RAAVEN UAS data processing for supercell thunderstorm observations during TORUS and TORUS-LItE

Adam L. Houston, Mark De Bruin, Céu Gómez-Faulk, Brian M. Argrow, Robert Sasse

The authors are grateful for the thorough and constructive feedback from the reviewers.

Inline responses to reviewers are included below. Reviewer comments in blue, response from authors in black, and verbatim text from the revised manuscript in green.

General response

We have added more specificity to the title in response to a comment from Referee #3. We have also added another author (Robert Sasse) who was acknowledged in the previous version but whose contributions we felt warranted co-authorship.

Editor comments

I have noticed that your tables contain coloured cells. Please note that this will not be possible in the final revised version of the paper due to HTML conversion of the paper. When revising the final version, you can use footnotes or italic/bold font.

Cell shading has been removed and replaced with unique fonts.

Referee 1

1. The authors correctly identify sensor wetting as a major challenge in thunderstorm UAS observations. The decision to flag questionable wetting periods rather than apply uncertain corrections is reasonable. However, the criteria for identifying wetting events remain partly qualitative. The manuscript states that typical wetting indicators include sudden temperature reductions, slower relaxation back to ambient conditions, increased relative humidity, and coincident radar-indicated precipitation. This is scientifically reasonable, but the exact thresholds or decision rules are not sufficiently formalized. The authors should clarify whether wetting identification was fully manual, semi-automatic, or automatic. If thresholds were used, they should be provided. If expert judgment was required, the manuscript should state how consistency was ensured across cases. Because wetting affects a substantial fraction of observations, especially in left-flank flights, this issue is important for downstream data users. Table 4 shows wet MHP occurrence of 18.4% for all storms and 30.7% for left-flank missions, and wet RSS-421 occurrence of 19.1% for all storms and 22.2% for left-flank missions.

We agree that we should have included more details on our approach to flagging potential sensor wetting. Our approach was entirely manual based on expert judgement and so no thresholds were used. We've added the following to the manuscript:

The approach was manual and iterative (i.e., cases were examined multiple times as author judgement was refined) and involved concurrent assessment of the radar reflectivity at the location of the aircraft and the character of the temperature and RH time series. Both WSR-88D and/or mobile radars were considered for coincident reflectivity, but set thresholds were not used since values of reflectivity associated with precipitation and not, for instance, biological scatterers, depend on radar characteristics, geography, and time of day.

2. The manuscript notes that the dedicated iPTH temperature thermistor functioned only 26.6% of the time and that a backup RH temperature sensor was used after hysteresis correction when the dedicated thermistor was unavailable. This is a reasonable solution, but the resulting variable may have heterogeneous response characteristics. The manuscript should more explicitly discuss the possible consequences of switching between the dedicated temperature sensor and the corrected backup RH temperature sensor.

The possibility of heterogeneous response characteristics for the RH temperature sensor was something we considered. More sophisticated corrections may be possible but the simpler first-order correction seemed to produce adequate values when compared to the dedicated temperature sensor (when it was working) and the RSS-421 sensors (when they weren't wetted by precipitation).

3. The introduction could more clearly state why RAAVEN observations are complementary to mobile mesonet, radar, lidar, and balloon-borne observations. The manuscript mentions the broader TORUS observing system, but the unique role of UAS observations in filling near-surface and low-altitude sampling gaps should be emphasized more strongly. Table 1 is useful, but it would be helpful to add whether each sensor was exposed or shielded, and whether the sensor was used for final reported variables, backup variables, or only quality-control purposes. Table 2 is very large and difficult to read. Consider moving part of it to supplementary material and retaining a condensed summary table in the main manuscript.

Regarding the "unique role of UAS observations" during TORUS, we've added the following to the Introduction:

To satisfy TORUS objectives, it was necessary to expose the 4D character of storm-generated airmass boundaries and coherent features and relate their 4D character to the thermodynamics and kinematics of supercell outflow. Since their cross-boundary scale is < 1 km, direct sampling of outflow thermodynamics and vertical motion is required. Neither remote platforms nor unguided in situ platforms are capable of consistently collecting thermodynamic and vertical motion observations at such small scales. Fixed-wing UAS like the RAAVEN are ideally suited to collect these data.

Regarding sensor information provided in Table 1, the additional elements requested by the reviewer are provided in Table 3. We believe the information should remain separated into two tables.

Regarding the size of Table 2, we agree that this table is quite long. It has gone through many iterations to keep it as concise as possible. We have moved it to an appendix.

