Dr. Yugo Kanaya

Editor

Atmospheric Chemistry and Physics

June 1, 2021

Dear Dr. Kanaya,

Subject: Revision of manuscript #acp-2020-1193

Thank you and the two reviewers again for reviewing our manuscript. We appreciate your thoughtful comments and suggestions. We have carefully reviewed the comments and revised the manuscript accordingly. Our responses are given in a point-by-point manner below. We have also attached a version of the manuscript and supplement with tracked changes and hope the revised manuscript is suitable for publication.

Please address all correspondence concerning this manuscript to Dr. Yuhang Wang (yuhang.wang@eas.gatech.edu). Thanks again for your time.

Sincerely,

Jianfeng Li

Atmospheric Sciences and Global Change Division

Pacific Northwest National Laboratory

Richland, Washington, US, 99354

Response to Editor

Thank you for your careful and thorough reading of this manuscript and your thoughtful comments and suggestions. Our responses follow your comments (in *Italics*).

Comments to the Author:

Dear Authors,

The two reviewers found that the manuscript has been improved based on the original comments and the authors' additional analyses/revisions. However, one reviewer still suggests minor revision is necessary before publication. I would appreciate it if the authors could further consider the comments from Reviewer #2. Thank you very much.

Reply:

Thank you for your comments. We have revised the manuscript based on the suggestions from the two reviewers. Besides, we have made some other slight changes to the language. Please see Lines 522 and 747 in the revised main manuscript and Lines 117 – 119 in the revised supplemental figure file.

Response to Reviewer #1

Thank you for your careful and thorough reading of this manuscript and your thoughtful comments and suggestions. Our responses follow your comments (in *Italics*).

The authors have addressed most of my concerns/suggestions. I agree with the authors that night-time/early-morning PBLH is one of the difficulties the modelling community needs to study/solve. The mismatch found between ground-based remote sensing instruments and modelling results is interesting and important for air-quality researchers. I cannot say I agree with all the findings and results in the manuscripts, but I think the authors have provided enough descriptions, thoughts, and reasoning for readers and researchers. I would like to support the publication of this good-quality work on ACP.

Only a few minor technical corrections are left.

Reply:

Thank you for your positive comments. After addressing the reviewer's suggestions, we have also updated two references from Lamsal et al. (2021) and Judd et al. (2020). Please see Lines 171, 174, 503, 981 – 982, and 1017 – 1020. We have deleted the R^2 values for wind direction in Line 321 because wind direction is a periodic variable which is not suited for computing a linear correlation. Figure S5 shows that our WRF simulation predicts wind direction well.

Technical corrections:

Figs. 10 and 13 (revised manuscript). I expect the shading areas (e.g., red areas on Fig. 10) are also the 1 std of the data (same as the error bars). But, I would suggest the authors make this clear. Also, I did not see any reason why the KNMI-GOME2 results in the revised manuscript became "better" (in terms of agreement with other models/observations). I would expect just some outliers been removed. But, if any criteria have been changed, one should provide such details at least in the response file.

Reply:

Thank you for your suggestions. Yes, the shading areas in Figures 10 and 13 show the standard deviations of 36-km REAM simulation results. We clarified the captions of Figs. 10 and 13. Please see Lines 1274 – 1276 and Lines 1296 – 1297 in the revised main manuscript.

There are two reasons for the "better" KNMI GOME-2A results. One is the removal of some outliers, as the reviewer mentioned. The other one is the inclusion of negative NO_2 TVCDs. In the first version of the manuscript, we only selected positive NO_2 TVCDs. However, according to the user guide of NASA $OMNO_2$ (v4.0) (https://aura.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level2/OMNO2.003/doc/READ

ME.OMNO2.pdf), we should consider all valid measurements, regardless of the sign, to avoid biases. Negative NO_2 TVCDs are produced when DOAS-derived total NO_2 slant column densities (SCDs) are lower than stratospheric NO_2 SCDs. Stratospheric NO_2 SCDs can be derived from model results or measurements from unpolluted areas. We want to emphasize that we used the same criteria (Lines 179 – 180 in the revised main manuscript) for all satellite NO_2 TVCDs, including KNMI and NASA products and our retrievals with 36-km REAM simulated NO_2 vertical profiles, to keep sampling consistent among these results.

