We thank the reviewers for their constructive and helpful comments. As a result of these suggestions, the revised manuscript now has clearer methodology and interpretation of results and conclusions, thereby enhancing the reproducibility and application of the results. Responses to individual comments are provided below each comment in blue text.

Review of manuscript "Characterizing Thermodynamic Observations from Unshielded Multirotor Drone Sensors" by S. W. Freeman

In the manuscript "Characterizing Thermodynamic Observations from Unshielded Multirotor Drone Sensors" by S. W. Freeman et al., the authors used observations from a tethersonde and from an instrumented flux tower to investigate the accuracy of temperature and humidity observations derived from instrumented UAS and how the errors varied as a function of sensor position on the UAS as well as a function of radiative regime. The following suggestions should be addressed.

We thank the reviewer for their detailed comments. In addressing these comments, our methodology is now clearer, thus enhancing the reproducibility of the results.

Specific Comments:

Line 88–90: In panels (a) and (d) of Figure 1, it is difficult to distinguish the UAS and sensor, respectively, from the background in the photograph. I recommend including a better photograph in both of these instances to enhance legibility.

Thank you for your comment. We have changed these photographs by removing or dimming the background to emphasize the UAS and sensors. We have also added two additional photographs of the UAS in the new Figure 1 panels (c) and (d), and have labeled the visible sensors, to show the sensor placement from additional perspectives. The new Figure 1 and its associated caption are repeated below:

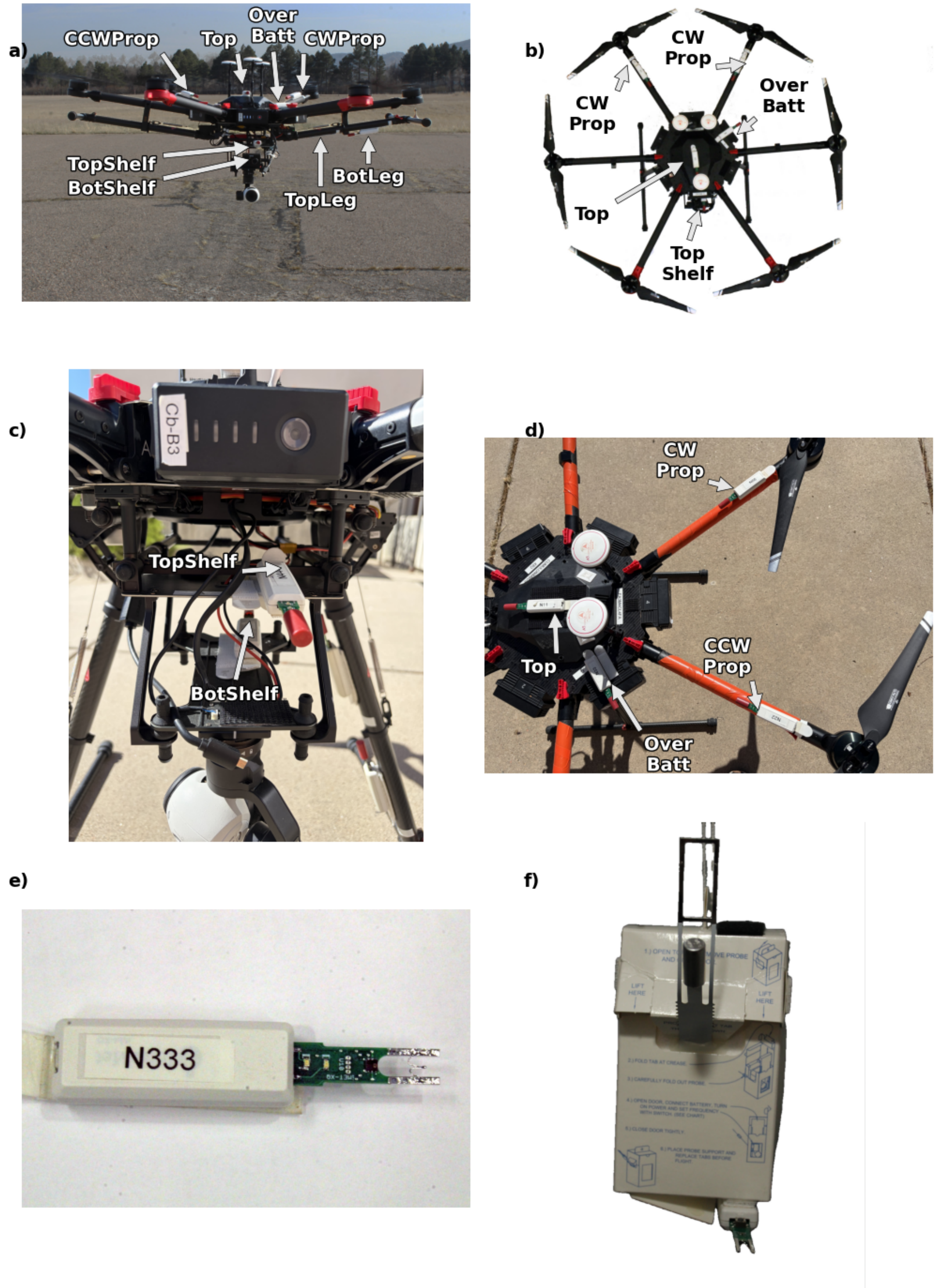

Figure 1: (a) Front-view picture of the DJI Matrice 600 Pro drone in flight with the legs folded up. The 8 sensor positions are labeled. (b) Top-down photo of the drone with the legs down; the position of all visible sensors are labeled. (c) Side profile of the drone, showing a close-up of the two shelf positions. (d) Close top-down profile of the drone, showing the four

positions on the drone's top. (e) Photograph of an iMet XQ sensor. (f) Photograph of the iMet-1 radiosonde with an iMet XQ sensor attached to the bottom.

Line 135–136: Vague; please quantify how this was determined.

We have clarified that this was done via real-time onboard GPS in lines 148-149:

"After takeoff, the drone was flown directly upwards until the altitudes of the tethersonde and the drone were within a few meters of each other, as determined by the onboard GPS of both platforms."

Line 139–140: Please quantify how similar the measurements were in this instance.

When examining only those flights when the tethersonde was located below the drone, the overall average temperature error increased by 0.06 K, which is well within the sensor accuracy of ± 0.3K, We have added this specification in lines 151-153:

"When comparing the data from only those flights when the drone was located higher in altitude than the tethersonde, the thermodynamic measurement errors are similar, with average temperature error across all sensors increasing by 0.06 K, which is well within the sensor accuracy."

Line 169–170: More details should be provided about where and how the iMet XQ sensor was installed on the tower.

We have added additional details about the exact location of the iMet XQ installed on the tower (lines 184-188):

"At this location, the temperature measurements were compared with both an unshielded, unaspirated iMet XQ sensor mounted to the flux tower (not shown), and the triple aspirated, one-minute-averaged air temperature (NEON 2024a) and humidity measurements (NEON 2024b) taken at the *Top* of the flux tower. The iMet XQ sensor was mounted to the railing at the top of the tower, near the existing NEON temperature measurements, with the sensor oriented parallel to the ground and pointed away from the tower such that airflow to the sensor was as unimpeded as possible."

Line 182–183: Justification is needed as to why the sensors were rotated daily rather than after each flight.

This was done for logistics reasons as unlike the previous generation, the new iMet XQ2 sensors are not easily movable when mounted to the drone. Given that the sensors were still rotated between days, such that no sensor was in the same position on each of the three days, we believe that a similar result was achieved, particularly as these flights were designed for validation purposes only. We have clarified our language (lines 200-204):

"However, due to logistical reasons, the sensors on the flux tower validation flights were not rotated between each flight as for the tethersonde flights, but rather on a daily basis. Despite the more limited changing cadence, no sensor on the drone was kept the same position between days."

Line 198: How were the wind gusts determined?