4. The humidity correction equation is important, but the notation should be checked carefully. The text should clearly define all variables, units, valid ranges, and whether temperature is in Kelvin or Celsius. Since RH values are clipped at 99.9%, the authors should discuss whether this may affect analyses of near-saturated air, especially in storm flank or precipitation-adjacent environments.

The units for state variables are now included along with more information on the testing:

> Testing was performed in an environmentally controlled chamber over temperatures of 283-308 K and RH of 20-95%.

As noted in the manuscript, RH >98% only occurs 2.8% of the time, suggesting that clipping isn't occurring often enough to overly impact the resultant analyses. Importantly, this was for the left-flank mission only (the near-inflow mission didn't have an iPTH sensor) and, between the two mission areas active in 2023, it's the left-flank mission that would be most likely to encounter precipitation. This point was emphasized in the modified text.

5. The wind quality-control thresholds should be justified. For example, why were wind angles greater than 45 degrees, component magnitudes greater than 50 m s^{-1}, and uncertainties greater than 30 m s^{-1} chosen as thresholds?

Wind quality control thresholds were chosen based on reasonable bounds unlikely to be exceeded in flight; for instance, wind angles (which are actually the angle of attack and sideslip angles, a clarification we've made in the modified text) greater than 45 degrees are physically unfeasible for the platform in controlled flight. Uncertainty thresholds are similarly based on reasonable (manual) bounds. This has been added to Section 5.2.

Referee #2

1. The current introduction only vaguely mentions 'studying near-surface rotation,' which is too bland. It's suggested to clearly point out the scientific bottleneck—for example, that existing radar and soundings cannot resolve centimeter-scale wind shear/temperature gradient evolution in the left-flank boundary layer before a tornado—and then logically justify the irreplaceability of UAS, finally leading to the critical nature of data quality. This way, the QC standards directly match scientific needs.

We've included the following in the Introduction to add specificity to the motivation behind using UAS for this work:

> To satisfy TORUS objectives, it was necessary to expose the 4D character of storm-generated airmass boundaries and coherent features and relate their 4D

character to the thermodynamics and kinematics of supercell outflow. Since their cross-boundary scale is < 1 km, direct sampling of outflow thermodynamics and vertical motion is required. Neither remote platforms nor unguided in situ platforms are capable of consistently collecting thermodynamic and vertical motion observations at such small scales. Fixed-wing UAS like the RAAVEN are ideally suited to collect these data.

2. Currently, Section 6 and Figures 10/11 only list statistics showing that left-flank wind speed is higher than the inflow. It's recommended not to just report uncertainty percentages, but to clearly tell readers that, based on this QC dataset, these observational constraints can be directly used to verify the reference value of convective parameterization in high-resolution numerical models.

We didn't test whether the winds in any of the sectors were statistically significantly different from each other so while some differences are apparent in Figure 11, particularly at the higher speeds, we're not able to say that they're significantly higher in one or more sectors. The application of these results to convective parameterization, while an intriguing avenue for future research, is well beyond what we aim to achieve with this analysis but could be undertaken by the community with these data.

3. For the 'moisture flag' and 'questionable wind field flag,' only the trigger conditions are defined, but users are not told how to use them. It's suggested to add guidance at the end of the corresponding sections. For example, when RH is capped at 99.9%, advise users to treat this data only as a 'saturation indicator' rather than an exact humidity value. This kind of guidance can greatly enhance the usefulness of the dataset.

It is true that we generally adopt a nonprescriptive approach to data flags, instead allowing the user to decide how to handle flagged data. However, the reviewer makes a good point regarding the lack of guidance we provide. We have added the following to the description of "flag_shield":

> Ultimately, users of these data should have somewhat higher confidence in the quality of temperature observations collected in 2023 when "flag_shield" = 1.

We've also added the following to the description of iPTH RH correction:

> Nevertheless, the accuracy of reported RH values greater than ~98% should be considered less than the accuracy of values less than ~98%.

4. Was there any cross-calibration or consistency check performed between the 2019 and 2023 sensor configurations?

All PTH sensors were calibrated to references, often the RSS421, across several iterations of RAAVEN setups that were used in 2019 and 2023. Starting with the 2019 TORUS configuration, each successive RAAVEN configuration differed from the previous with the integration of additional sensors, based on opportunities to improve measurement quality, reliability, and redundancy. All aircraft underwent check-out flights for sensor calibration and cross-checks prior to deployment. This has been added to the beginning of Section 5.

5. Section 5.1, Equation (2) and subsequent clipping description. The correction Equation (2) is based on humidity chamber experiments, but the humidity/temperature ranges of the experiments are not clearly stated. Users cannot judge the reliability of the correction results when input conditions exceed the experimental range. Values of 100.5% and 105% are all clipped to 99.9%, which affects the variance among samples.