Fig. S23 (revised manuscript). Please consider modifying the subpanels to use the same y-range (e.g., 0-24). I think the authors want to show the results clearly for each site, but the scale of the difference (between model and observations) might be more important in the current stage.

Reply:

Thanks. We now use the same vertical scale in Figure S23. Please see Lines 174 – 175 in the revised supplemental figure file.

References:

Judd, L. M., Al-Saadi, J. A., Szykman, J. J., Valin, L. C., Janz, S. J., Kowalewski, M. G., Eskes, H. J., Veefkind, J. P., Cede, A., Mueller, M., Gebetsberger, M., Swap, R., Pierce, R. B., Nowlan, C. R., Abad, G. G., Nehrir, A., and Williams, D.: Evaluating Sentinel-5P TROPOMI tropospheric NO_2 column densities with airborne and Pandora spectrometers near New York City and Long Island Sound, Atmos. Meas. Tech., 13, 6113-6140, https://doi.org/10.5194/amt-13-6113-2020, 2020.

Lamsal, L. N., Krotkov, N. A., Vasilkov, A., Marchenko, S., Qin, W., Yang, E. S., Fasnacht, Z., Joiner, J., Choi, S., Haffner, D., Swartz, W. H., Fisher, B., and Bucsela, E.: Ozone Monitoring Instrument (OMI) Aura nitrogen dioxide standard product version 4.0 with improved surface and cloud treatments, Atmos. Meas. Tech., 14, 455-479, https://doi.org/10.5194/amt-14-455-2021, 2021.

Response to Reviewer #2

Thank you for your careful and thorough reading of this manuscript and your thoughtful comments and suggestions. Our responses follow your comments (in *Italics*).

The manuscript was majorly revised and improved. Detailed analysis in the manuscript and supporting information would be valuable for the science community in this field. However, I would like to recommend authors to rewords abstract and main text to reflect what is in the manuscript considering uncertainties in the data and analysis. Main concerns are addressed below.

Reply:

Thank you for your comments and suggestions. We have added more discussion of the uncertainties/limitations of the ELF mixed-layer height and ACAM NO_2 VCD data. Details are as follows.

I would like to ask the authors to inform uncertainties in mixing heights retrieved from lidar during nighttime (or stable condition). The mixing height determined by aerosol backscatter may not be a direct indicator for stable boundary layer height, but the residual layer of aerosols. Compton et al. (2013) evaluated the lidar mixing height observed only during daytime. Because ELF data are not a reliable stable boundary layer height, the updates in this manuscript should be regarded as a sensitivity test, not a correction. In this regard, abstract and main text need to be reworded carefully. More information on the location and characteristics (number of data and uncertainty) of ELF observations would be helpful. Note that nocturnal boundary layer height from original YSU scheme is within the range of boundary layer height measured and modeled (Steeneveld et al., 2007; Koracin and Berkowics, 1988; Nieuwstadt and Tennekes, 1981). In the abstract, "However, nighttime mixing in the model needs to be enhanced to reproduce the observed NO2 diurnal cycle in the model". Based on the comments above, I think this should be rephrased. Furthermore, uncertainties in the NOx emissions in nighttime were not well estimated or understood. This part needs to be mentioned.

Reply:

Thank you for your comments and suggestions. We have added more explanations and discussion about the ELF data, substituted PBLH with mixed-layer height (MLH), and revised the abstract and main text accordingly. Please see Lines 38 – 40, 356 – 358, 360 – 385, 387, 390 – 394, 423, 446, 456, 595, 765, 996 – 998, 1237 – 1243.

Compton et al. (2013) indeed only evaluated the ELF mixing height with other measurements in the daytime, and the covariance wavelet transform (CWT) method is also designed for daytime (sunrise to sunset) but not nighttime. However, the residue layer (RL) issue is considered in the algorithm. When a RL is encountered, mixing height will be generally searched for below the RL. The limitation is that they set RL = 1 km constantly in the algorithm. According to Compton et al. (2013), one major problem during nighttime is the insufficient vertical resolution of the CWT technique as nighttime mixing heights are much lower than daytime.