Wind gusts from both stations shown in Figure 2 were the average of the 1-minute maximum wind speed reported by the weather stations during each flight (i.e., for a 9-minute flight, the gust reported per flight is the average of the nine 1-minute maximum wind speeds). In both cases, the gusts are separately reported by the station and are the maximum wind speed observed during each station's sampling period. We have clarified this in the caption for Figure 2 (lines 215-219):

"Figure 2: Histograms of conditions observed by the weather stations during all flights from Christman Field (blue) and the NEON Flux Tower (purple; NEON 2024a,b,c,d,e). The conditions are averaged over the duration of each flight. Shown are surface (a) temperature, (b) mixing ratio, (c) pressure, (d) solar radiation, (e) sustained wind speed, and (f) wind gust, which is determined by taking the average station-reported wind gusts at the 1-minute sampling interval of both surface weather stations."

Line 211–212: More detailed are needed about the data filtering and data quality control here to clarify how the authors determined data that were considered to be invalid.

Only data that were unphysical were deemed to be invalid (0%, negative, or >100% relative humidity and temperatures >373 K). We have clarified lines 232-234 to now read:

"In addition to removing data points where the drone and tethersonde were too far apart, data points were also removed when any sensor recorded unphysical (e.g., negative relative humidity, negative or zero pressure) data."

Line 518–519: The drone data and tethersonde data should presently be ready for review rather than being added upon the article's acceptance.

We have uploaded these files to Zenodo, and these files are now available at DOI: 10.5281/zenodo.17497401.

Technical Corrections:

Line 73: Missing period.

Fixed.

Line 242: The symbol μ is missing from the parentheses.

Fixed.

Line 280: Extra space.

Fixed.

Line 442: Extra space.

Fixed

Line 460: “Nevertheless” is one word.

Fixed.

General Comments

The paper discusses test results from an experiment using a drone with thermodynamic sensors mounted in various locations. The drone is flown next to a tethersonde and an instrumented flux tower. The work builds on prior work of a similar nature which is discussed. The experiments are described, but some of the information should be consolidated in one place. Data quality control is performed and discussed. The relevance of the work is to correct for biases in field experiments using drone-based measurements, which is an important topic. Shortcomings are discussed and ideas for future work are presented. Overall the work is solid, the paper well-written, but some reorganization is recommended. Specifics are provided below.

We thank the reviewer for their detailed and insightful comments. In addressing these comments, we believe that our manuscript is now clearer, with enhanced reproducibility due to improved descriptions of the methodology. Furthermore, the interpretation and applicability of the results has also been improved as a result of addressing the reviewer's comments.

Specific Comments

On page 2 (lines 55 -59) the authors discuss absolute temperature measurements versus temporal changes and gradients. At this point the difference between the two is not clear, because as a drone flies temperature changes occur (due to e.g. clouds passing or turbulence) and gradients (spatial changes) in temperature and humidity exist and these are measured. In section 3.4 (pages 18 and 19), the authors clarify the difference. It appears that they are talking about absolute and relative measurements. Using those terms on page 2 instead of absolute and temporal changes, would probably make more sense to the reader. The same should happen on lines 298 and 299. Also, technically the term "gradient" is associated with a spatial change and since those are not explored in the paper, I would recommend using "temporal change" instead of "temporal gradient" throughout the document. E.g. lines 398, 477.

Thank you for raising this excellent point. We have clarified our language throughout the document to be clear that we are examining relative changes in thermodynamics over time. Below, we have highlighted several instances of where we have made such changes.

The term “temporal gradient” has also been changed everywhere to “temporal changes” as suggested by the reviewer.

Lines 57-59 now read: “Quantifying the properties of relative temporal changes in thermodynamic measurements is therefore necessary to enhance confidence in these drone-based measurements."

Lines 69-71 now read: “Further, by investigating these errors, we aim to provide field campaigns with a characterization of drone-based temperature and water vapor mixing ratio measurements, both for absolute measurements and the temporal changes in these measurements.”

