<u>AMT-2022-85-RC1</u>

amt-2022-85: Evaluation of the New York State Mesonet Profiler Network Data

Bhupal Shrestha, Jerald A. Brotzge and Junhong Wang

The paper examines data from 3 out of 17 NYSM Profiler Network stations and analyzes wind speed from the Doppler Lidar (DL), temperature and humidity profiles from the microware radiometer (MWR), and AOD from AODeSIR. Data from the DL and MWR are evaluated against radiosondes soundings and the usefulness for prediction of severe weather events evaluated. AOD at 2 sites is evaluated against data from the Aerosol Robotic Network.

The paper is interesting, and it presents a good overview of the Network data and capabilities. I do have some general major comments on the results and related discussion. All my comments are related to the MWR evaluation that in my opinion is the most problematic part.

Major general comments

The purpose of the paper as stated by the authors is *"to determine the robustness and accuracies of the instruments deployed with respect to well-defined measurements"* in view of their extensive use to *"complement radiosondes and satellite systems"*. The purpose of the paper is therefore very relevant considering the need for such measurements.

With this perspective in mind, I feel that the conclusions that: "*Overall, the MWR is a robust and reliable tool for the continuous measurement of atmospheric data and derived forecasting parameters*" <u>is not supported by the data presented.</u> The MWR data in the paper unfortunately present a bleak outlook for those of us hoping to use microwave radiometers for temperature and humidity monitoring. In fact, the results seem to suggest that, unless radiosondes are available at the site (to help with corrections or debiasing) the RMS errors shown in this paper (Figs. 6 and 10) ranging between 2K and 8K are well beyond what is required for any purpose.

Response: We agree that the "robust and reliable" statement is not an accurate representation of the results. So, the above stated sentence is reworded to "Overall, the MWR provides continuous and real-time measurement of atmospheric data and can be a valuable nowcasting tool for high-impact weather events despite cold biases in the temperature data" in the revised manuscript (line 653-656).

Therefore, the most important questions the authors should address in my opinion are the following:

1. Are these results due to poor calibration of the instruments and to poor retrievals? Are these issues resolvable?

Response: The 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 neural network, calibration uncertainty, gas absorption model, radiosonde biases and radiosonde drift distances. However, the near identical error profiles from three selected NYSM Profiler sites suggest that those errors are not likely due to calibration issues or poor neural network performance, as those errors are sensor specific. The cold bias error has not been fully understood yet, and it is still under investigation. Though overall off-zenith (20 degrees elevation) retrievals are better than zenith retrievals (Xu et al., 2013), off-zenith retrievals have reduced MWR lower v-band weighting functions; for example, the zenith weighting function of 2 km is reduced to off-zenith weighting function of ~0.8 km. The observed vertical profile errors are consistent with such rapidly fading weighting function above ~1 km for off-zenith retrievals.

These issues are resolvable with the application of the one-dimensional variational (1-DVAR) technique (Cimini et al., 2011). While that retrieval technique is under evaluation and consideration for long-term use, the radiosonde correction method discussed in the manuscript will be implemented as a temporary solution. Since our studies have shown that limited clear sky days radiosonde data can help to reduce biases in MWR retrievals, the NYSM is planning to launch several radiosondes to bias correct remaining 14 MWRs and to further investigate the performance of the MWR during the daytime (7am – 7pm LT) apart from the twice-daily NWS launches. This information is added in line 670-672.

2. What are the accuracy requirements we need to strive for in a ground-based network and how far are the Mesonet radiometers from these requirements? For example, in Table 1 of Wulfmeyer et al. (DOI: 10.1002/2014RG000476) a bias < 0.5 K with noise error of < 1K are reported as desirable for temperature profiling. I understand that microwave radiometers can't achieve that, but what are the accuracy requirements for the Mesonet network?

Response: The accuracy requirements are set by our users of the data and vary with application. Based on WMO OSCAR (https://space.oscar.wmo.int/requirements), the "ideal" and "optimal" requirements for temperatures and humidity in PBL and free troposphere are 0.5K and 1 K, and 2% and 5% respectively, for high resolution numerical weather prediction applications. But, considering the brightness temperature accuracies and the ill-posed retrieval problem, the expected accuracies for MWR temperature and water vapor density are $0.5 - 2$ K and $0.2 - 1.5$ g m^{-3} that further decrease with height (Illingworth et al., 2015). The NYSM aims to achieve similar accuracies of temperature and water vapor density, particularly within the boundary layer by minimizing any significant biases. Therefore, we presented the linear correction technique using the radiosonde data to bias correct the MWR retrievals as demonstrated in the manuscript.

