Point-by-point Responses to Referee #2 for # amt-2020-302

Authors used regular fonts for Referee #2 comments and used blue fonts for author's response.

Anonymous Referee #2
This manuscript is topical and informative and should be useful to the micromet community. It is certainly appropriate for AMT. My specific comments and recommendation follow.

1 Overall the paper is clear enough, but I still think the use of English could be improved.

Response: Thank you. Yes, we have made improvements in our revision.

2. Lines 133-134 - The authors state "Implausible values of 20 Hz data, defined as greater than 30 g m−3 or less than 2 g m−3 , were removed . . . ." This range of values poses a bit of a puzzle to me. Why/how were these max/min values chosen and why are they implausible? I think the authors should include a histogram of the noise spikes. They really need to say more about their criteria for noise/spike removal. I will also note that mentioning "Welch's periodogram method" and citing Blanken et al. (2003) (Lines 139-140) does not really address or answer my concern here. Is it possible to show a Welch periodogram and discuss the details relevant to how these "implausible" values were determined?

Response: We appreciate you for making this point. It would be worth noting that we used our own Matlab codes to process data and conducted all data analysis including spectral analysis. We did use Eddy-Pro software in this study as well and used to double-check our flux estimates. In fact, water vapor fluxes calculated from both data processing tools were nearly the same. The 2 gH_2O m^{-3} (equals to 111 mmol H_2O m^{-3} or 2.7 mmol mol^{-1}) is the lowest water vapor density during the growing season at Bushland, Texas. The 30 gH_2O m^{-3} is equivalent to 1,666 mmol m^{-3} or 42 mmol H_2O mol^{-1} water vapor density as an upper bound, which covered any possible highest water vapor density readings in Bushland, Texas. In Eddy-Pro software, the de-spiking thresholds for both water vapor and CO_2 are +/- 3.5 standard deviations of a moving window (usually a 5-minute window or 1/6 of flux averaging period with half window overlapped).

We revised the sentence in our revision as:
"The data de-spiking process set all data beyond the upper (30 g m^{-3}) and lower (2 g m^{-3}) values as missing. Both upper and lower bounds were estimated by all possible water vapor density observations during the growing seasons in Bushland, Texas."

Regarding Welch's periodogram, it is a method for calculating the power spectral density and co-spectral density in Fourier transform computations. For example, Blanken et al. (2003) used this method for estimating the power spectral density and cospectral density in their 20 Hz time series. This method, per our understanding, is not associated with the upper and lower bounds of water vapor density.

3 Lines 156-157 - The authors state "Based on spectral losses and other corrections, E was calculated iteratively." This statement needs some clarification. What other corrections are involved and why does E need to be calculated iteratively? It would be helpful to show the equations and explain the need for the iterative approach.

Response: Thank you for your insight. Our intent here is to briefly describe the standard flux computation procedures and corrections. We agree that this sentence was not well written and we deleted this sentence to avoid possible confusion.

There are many papers and textbooks that describe iterative approach equations and other standard corrections used in eddy covariance methods (e.g., an excellent software manual by Mauder and Foken, 2004). The basic rationale for having iteration approaches is because the sonic anemometer is directly measuring sonic virtual temperature (T_s, $w'T_s'$) rather than absolute thermal temperature (T_{air}, for $w'T_{air}'$).

<u>4 Lines 165-166</u> - Here the authors state "The measured λE is assumed to be the difference between the actual flux and these errors (Lasslop et al. 2008)." This statement also needs some clarification. I do not understand the point of the referring to Lasslop et al. (2008). What exactly does Lasslop et al. (2008) show that is relevant to the authors' study in general and this specific statement in particular? What is the significance of or the need for the Lasslop et al. (2008) C2 AMTD Interactive comment Printer-friendly version Discussion paper reference. Do Lasslop et al. (2008) state something, either explicitly or implicitly, that is relevant to manuscript that could be restated for clarity?

Response: Many thanks for your comments and constructive questions. Lasslop's paper addressed random errors and systematic errors in the eddy covariance system, in which the random errors were estimated by using the gapfilling algorithm (Reichstein et al. 2005). Our objective in this study is to evaluate three generations of IRGAs by inter-comparison, spectral analysis, and direct comparison against an absolute reference – the world-class weighing lysimeter in Bushland, Texas. We also evaluated the systematic errors based on Mauder et al. (2013) and random errors where the estimates were from Finkelstein and Sims (2001). Therefore, we deleted the citation of Lasslop et al. (2008) which is an inaccurate citation in our original manuscript.

<u>5 Lines 166-173</u> - The definition and discussion of the systematic error must have at least one unstated assumption, i.e., that there are no comparable errors in the heat flux. While this may be true for many eddy covariance systems I don't think one can assume, a priori, that it is universally the case. Could I not define a systematic error (say δH) associated with the heat that mimicked Equation (2), i.e., $\delta H = H(1/ERB - 1)$? If so, what exactly does this mean to the value and utility of using Equation (2) to define the systematic error associated with λE?