Lines 323-324 now read: “Section 3d contains an analysis of the errors when measurements of temporal changes in temperature are sought instead of measurements of absolute temperature values.”

Lines 422-423 now read: “The 95% confidence error for temporal changes in temperature at night…”

On page 2, the sentence spanning lines 61 and 62, states that the “accuracy of thermodynamic measurements made from multiple sensor positions aboard a multirotor while in flight, during both day and night, has not yet been investigated”. I take it the authors mean that the “both day and night” item has not yet been investigated, because Kimball et al and others certainly investigated the other items mentioned in this sentence. Please clarify that specifically the night-time measurements are new and relevant, because they allow the investigation of the effects from solar radiation.

Yes, indeed. We have now revised this section to read (lines 62-65):

“However, separating the solar radiative effects on sensors on the accuracy of thermodynamic measurements from those effects specifically due to multirotors, has not yet been investigated. As such, observations from multiple sensor positions on board multirotor drones have, for the first time, been made during both day and night to address these impacts.

On page 4, lines 97 through 102, the authors discuss the justification for comparing water vapor mixing ratio instead of RH. Please clarify this statement: “…. RH depends not only on accurate temperature measurements but also on the assumption that the temperatures of the drone and the comparison sensor are the same”. If you use the temperature of the RH sensor, both for the drone and the tethersonde/flux tower, to convert RH to water vapor mixing ratio doesn't that provide an independent representative measure of water vapor on each platform? Please elaborate.

We agree that using the RH sensor temperature on all platforms to convert RH to water vapor mixing ratio provides an independent, representative measure of water vapor. We have clarified why we are using water vapor mixing ratio on all platforms, and have revised this sentence to read (lines 103-105):

"This is because RH, without conversion to a temperature-independent measure of water vapor, depends not only on accurate temperature measurements but also on the assumption that the temperatures of the drone and the comparison sensor are the same—an assumption we investigate in this study."

Close-up pictures of the various sensor locations on the drone would be helpful. The relative locations of the sensors are shown in Figure 1, but the specifics of how the sensors are mounted and where they are positioned relative to the support structures on the drone are unclear. E.g. TopShelf and BotShelf are on platforms mounted below the main electronics (lines 123 and 124). What do these platforms look like and where are they relative to the body of the drone? Is there any chance of shading by the drone parts?

We have made two changes to Figure 1, both in response to this reviewer's comment and to a comment from Reviewer 1. First, we have darkened the background of Figure 1a to help indicate the locations of the sensors relative to the drone. Second, we have added two additional photographs of the drone in the new Figure 1 panels (c) and (d), and have labeled the visible sensors, to show the sensor placement from more perspectives. Specifically, panel (c) shows the TopShelf and BotsShelf sensor mount locations. The new Figure 1 and its associated caption are repeated below. As for shading, we address this below.