Specifically, the NYSM is developed primarily with an aim to monitor and provide situational awareness and warning of severe high-impact weather events. Despite the limitations of MWR retrievals, the vertical profiles of temperature and water vapor density have shown similar trends to that from the radiosondes, particularly within the boundary layer. Thus, the high temporal resolution of vertical profiles provided by the MWR are helpful to understand the evolution of weather events and provide much needed real-time information to monitor and nowcast severe high-impact events as presented in the manuscript.

3. What are the implications of this analysis for the network itself? It seems to me that with those uncertainties only radiometers co-located with radiosondes can be reasonably used. How about the remaining 14 radiometers for which there are no radiosondes available?

This question brings the following consideration: what is the expected uncertainty in the forecast capability at a site where there are no radiosondes for correction? To this end, in my opinion, the paper would be more informative if the radiosondes were used only for evaluation purpose and the analysis was carried out entirely without the correction part. Table 5 and the whole case study (i.e. sections 4.6 and 4.7) should contain the results from the uncorrected MWR profiles. This would give us an idea of what can be reasonably expected from a profiler at a site without radiosondes.

Response: As a short-term plan, the NYSM will launch a several-day series of radiosondes near each profiler site to bias correct the MWR retrievals. Bias corrections will be developed for the remaining 14 Profiler sites to yield the accuracies discussed above in (2). Our data has shown that bias corrections are stable with time, and so a limited set of radiosondes collected at each site during clear sky days are sufficient, until a 1-DVAR technique or other new retrieval method is implemented in near future.

A height-dependent correction method applied to the MWR retrievals based on nearby radiosondes is presented in the manuscript with sole aim of demonstrating that a simple correction can be applied with just limited radiosonde data.

Since the correction method primarily reduces systematic biases (but not much random biases, Fig. 11), the results presented in Table 5 and sections 4.6 and 4.7 are similar before and after the correction, but with major differences in mean bias error (MBE) only. So, the MBE for original data is added alongside those from corrected data in Table 5.

Major specific comments:

1. Throughout the paper the MWR retrievals are called "measurements". Please change this.
Response: Done.

Some specific examples are reported later.

2. **Section 4.4** This section is a little bit puzzling, and I am not sure I understand it. I understand that, because the radiometer is pointing at 20-degree elevation, you may have a better chance of having less measurement degradation during rain. However, why should the radiometer do so much better during cloudy conditions than during clear sky? Or during rain than during cloudy? I assume that, if the effect of rain deposition is eliminated, the radiometer shouldn't distinguish between rainy and cloudy. To me the results could be easily explained by compensating measurement biases during precipitation, I am not sure this section should be kept at all as it is difficult to interpret.

Response: The observed larger cold biases during cloudy days than precipitation days are found to be 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. It is not quite clear why there are such differences in the performance of the MWR during three different weather conditions, although we can speculate that better accuracies during precipitation and cloudy days than clear sky days could be due to the temperature profiles trending towards the moist adiabat and reduced temperature inversion layers. As demonstrated in the manuscript, the MWR consistently fails to detect elevated temperature inversion layers giving rise to a marked increase in cold biases in the profiles, which are more prominent on clear sky days. Due to the varying performance of the MWR under different weather conditions, there is also a possibility of shifting in the weighting function which is now under investigation.

3. **Section 4.4** How do you know what are the conditions off-zenith? The IRT is looking at zenith so how do you know whether off-zenith, where the radiometer is pointing, is cloudy or clear? Perhaps there are clouds over the radiometer and not off zenith, or the other way around. In my opinion this entire section 4.4 should be eliminated.

Response: Yes, that is correct. The three different weather condition classifications are based on the IRT looking at zenith. Since the average of two off-zenith retrievals pointing in opposite directions is used in the study, the classification based on three weather conditions are carried out with an assumption that observations from the zenith pointing IRT still hold true for off-zenith retrievals, though this assumption will lead to some limitations and uncertainties (note added in line 422 – 426). Nevertheless, the observed but varying cold biases in MWR temperature profiles during different weather conditions are found to be consistent with previous studies by Cimini et al., 2011 and Xu et al., 2015.