We agree with you that there are potential larger uncertainties for the ELF data at night than in the daytime. However, as explained in the revised main manuscript in Lines 373 – 378, Figure 6 still shows underestimated WRF-YSU MLHs compared to ELF observations in the early morning after sunrise and the late afternoon before sunset. They are daytime measurements and have been evaluated by using Radar wind profiler observations and Sigma Space mini-micropulse lidar data. Moreover, the nighttime MLHs in Figure 6 are comparable to those measured by the Vaisala CL51 ceilometer at the Chemistry And Physics of the Atmospheric Boundary Layer Experiment (CAPABLE) site in Hampton, Virginia (Knepp et al., 2017). We now use the term, MLH, following Knepp et al. (2017).

UMBC is an urban site surrounded by a mixture of constructed materials and vegetation. Maybe urbanization-associated surface roughness change, anthropogenic heat release, and heat storage are potential causes of the high nighttime MLHs. We have reworded the sentence in Lines 38 – 40 in the abstract to avoid overemphasizing the importance of k_{zz} adjustment.

In the previous submission, we mentioned the potential uncertainties of nighttime NO_x emissions in Lines 390 – 394 in the revised main manuscript, and our previous sensitivity tests in Figure R1 indicate that nighttime NO_x emissions need to be reduced by at least 50% to match REAM results with NO_2 observations within uncertainties, and the reduction needs to be at least 67% to match O_3 concentrations between REAM and observations. Without additional robust evidence, such significant reductions are unreasonable, as we stated in Line 394 in the revised main manuscript. We have rewritten the sentences so that they are more apparent. Please see Lines 390 – 394 in the revised main manuscript.