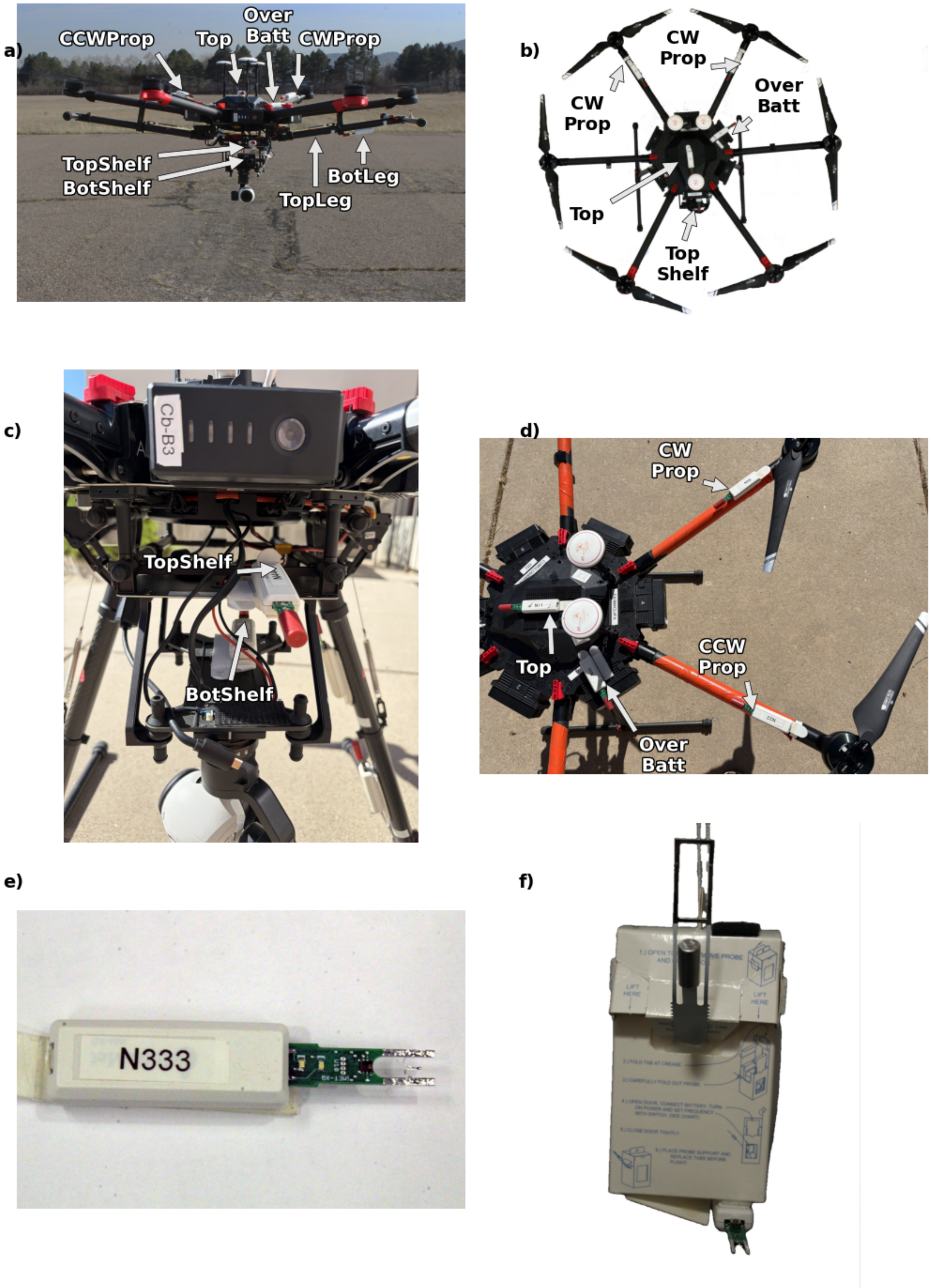

Figure 1: (a) Front-view picture of the DJI Matrice 600 Pro drone in flight with the legs folded up. The 8 sensor positions are labeled. (b) Top-down photo of the drone with the legs down; the position of all visible sensors are labeled. (c) Side profile of the drone, showing a close-up of the two shelf positions. (d) Close top-down profile of the drone, showing the four

positions on the drone's top. (e) Photograph of an iMet XQ sensor. (f) Photograph of the iMet-1 radiosonde with an iMet XQ sensor attached to the bottom.

In section 2.2 the tethersonde tests are described, but no dates are given. They are elsewhere in the manuscript, but the really should be included in this section as well. Please put all specifics (dates, location, sensor specifics, mounting platform details, etc.) about the experiments in this section for easy reference for your audience. On line 128 a surface weather station is referenced which had not been mentioned or described before., Please give some specifics about what this looks like, how tall is the tower/platform, what sensors are included, is it portable or permanent etc. The drone's position is mentioned in line 141, but specifics are not given i.e. distance from the base of the tethersonde or latitude/longitude, height above the surface. This information should also be included. Make sure the same information is included in section 2.3 where the flux tower experiments are described.

Thank you! This, and the subsequent suggestions in section 2 are very good suggestions. We have now included all specifics for each location and platform in Section 2 for easy reference.