4. **Section 4.5** This section should be eliminated. Of course, it comes without saying, that if we do have radiosondes at the site and we correct the radiometer based on the radiosondes we get good results, we already know this. But this defeats the entire purpose of having a network of radiometers. As mentioned in my general comments, I think the manuscript would be much more valuable if the authors used the radiosondes only for evaluation and not for correction. With the correction in place the conclusion of the analysis should be that the MWR are a good "*complement*" to radiosondes and should be deployed at radiosondes sites to increase the temporal coverage but have no value at sites where there are no radiosondes.

Response: The correction of MWR retrievals based on radiosonde data is a temporary solution to minimize errors and meet our accuracy specifications discussed above. Since the linear correction for the MWR can be developed from a rather limited sample size of radiosonde data, the MWR need not be co-located near an operational NWS radiosonde site. Indeed, the fact that cold bias remains relatively steady over time, and which can be fixed with a single linear fit is surprising. We believe that this information provides other researcher with a method to replicate this simple linear correction method until a better retrieval method is available and implemented. Furthermore, our aim of presenting this correction method aligns with another reviewer who believes that it is the main novelty of this study, one can be implemented easily by other researchers.

Despite the limitations of the MWR, the MWR retrievals have been found to follow similar vertical trend to that of radiosonde profiles, particularly within the boundary layer, where degrees of freedom are much greater. The MWR profiles provide much needed real-time evolution of thermodynamic properties diurnally, that is unavailable from the twice per day radiosonde data. Therefore, the MWR has significant value at sites where there are no radiosondes (and even at sites with radiosondes) and can be a critical nowcasting tool as demonstrated by our case study and other studies such as Madhulatha et al., 2013 and Chan et al., 2011.

Minor comments:
1. Page 6 L138 Please change vertical resolution to vertical grid.

Response: Done (line 140)

2. Page 9 Line 209 Please change MWR measurements with MWR retrievals.

Response: Done (line 205)

3. Page 9 Line 225 and following. It is important to understand that MWR measurements have no *vertically resolved information* above roughly 2 km. It will be enough to keep the MWR comparison between 0 and 3 km as well.

Response: It is true that the accuracies of MWR retrievals decrease from the surface upward as our comparison results of the MWR retrievals with the radiosondes are better at lower heights than at higher heights. However, there are still some useful information above 2 – 3 km that are needed for deriving forecasting parameters as demonstrated in the manuscript (in our case at least up to 500 mb or 5 – 6 km). Despite limited vertically resolved information above 2 km, studies (Cimini et al., 2015, Chan et al., 2011, Madhulatha et al, 2013) have shown the value of MWR retrievals in deriving valuable thermodynamic indices that require data above 2 km.

4. Page 10 line 230: "directly measured" please change to *retrieved* as you are comparing retrievals of temperature and humidity from the MWR with those measured by radiosondes.

Response: Done (line 229)

5. Page 15 section 4.3: The high biases and RMSE of the MWR retrievals are obviously concerning and are probably due to poor calibration of the instruments combined to an inadequacy of the neural network retrievals. Both aspects can be improved operationally to lower the RMSE to less than 1.5 K between 0 and 3 km. Is the mesonet network planning to do that?

Response: Yes, the NYSM is working with the sensor vendor on improving sensor calibration and neural network retrievals. In meantime, the NYSM is using the bias correction technique as discussed in the manuscript until the application of an improved calibration and 1-DVAR technique are available. The RS based correction helps to reduce both temperature MAE and

RMSE below 2 °C within 2 km (Fig. 11). Therefore, we plan to launch limited radiosondes at the remaining 14 sites.

More importantly, our consistent results from across 3 different sites (MWR retrievals compared against two different radiosondes systems) indicate that the observed errors in the MWR retrievals are not primarily due to poor calibration or neural network performance but rather due to some inherent issues in the MWR with an ill-posed retrieval problem and rapidly fading weighting function. As mentioned before, we are exploring other retrieval algorithms, such as 1-DVAR technique.

6. Line 363-406. The discussion of the vapor density profiles would be more informative if the author could provide the range of vapor density at the sites. This would give the reader an idea of the error percentage on the profiles (for example an RMSE of 1 g/m_3 would be about 10% if the average vapor density at the site is 10 g/m_3).