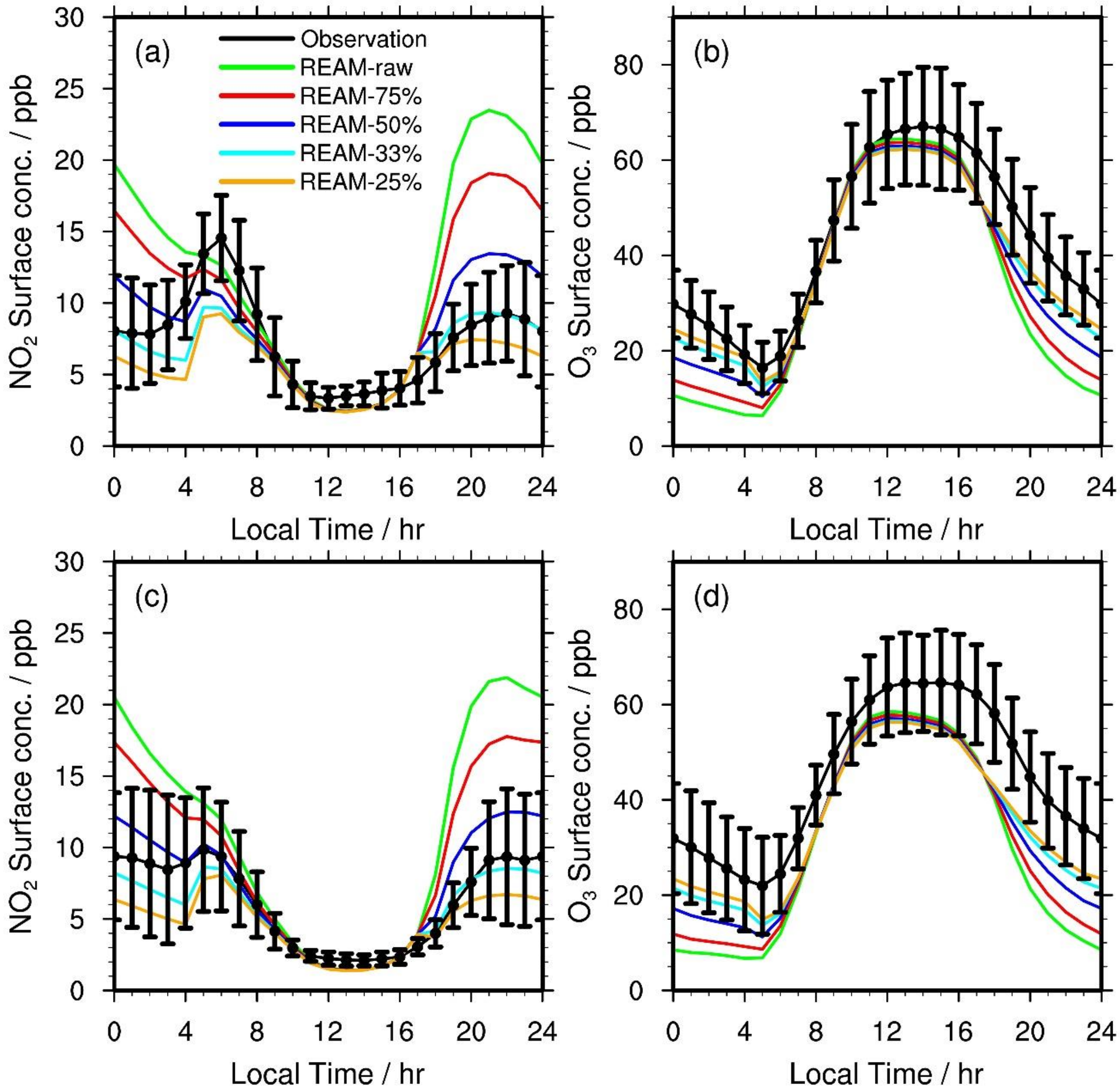

Figure R1. Diurnal cycles of surface NO_2 (a, c) and O_3 (b, d) concentrations on weekdays (a, b) and weekends (c, d) during the DISCOVER-AQ campaign in the DISCOVER-AQ region. "REAM-raw" (green lines) denotes the REAM simulation with NEI2011 emissions; "REAM-75%" denotes the REAM simulation with NO_x emissions from 18:00 – 5:00 LT reduced by 25%; "REAM-50%" denotes the REAM simulation with NO_x emissions from 18:00 – 5:00 LT reduced by 50%; "REAM-33%" denotes the REAM simulation with NOx emissions from 18:00 – 5:00 LT reduced by 67%; "REAM-25%" denotes the REAM simulation with NO_x emissions from 18:00 – 5:00 LT reduced by 75%. Black lines denote the observations during the campaign, and black vertical bars denote corresponding standard deviations.

Nighttime biases in NO2 and NOy at the surface were much reduced when the updated YSU Kzz was used. However, column NO2 concentration in Figure 10 during nighttime are much larger than those from PANDORA. It is possible that the PANDORA *observations are underestimated in the morning and late afternoon, but it is also possible that the model columns are overestimated due to the emission uncertainty and this problem could not be fixed with the updated Kzz. Figure S2 exhibits dynamic changes in PANDORA NO2 columns during daytime and the model well reproduced these changes except late afternoon to nighttime (Figure S23). In the abstract, "Another discrepancy is that Pandora measured NO2 TVCDs show much less variation in the late afternoon than simulated in the model". Can it be the case that the model columns vary too much?*

Reply:

Thank you for your suggestions. We think it is necessary to clarify that when we talked about the discrepancies between REAM simulation results and Pandora, we didn't mean that Pandora was wrong. What we did was to describe the results and explain why they are different. As illustrated in our responses in the first round of the review process and Lines 523 – 540 in the revised main manuscript, Pandora measured different columns of air at different times of the day and is sensitive to local conditions. It is highly possible that Pandora measurements cannot represent the average conditions of a REAM grid cell, considering the potential significant spatial heterogeneity of NO_x, especially in the early morning and late afternoon when NO_x lifetime is long, and NO_x accumulation is much more apparent than mid-day. We can say that Pandora has limitations in evaluating current model results; however, it doesn't mean that Pandora is wrong, or REAM simulations are wrong because the measurements and model results are not directly comparable. The performance of REAM simulations in the afternoon was validated using P-3B aircraft derived NO_2 VCDs, as shown in Figure 10 in Line 1267 in the revised main manuscript. We have reworded the sentence in Lines 517 – 518 in the revised main manuscript to avoid potential misunderstanding that we meant Pandora was wrong.

In Figure 14, the authors indicated that the purple circle denote a small region surrounded by high-NOx emission pixels and with high NO2 VCDs in the 4-km REAM but low NO2 VCDs in ACAM. But Figure S25 shows enhanced columns in the purple circle region in ACAM (purple circle not shown in the figure). The purple area is on the edge of land and is filtered out in Figure S24 and S26 (potentially due to clouds). Is it possible that undersampling issues in this area highlight the differences in spatial distributions in Figure 14? R square values in Figure S24-S26 are quite reasonable, considering that it is comparisons at fine-resolution. It is not convincing that the results in the manuscript suggest spatial allocation problems in NEI as written in the abstract.