The units for state variables are now included along with more information on the testing:

> Testing was performed in an environmentally controlled chamber over temperatures of 283-308 K and RH of 20-95%.

6. According to the guidelines, reviewers should verify the dataset itself as much as possible. Due to time constraints, I was unable to fully download and test the complete dataset, but based on the statistics provided in the manuscript and the transparent disclosure of data quality flags by the authors, the overall quality of the dataset seems reliable.

We believe that our verification was comprehensive, but we also hope for further assessment by the community.

Referee #3

Major Comments

1. The title contains several uncommon abbreviations but lacks more general terms that directly describe the dataset and its scientific content. This may make the manuscript less discoverable in literature searches and may also make it difficult for readers to immediately understand the main subject of the paper.

This is a good suggestion. We've changed the title to "RAAVEN UAS data processing for supercell thunderstorm observations during TORUS and TORUS-LItE".

2. The Introduction would benefit from a more comprehensive background description. At present, the manuscript moves rather quickly to the RAAVEN deployments and data-processing procedures, while the broader scientific context is not fully developed. The authors may consider adding some background paragraphs that introduce supercell thunderstorms and the importance of in situ observations, briefly compare conventional severe-storm observing platforms with UAS-based measurements, and explain the advantages and challenges of RAAVEN observations in hazardous near-storm environments. These additions would help readers better understand the motivation for the dataset and the specific contribution of the RAAVEN observations.

We've included the following in the Introduction to add specificity to the motivation behind using UAS for this work:

> To satisfy TORUS objectives, it was necessary to expose the 4D character of storm-generated airmass boundaries and coherent features and relate their 4D character to the thermodynamics and kinematics of supercell outflow. Since their cross-boundary scale is < 1 km, direct sampling of outflow thermodynamics and vertical motion is required. Neither remote platforms nor unguided in situ platforms are capable of consistently collecting thermodynamic and vertical motion observations at such small scales. Fixed-wing UAS like the RAAVEN are ideally suited to collect these data.

3. In Section 3, the authors introduce the storm-relative mission areas used in the RAAVEN deployments, including the left flank, right flank, near-inflow, and far-field regions. However, the manuscript would benefit from a brief explanation of how these regions differ in terms of their dynamical and thermodynamical characteristics and how these differences motivated the deployment strategy. Such clarification would help readers better understand why the UAS observations were organized around these specific mission areas and how the resulting data may be used to examine different storm-scale processes.

We have added the following to Section 3:

> Left and right flanks are treated separately in the TORUS concept of operations because, while both flanks contain potentially distinct airmass boundaries and coherent structures, bridging them with a single set of UAS would ostensibly require operations in or very near the low-level mesocyclone where tornado formation is most likely. The near-inflow mission area was a focus for data collection because it would manifest storm-induced changes to supercell inflow and the far-field mission area was observed to characterize the "large-scale" environmental conditions within which targeted supercells resided.

4. Since different sensor configurations were used in 2019 and 2023, the authors should clarify whether any consistency check was performed between the two in situ observational datasets.

All PTH sensors were calibrated to references, often the RSS421, across several iterations of RAAVEN setups that were used in 2019 and 2023. Starting with the 2019 TORUS configuration, each successive RAAVEN configuration differed from the previous with the integration of additional sensors, based on opportunities to improve measurement quality, reliability, and redundancy. All aircraft underwent check-out flights for sensor calibration and cross-checks prior to deployment. This has been added to the beginning of Section 5.

5. The criteria used for wind quality control and subsequent statistical analyses should be better justified. It would be helpful for the authors to clarify whether these thresholds are

based on instrument limitations, previous studies, empirical testing, or sensitivity analyses, and whether the main conclusions are sensitive to these choices.

Wind quality control thresholds were chosen based on reasonable bounds unlikely to be exceeded in flight; for instance, wind angles (which are actually the angle of attack and sideslip angles, a clarification we've made in the modified text) greater than 45 degrees are physically unfeasible for the platform in controlled flight. Uncertainty thresholds are similarly based on reasonable (manual) bounds. This has been added to Section 5.2.

Minor Comments

1. Table 2 is very long and difficult to read in the main text. The authors may consider moving it to the supplementary material or the data repository webpage. If possible, basic information on each observed storm, such as its lifecycle, development stage, and intensity, could also be added.

We agree that Table 2 is quite long. It has gone through many iterations to keep it as concise as possible. We also agree that it is useful information to include in the data repository and we will add it to the existing documentation. We're reluctant to remove it from this manuscript given that no other reviewer suggested its removal, but we can appreciate that it takes up a lot of real estate in the manuscript and have moved it to an appendix. In terms of adding "basic information on each observed storm such as its lifecycle, development stage, and intensity", we can appreciate the value of this information, but it in the interest of brevity, it would require removing columns and we feel strongly that all columns currently included are important.

