolution of the retrieval)

Response: We agree that the performances of the MWR retrievals are height dependent. However, the sole purpose of showing these scatterplots is to demonstrate an overall one-to-one relationship for providing the instant shape of the relationship between MWR and RS data cumulatively from all possible height levels. For any information on the errors along the height, mean bias error (MBE) with one standard deviation error bars, mean absolute error (MAE) and root mean squared error (RMSE) are presented alongside for specific height dependent error information.

• p.19, l.396-397: Yes, I can see that the biases are significant, but you should further investigate the reason for that.

Response: This is already discussed above. The changes are made accordingly in the paragraph (line 401-422). As already mentioned above, the exact reason for the significant biases is under investigation, but we believe it is most likely due to instrumental and calibration issues that inherently introduce biases in the brightness temperature.

• p. 19, l. 398ff.: *"This is likely because the MWR measured V-band observations are ingested into the neural network with greater weighting function at lower heights than at higher heights to produce a finer vertical resolution at lower heights."* This phrase makes no sense to me. The neural network itself will determine the weights of the single V-band brightness temperature observations for the temperature profile. This has nothing to do with the vertical resolution.

Response: The better accuracy of MWR retrieved temperature profiles in the lower altitude (within the boundary layer) than higher altitude is due to the weighting function of V-band channels, peaking near the surface and fading rapidly above the boundary layer (Hewison, 2007, Cimini et al., 2011, Westwater, 1993). Similarly, it could also be due to the fact that the temperature information is concentrated in the lowest few kilometers but drops off steadily with height (Hewison 2007). So, we agree that the sentence is not precisely correct and has been changed with the above information in line 405-407.

• p.19, l.401-402: *"In contrast, the water vapor density comparisons show better results at higher altitudes than at lower altitude"*

This is not surprising, as there is not much water vapor (absolute) in upper levels. If you compared the relative error, the result would look very different.

Response: We totally agree on that. Sentence is reworded in line 409-410.

• p.20, table 4: This table would be more interesting as a function of height (or limited to the lower troposphere, such as up to 3 km)

Response: Without a doubt, it is true that the accuracies of MWR retrievals decrease from the surface upward. It may be enough to keep the comparison results within 3 km. But, because we are also deriving forecast indices that require data from at least up to 500 mb or 5 – 6 km height, we decided to present data from all the heights. For any specific height dependent errors, vertical profiles of MBE, its statistical significance and one standard deviation error bars along with MAE and RMSE are presented, which provides readers with the height dependent errors, though we agree that overall comparison results would improve by considering height up to 3 km only. Moreover, being a leading profiler network in the nation, we thought that we should present analysis throughout the profiles for the readers and for those planning to deploy such networks in near future so that they are aware of advantages/disadvantages of deploying DL/MWR and the quality of data along the height from diverse geographical locations. Furthermore, the data requests from our users have shown interest in data along the profile and have frequently asked about the accuracy of data throughout the profile.