Reply:

Thank you for your thoughtful and valuable suggestions. We indeed didn't consider the sampling issue for the ACAM dataset. We have updated Figures 14, S24 – S27 and their captions (Figure S25 is new). Please see Lines 1299 – 1308 in the revised main manuscript and Lines 179 – 200 in the revised supplemental figure file. Figure numbering has been updated accordingly. The updated figures have included the distributions of the relative differences of NO_2 VCDs between REAM and ACAM and the number of data points used to calculate grid cell mean NO_2 VCDs. The purple area in the old Figure 14 indeed has limited samplings (≤ 3); therefore, we have deleted the purple circle and relevant discussion about it. Please see Lines 687 – 690, 1299 – 1300 (Figure 14), and 1303 – 1304 (Figure 14 caption) in the revised main manuscript. However, these modifications do not affect our analysis results or conclusions, as we identified NO_x emissions and gradients as a major factor causing the discrepancies among the 36-km REAM, the 4-km REAM, and observations through comprehensive evaluations and diagnostics of NO_x related chemistry and physics (sections 3.6) and excluded the horizontal transport effect in Lines 682 – 687. The purple circle discussion is a detail that is not essential for the analysis.

The new Figure S25 shows the results only for those grid cells with samplings ≥ 10. As expected, we can still find that the 4-km REAM has more concentrated NO_2 VCDs in Baltimore and Washington, D.C. urban regions but less concentrated in rural areas than ACAM in Figure S25. The following figure R2 shows the results without scaling NO_2 VCDs by corresponding domain averages. It considers the real differences of NO_2 VCDs between the 4-km REAM and ACAM. It is clear that the 4-km REAM generally has larger NO_2 VCDs in Baltimore and Washington, D.C. urban regions but lower NO_2 VCDs in rural areas than ACAM. Please see Lines 693 – 698 and 703 – 704 in the revised main manuscript.

R^2 or Pearson correlation coefficient is widely used in atmospheric science studies and our study. However, R^2 has limitations. The criteria of a good R^2 is arbitrary, depending on the purpose. Since NEI can identify urban regions using population densities, capturing the essential urban-rural contrast in NO_x emissions, that R^2 value should be high. But that's not enough for our spatial distribution discussion, which is to say that R^2 does not describe how NO_x emission decrease from urban to rural regions in the model as compared to the observations. For example, in Figure S27d in Line 199 in the revised supplemental figure file, we have a regression line $REAM = 2.11\ ACAM - 1.11$ with $R^2 = 0.61$. If we only focus on R^2, the regression is good. But this regression line generally shows larger REAM NO_2 VCDs than ACAM on high NO_2 VCD grid cells but lower REAM NO_2 VCDs than ACAM on low NO_2 VCD grid cells, as shown in Figure S27d and S27e. Even if Figure S27d shows a perfect regression line $REAM = 2.11\ ACAM - 1.11$ with $R^2 = 1$, the systematic biases between high NO_2 VCD grid cells and low NO_2 VCD grid cells persist. It is an obvious spatial distribution problem. In this situation, a perfect model-observation correlation requires $y = x$ with $R^2 = 1$. Therefore, we have added the distributions of the relative differences of scaled or unscaled NO_2 VCDs between REAM and ACAM in the updated figures to show the distribution issue directly. Besides comparing NO_2 VCDs between 4-km REAM and ACAM, we also showed other evidence for the NEI NO_x emission distribution issue, including site-comparison results and the comparison of NO_2 VCDs between the 36-km REAM and satellite products (section 3.7).