The modified information contained in the first paragraph of section 2.2 is as follows (Lines 134-140):

"In the first part of this experiment, we flew the drone adjacent to a tethersonde during both daytime and nighttime hours during several operating periods in October and November 2018 and January 2019. All night flights were conducted after nautical twilight. The tethersonde was anchored in dead scrubland at the now-retired Christman Field Airport (hereafter Christman Field) in Fort Collins, CO (40.59077° N, 105.14389°W, 1572 m above sea level). The meteorological station at Christman Field was used to quantify the spread in local meteorological conditions for each of the flights, and to validate that the measured insolation during nighttime flights was 0.0 W m^{-2}. The Christman Field weather station is a permanent weather station that measures temperature at 2 m above the surface and winds 10 m above the surface."

Similarly, the first sentence of section 2.3 now reads (Lines 181-183):

"The second part of the experiment was to conduct additional flights in June and July 2022 at the National Ecological Observatory Network (NEON) North Sterling, CO site (Metzger et al. 2019; NEON 2024a,b,c,d,e; 40.461894° N, -103.02929° W, 1365 m MSL)."

In section 2.2 the authors state that “this mean altitude does not substantially affect the thermodynamic errors measured” (lines 138 and 139) without explaining why. On page 10 (section 2.5) they do explain this very well. It is recommended that the explanation in section 2.5 either be moved to section 2.2 or a statement be added to section 2.2 referring the reader to an explanation in section 2.5.

We have added additional information to this statement in Section 2.2 (lines 151-154):

“When comparing the data from only those flights when the drone was located higher in altitude than the tethersonde, the thermodynamic measurement errors are similar, with average temperature error across all sensors increasing by 0.06 K, which is well within the sensor accuracy.”

On lines 154 and 155, the authors state that “while the tethersonde may also have errors in the temperature measurements arising from solar radiation, these should be similar to those on the drone”. This is where the exact placement of the sensors comes into play. The issue of solar angle relative to the location of the sensors and the possibility of them being shaded by components from the drone or facing directly into the sun, is not addressed. This should be included in the analysis.

In all flights, the drone was oriented to face south, which can cause slight differences in shade versus other orientations. The new Figure 1, which demonstrates the positions of the sensors in more detail on the drone puts this into context, and we have also addressed this in the manuscript text (lines 129-132):

“In all flights, the drone was flown such that its “front” (the same direction that the TopShelf sensor points) was facing south. In all daytime flights, this resulted in the front-facing sensors being directed into the sun and other sensors potentially being shaded by the drone. This uncertainty is mitigated by the range in solar radiations of the flights, which are discussed below.”

On line 180/181 the authors mention that the flux tower measurements were used primarily to validate the tethersonde measurements, but do not explain how. Thhis has the reader wondering which specific sensors were validated, if the tethersonde was deployed next to the flux tower, and when these experiments took place. Again, this is explained later in the document, but really belongs in this section.

We’ve now clarified the purpose of the flux tower tests in this section (lines 198-200):

“Therefore, we use these measurements primarily to validate the tethersonde comparison experiments, ensuring that the tethersonde flights yield robust results when compared to flights conducted with gold-standard, forcefully aspirated temperature measurements.”

In Figure 3, the authors present box plots and frequency distributions which are very insightful. However, because the distributions are not normal, the statistics of mean and standard deviation are not appropriate, instead median and IQR (inter-quartile range) should be used to quantify the center and spread of the distributions. Please replace (or at the very least add) these in the figures and text. Also the sample size (N) is given for most datasets, but not for the flux tower data. This should be included somewhere on page 12 as well in possibly table 2 (add N for all the experiments shown). N is also not given for the nighttime measurements; on line 304 the authors say there were 1.5 total flight hours, but how many individual temperature observations were there? Please add.

Yes, this is a good point. We have now added median and interquartile range to all figures, tables, and the text where mean and standard deviation appear. We have, however, retained the mean and standard deviation given their broad use in the community.

We have added the sample sizes for the tethersonde, daytime, and nighttime flights in Section 2, however, we had to omit them from Table 2 for space considerations.