Response: Across three selected sites, the mean water vapor density values are found to be 7.8 – 9.3 $g\ m^{-3}$ at the surface, 5.5 – 6.6 $g\ m^{-3}$ at 1 km and 3.9 – 4.6 $g\ m^{-3}$ at 2 km giving rise to error percentage of 9 – 14 % at surface, 20 – 25 % at 1 km and 26 – 33% at 2 km, with overall error of 17 – 23 % below 2 km. This information is added in line 378 – 381.

Conclusions

In conclusion, I think the information presented in the paper is valuable, however it would be more valuable if the authors could provide the reader with a true assessment of the value of the entire MWR network for forecast purpose. To this end the authors should not be afraid of presenting less than perfect results if that is what the MWR network is providing. Such information would be extremely valuable for those planning to deploy such networks where radiosondes are not available.

In order to evaluate the true capability of a network of radiometers for forecasting purposes the authors should use radiosondes only for evaluation and not for correction. After the uncertainties in the forecast parameters have been established (with the help of radiosondes) data from sites without radiosondes should be used to forecast events and to establish realistic uncertainties. If the results are not satisfactory the authors should discuss how the poor radiometric performance can be improved without help from radiosondes, i.e. by reviewing the calibration procedures and improving the retrievals.

We are evaluating the MWRs to determine their value for providing real-time data for weather operations and warning on high-impact weather. Specifically, we are evaluating the ability of the MWRs to provide enhanced situational awareness, a critical tool for improving short-term warnings and nowcasts. Due to the limitations of the MWR, we are in the process of understanding the issues of the MWR, which appear to be more than simply the calibration and neural network retrieval. We hope to resolve these issues with the application of better retrieval method such as 1-DVAR in near future.

Review of Evaluation of the New York State Mesonet Profiler Network Data

By Shrestha, Brotzge, and Wang (AMT-2022-85)

General:

This study presents an evaluation of part of the New York State Mesonet, with several stations providing microwave radiometer profiles of thermodynamic quantities and profiles of winds from a Doppler lidar. Multiple years of data are considered, and three sites located near radiosonde launch sites are evaluated by comparison to those soundings. Errors in thermodynamic quantities are large (and differ between clear-sky, cloudy, and precipitation conditions), and to resolve these errors a linear regression method is developed and applied to the MWR data. A brief case study is presented, that of a thunderstorm that would not have been characterized with the radiosonde network. Finally, some profile-related thermodynamic parameters are derived from the MWR and evaluated, but an opportunity was missed to evaluate important stability-related metrics like the Richardson number that would be available from conventional radiosonde datasets.

A manuscript like this, presenting a new, publicly available dataset, could be of interest to the AMT readership. The main novelty is the correction applied to the MWR, but insufficient detail on that correction is provided for a reader to understand the steps involved or how to implement a similar correction on a different dataset. I would recommend that the authors present this correction in more detail, and that the section on derived parameters be expanded to include the Richardson number and perhaps the boundary-layer height as well. The figures should be improved as discussed in my comments below.

Major:

1. The section on the "Correction to MWR biases" requires more detail. As lines 464-6 are written, it is not clear if the correction is derived separately for cloudy, clear, and precipitation days. Not enough information is given for a reader to attempt to replicate this correction. Are the profiles in Figure 11 taken from the 25% of the dataset used for testing, or from the 75% training? (I would presume the "testing" portion, but it is never explicitly stated.) Is this correction something that could be applied to other datasets in other locations? If so, what steps should a researcher take?

   Response: The correction is derived separately for precipitation, cloudy and clear sky days for both temperature and water vapor density. Profiles in Fig. 11 are only from the testing dataset (25% of total dataset for each weather conditions). The paragraph (line 475 – 486) is revised for better clarity.

   Our results have shown that limited clear sky days radiosonde data are found to be helpful to reduce biases in MWR retrievals. Therefore, the NYSM is planning to launch

several radiosondes to bias correct remaining 14 MWRs in summer of 2022. As a long-term plan, one dimensional variational (1-DVAR) technique (Cimini et al., 2011) is under consideration and will be implemented in near future.

2. Some profile-related thermodynamic parameters are derived from the MWR profiles and briefly evaluated in section 4.6. Because both winds and thermodynamics are available from the RS, why not combine the MWR and the DL datasets to calculate stability metrics like the bulk Richardson number (Ri)? That would be a very useful test of the utility of the network. An important opportunity was missed here. The Ri could have been incorporated into the discussion of the thunderstorm case study as well. Similarly, the planetary boundary layer (PBL) height is easily calculated from the DL dataset, and that could have been compared to sounding-based estimates as well. Or, if the radiosonde dataset is not adequate for calculating the PBL height, pointing out the utility of the new mesonet capabilities could be a nice addition to the paper.