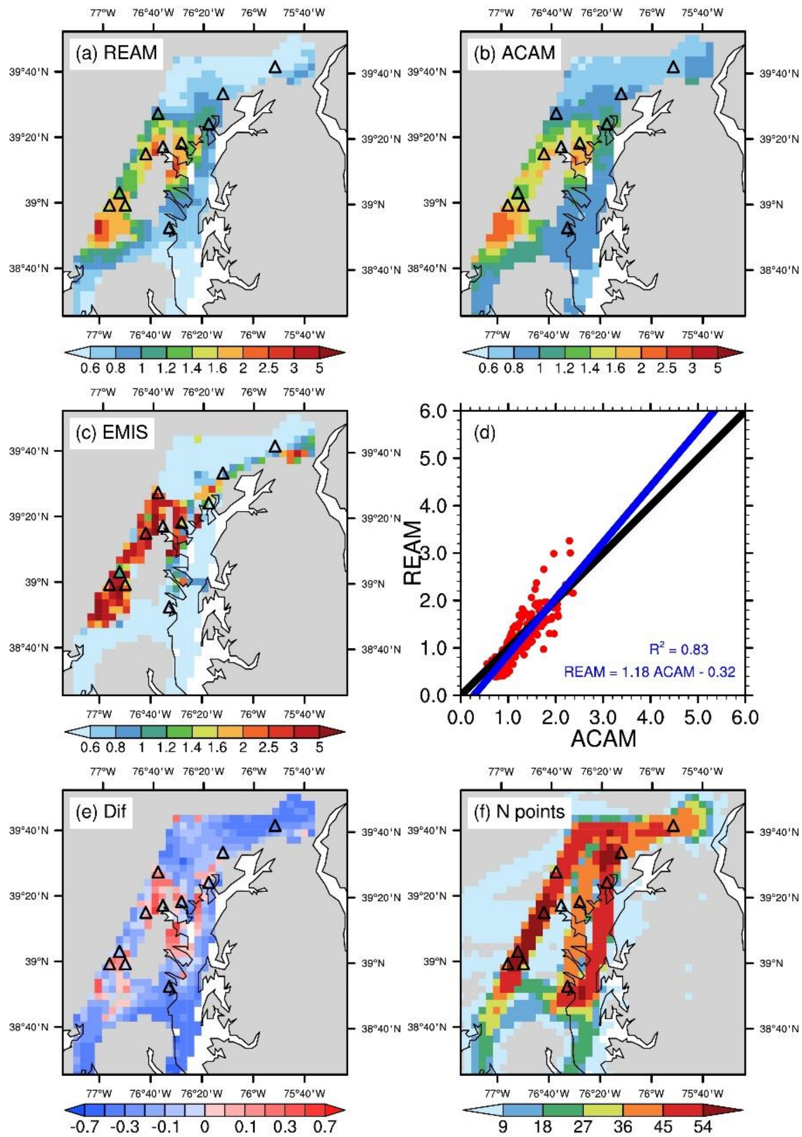

Figure R2. Same as Figure S25 in the revised supplemental figure file but without scaling NO_2 VCDs by corresponding domain averages in (a) and (b). The NO_2 VCD unit in (a) and (b) is 5×10^{15} molecules cm^{-2}.

References:

Compton, J. C., Delgado, R., Berkoff, T. A., and Hoff, R. M.: Determination of planetary boundary layer height on short spatial and temporal scales: A demonstration of the covariance wavelet transform in ground-based wind profiler and lidar measurements, Journal of Atmospheric and Oceanic Technology, 30, 1566-1575, https://doi.org/10.1175/JTECH-D-12-00116.1, 2013.

Steeneveld et al. (2007), Diagnostic equations for the stable boundary layer height: evaluation and dimensional analysis, Journal of Applied Meteorology and Climatology, 46, 212-225.

Koracin and Berkowics (1988), Nocturnal boundary-layer height: observations by acoustic sounders and predictions in terms of surface-layer parameters, Boundary-Layer Meteorology, 43, 65-83.

Nieuwstadt and Tennekes (1981), A rate equation for the nocturnal boundary-layer height, Journal of the Atmospheric Sciences, 38, 1418-1428.

References:

Compton, J. C., Delgado, R., Berkoff, T. A., and Hoff, R. M.: Determination of planetary boundary layer height on short spatial and temporal scales: A demonstration of the covariance wavelet transform in ground-based wind profiler and lidar measurements, Journal of Atmospheric and Oceanic Technology, 30, 1566-1575, https://doi.org/10.1175/JTECH-D-12-00116.1, 2013.

Knepp, T. N., Szykman, J. J., Long, R., Duvall, R. M., Krug, J., Beaver, M., Cavender, K., Kronmiller, K., Wheeler, M., and Delgado, R.: Assessment of mixed-layer height estimation from single-wavelength ceilometer profiles, Atmos. Meas. Tech., 10, 3963-3983, https://doi.org/10.5194/amt-10-3963-2017, 2017.