Response: This is an excellent point. We agree that our study has to assume that there are no comparable errors in the sensible heat flux. Per our understanding, this is a legitimate assumption. We used two IRGAs to share one cast3 anemometer so that $\delta H = H(1/ERB - 1)$ for the two IRGAs are the same. The second csat3 we used also shared identical homogenous footprints within a well-managed crop field. We tried to examine LE's systematic errors and random errors as our secondary objective in this paper because our main objective was to address intercomparison, spectral analysis, and direct comparison against the weighing lysimeter. We used Eq. (2) to evaluate systematic errors because (1) it can be used to examine the difference between two IRGAs due to insufficient sampling of large-scale air motion; and (2) the EBR in Eq. (2) exactly reflects the energy balance closure problem on a daily basis.

<u>6 Lines 215-216</u> - Here the authors state "After this time, the LI-7500RS appeared to be more stable, with steady rmsd over the final days compared to the other two instruments." This statement also needs some clarification. Because they define rmsd with Equation (1), but this does not seem consistent with their statement. The problem is that they claim that one sensor is

more stable than the others, but the rmsd is defined as the difference between two sensors. So how can they claim that the rmsd is a property solely of one instrument?

Response: We admit that the rmsd definition by Eq. (1) was not clear for readers in our original manuscript. To clarify, we slightly changed the $x_{RS,\ i}$ into $x_{REF,\ i}$ in Eq. (1) and reworded the sentence as below:

"$rmsd = \sqrt{\sum(x_{A,i} - x_{REF,i})^2}$, (1)

where $x_{A,i}$ is the i^{th} observation for the LI-7500/A and $x_{REF,i}$ is the i^{th} observation for the reference LI-7500RS. Interinstrument variability was also determined by *rmsd* except using the average value of three IRGAs or three EC systems as a reference value."

In Figure 3, the rmsd was determined by a reference from the average of three IRGAs water vapor density.

7 Lines 390-392 - Here the authors state "While it was paired with a different sonic anemometer than the other two IRGAs, flux differences were attributed to differences in variance of turbulent fluctuations of water vapor rather than sonic anemometer error." At the very least this statement is out of place. It should included in **2.2 Data processing and statistical analysis or 3.3 Water vapor fluxes** or maybe a separate section devoted to discussing the influence that uncertainties in the other Non-IRGA instruments might have on the present IRGA results. My concern is that there have been at least half a dozen papers in the last 8 years (starting with Kochendorfer et al.: 2012, Boundary-Layer Meteorology, **145**, 383-398 to the most recent Frank et al.: 2020, Boundary-Layer Meteorology, 175, 203-235) about sonic transducer shadowing errors causing systematic C3 AMTD Interactive comment Printer-friendly version Discussion paper underestimation of w 0 . (Note: the other recent sonic papers will be referenced in Frank et al. 2020.) So that means the some errors in the water vapor flux that are ascribed solely to the IRGA are in fact caused by the sonic itself. Just how much of an impact does this assumption make on the results of this study? In addition, if w 0 is biased low, the heat flux, H, will also suffer from this bias. So what impact does this have on the ERB, Equation (3), and the systematic error δ, defined in Equation (2) and ascribed solely to the λE? How certain are the authors that δ is not dominated by the bias in the sonic vertical velocity rather than errors inherent in the IRGAs? I think the paper would be strengthened if the authors performed a sensitivity or error analysis to estimate how much of δ is related to non-IRGA errors and how much of δ can reasonably be ascribed to an IRGA.

Response: Thank you for these insightful comments. We deleted the "While it was …" statement because it was out of place. In sections 2.2 or 3.3 we had similar statements.

We agree that the sonic anemometer's w0 underestimates (vertical component) have been (re)examined in many papers. In 2012 and 2013, two co-authors in this paper intensively discussed shadow effects with some of the authors that you mentioned. We also agree with your insight in terms of sonic uncertainties. However, such uncertainties as well as non-IRGA errors are not the objective for this paper. Our purpose was to address water vapor density measurements and corresponding flux estimates (i.e., latent heat flux) from three generations of IRGAs.

Recommendation

The paper is acceptably written, but the writing could be improved. I don't think that the statistical analysis is well described. Furthermore, I think the paper approaches this instrument performance problem in a manner that is a bit naive and simplistic. They use the energy balance ratio and its closure as a measure of hygrometer performance. But the measurements of Rn, G and J are not free of systematic error or bias. Nor is the sonic necessarily free of bias. How then can they be certain that just because the LI-7500 produces a better closure that it performs better that the other two generations of the instrument? Additionally, they do not discuss possible biases and errors in the lysimeter measurement of ET. I think all sources of errors and uncertainties need to be at least acknowledged in their study. And I think the paper would be further improved it the authors tried to quantify or partition δ into IRGA and Non-IRGA contributions. Finally, although I would not require a Bayesian statistical approach to their instrument comparison study, I think their efforts and analyses would benefit greatly from such an approach. A Bayesian analysis would allow the authors to build in estimates of the uncertainties associated with the lysimeter and the energy balance instruments.

Response: Thank you for your nice review and insightful comments which substantially improved our paper's quality. Our main objective was to address three generations of infrared analyzers with respect to water vapor density and water vapor flux by using intercomparison, spectral/co-spectral analysis, and direct comparison with the weighing lysimeter. The statistical method we used for systematic errors and random errors was a complementary method in our study. The sonic's uncertainties and non-IRGA errors are beyond for the scope of this paper. It would be our goal to further investigate these uncertainties in the near future including Bayesian analysis.

--- The END of point-by-point response for referee #2