Response: The single value bulk Richardson number used in convective storm forecasting that depends on CAPE and deep layer wind shear, 0 – 6 km AGL, (Weisman and Klemp, 1986; Evenson, 1993, https://glossary.ametsoc.org/wiki/Bulk_richardson_number) is not possible to derive due to lack of DL wind measurements at 6 km.

It is possible to derive vertical profile of Ri (Sorenson et al. 1998) and PBL height; however, as the RS launch times (7 am and 7 pm LT) are not optimal times for the DL data availability, those two parameters are not included in this study. However, the authors plan to derive and compare those two parameters when we launch our own radiosondes (during the afternoon) to bias correct the remaining 14 MWRs. The PBL height using Ri and other DL methodologies will be extensively discussed in our next paper. The utility of the NYSM Profiler Network to derive those parameters is briefly mentioned in this manuscript (line 530-531)

Minor:

1. Line 133: can you comment on how often the lidar can actually retrieve wind estimates from 7000m? We see later that almost no data is available above 3km, so that should be noted here.

Response: The DL data availability is very limited and rarely available above 3 km. There are occasional data availability above 3 km when long range transported wildfire smoke is detected in the region during the summer months. This information is added in line 133-135.

2. Figure 2 would be more intuitive with height on the y-axis. The colors are difficult to distinguish (especially for red-green colorblind readers) so please consider using different line styles.

Response: Done. Fig. 2 is revised with different colors and height on the y-axis.

3. Lines 200-205: how sensitive is the agreement between the radiosondes and the doppler lidar to the averaging time selected?

Response: Using ±10 min (20 minutes averaging) and ±30 min (60 minutes averaging) centered at the radiosonde launch time produced slightly better DL/MWR comparison results (R^2 increased by 1-2%) than using only +10 minutes and +30 minutes starting from the launch time.

4. Near line 210: please explicitly state which RS sites were used to train the neural nets for each of the MWR sites

Response: The RS sites used to train the neural network for three selected MWR sites are listed in line 210-212. The list of NWS RS sites used to train neural networks for all 17 MWR sites are as follows:

NYSM Site	NWS RS #1	NWS RS #2
Albany	Albany	
Belleville	Albany	Buffalo
Bronx	Upton	
Buffalo	Buffalo	
Chazy	Albany	
Clymer	Buffalo	Pittsburgh
East Hampton	Upton	
Jordan	Albany	Buffalo
Owego	Albany	Munich (Germany)
Queens	Upton	
Red Hook	Albany	
Staten Island	Upton	
Stony Brook	Upton	
Suffern	Albany	
Tupper Lake	Albany	Munich (Germany)
Wantagh	Upton	

5. Line 252: better to provide a textbook reference than a link that disappear over time.

Response: Paper references added in addition to web link (line 249-251).

6. Fig 3, 6: again, please don't rely on red-green differences. Use colors that are more easily distinguishable or also use line style differences

Response: All those figures are revised accordingly.

7. Why doesn't Fig 4 also include Stony Brook?

Response: DL data from Stony Brook was not available for same date and time as for other two sites. Stony Brook DL data from different date is added in Fig. 4.

8. Line 347: explicitly point out the elevated inversion layer near 1 km in Fig7a, near 2.5km in Fig7b, etc.

   Response: The heights corresponding to the elevated inversion layers are added in line 347-348)

9. Line 459-460: write these statements out explicitly instead of using the confusing parenthetical formulation.

   Response: Done (line 469-470)

10. Figure 12 relies on the rainbow color table although extensive literature is available showing that it is suboptimal (Light and Bartlein, 2004; Stoelzle and Stein, 2021)

    Response: Thank you very much for the suggestion! We have made extra efforts to test different color schemes and replaced rainbow color map with plasma as suggested by Stoelzle and Stein, 2021.

References

Light, A. and Bartlein, P. J.: The end of the rainbow? Color schemes for improved data graphics, Eos Trans. AGU, 85, 385–391, https://doi.org/10.1029/2004EO400002, 2004.

Stoelzle, M. and Stein, L.: Rainbow color map distorts and misleads research in hydrology – guidance for better visualizations and science communication, Hydrology and Earth Systems Sciences, 25, 4549–4565, https://doi.org/10.5194/hess-25-4549-2021, 2021.