Lines 177-179:

“In total, there were N=38593 individual samples during daytime flights and N=4872 individual samples during nighttime flights.”

Lines 195-197:

“However, these measurements are only available at a lower, one-minute resolution compared with the 1 Hz resolution on the drone. This leads to a total sample size N=228, versus the substantially larger sample size from the tethersonde measurements (N=43,465).”

In lines 268 and 269, the authors state that “random instrument errors are typically symmetric around zero after being corrected for any mean bias”. I assume this is a general statement and not something pertaining just to your dataset? If that is the case, it makes more sense to mention this at the start of your results section so the reader is aware when reading the rest of this section. A good place would be page 10, line 217.

We have included this information where suggested, now lines 238-240:

"The instrument error is defined as any error caused by fluctuations in the instrument itself other than the calibration biases that have already been corrected, and are typically centered around 0 after mean biases are corrected."

For the temperature range given on line 296 it is unclear whether the -1.34 K refers to the lowest 2.5th percentile in Figure 3b-i or something else. Similarly, is 1.88 K the highest 97.5th percentile in Figure 3 b-i? Please clarify.

We have clarified this in the text (lines 319-321):

"The temperature results demonstrate for the worst-case scenario for any single temperature measurement made during the daytime that the accuracy at a 95% confidence interval (hereafter CI), from the smallest 2.5% to the largest 97.5% has a range of -1.34 K ≤ T ≤ +1.88 K when not removing the mean biases (Figure 3b-i)."

In lines 333 and 334 the colder means of the positions with the most aspiration (CWProp, CCWProp, and BotLeg) are used to demonstrate that forced aspiration can reduce the bias induced by the drone's heating. The means of these 3 location are 0.08, 0.13, and 0.05 respectively (see Figure 5). What about the low mean (0.07) of position Top? That is the second lowest value and disproves the above statement. Also, the mean of 0.13 (CCWProp) is one of the highest mean values, so this further negates the above argument. Please explain or address.

Yes, this is a good point. We agree with the reviewer and have therefore removed the statement from the manuscript.

In line 355 it states that the RH sensor measures the temperature that the RH is based on. Add that that was the temperature used to calculate the water vapor mixing ratio. This will strengthen your statement.

We have added this as suggested (lines 378-380):

“This is despite the fact that the RH sensor measures the temperature that the RH is based on and this temperature is used in the calculation of the water vapor mixing ratio.”

Is the temperature range given on line 431 for the TopShelf location (i.e. Figure 8h) or something else? Please clarify and add the TopShelf range if this is something else. Also at the end of the sentence on line 443 you could state that: “because the temperature difference errors for OverBat range between x and y”. Then provide the values for x and y from Fig8e.

We have clarified how the temperature range, which represents the TopShelf range, is calculated after Figure 8h (lines 455-457):

“The results presented here indicate that $\Delta_{60s}\overline{T}$ errors onboard the drone for the *TopShelf* position are at worst -1.08 K ≤ $\Delta_{60s}\overline{T}$ ≤ +1.12 K at the 95% confidence level, representing the largest possible range in error.”

We have also clarified the text on LAFE (lines 466-469):

“Our study indicates that the magnitudes of the variance in temperature measured by the drones in the LAFE campaign are likely to be real atmospheric features at the 90% CI, as half of the sensor siting locations (*BotLeg, CWProp, OverBatt, Top)* had a 90% CI for $\Delta_{60s}\overline{T}$ <±0.5 K.”

Technical Corrections

Thank you for catching these technical corrections. We have resolved all of them.

Caption of Figure 1: no hyphen needed in in-flight. Just use in flight.

On line 134 add the words “the drone” after maneuver.

Line 178: Add “resistance” after the word platinum.

On the last line of the caption of Figure 3, the symbol for ‘mu’ (mean error) is missing from the parentheses.

Line 295: Add ‘that’ after the word demonstrate.

Line 460: nevertheless is one word.

Line 497 onboard is two words: on board.