2. Please check all abbreviations throughout the manuscript and ensure that they are defined at first use.

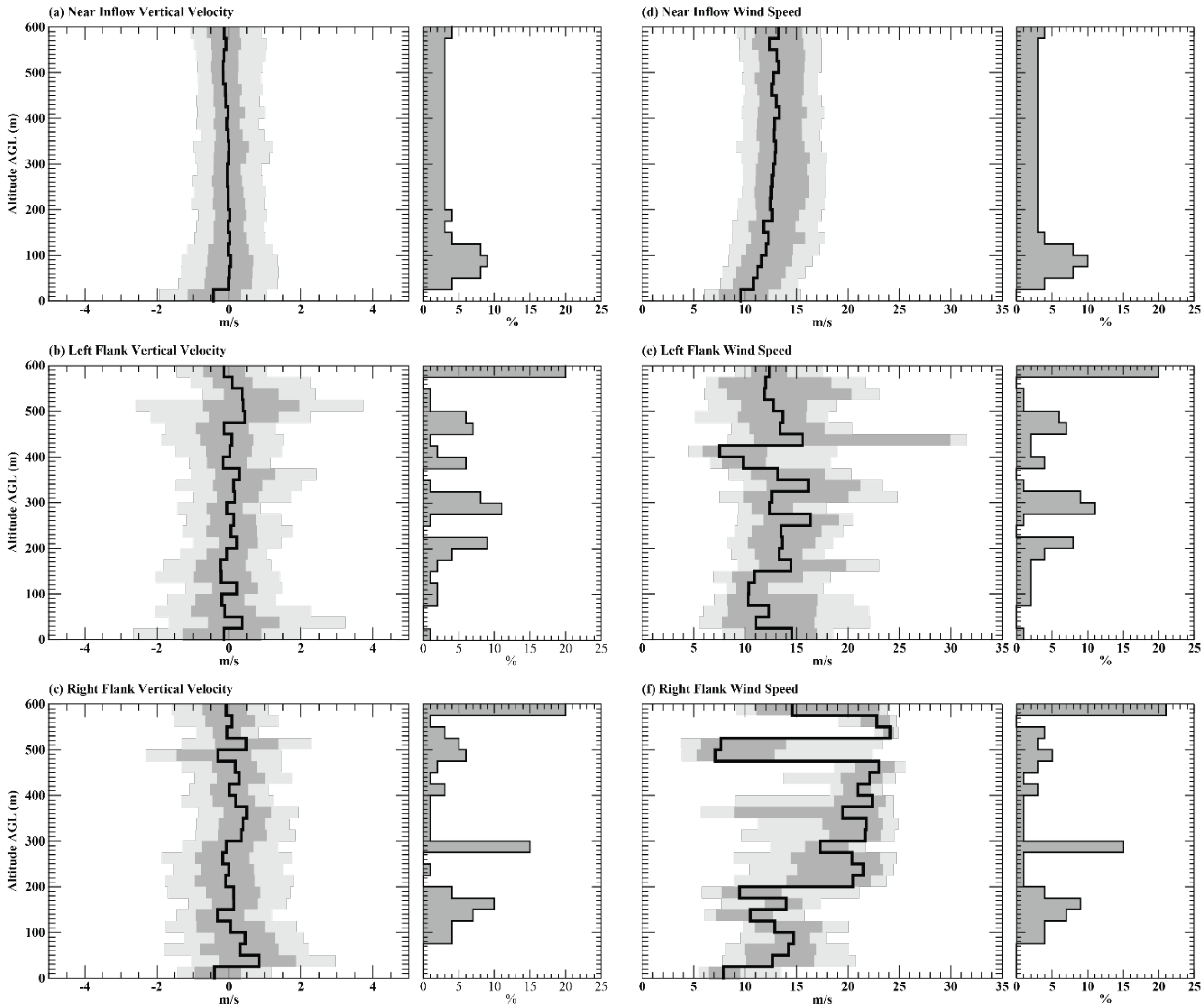
Corrected.

3. The caption of Figure 8 could be made more detailed to improve readability.

Changed as suggested.

4. Please briefly explain the apparent abrupt change shown in Figure 11f.

Most, if not all, of the discontinuities are attributed to sample size. Specifically, if more data are considered, the median vertical velocity values tend to be smaller. This is nearly impossible to discern from the previous version of this figure so additional observation frequency distributions have been added:



5. The summary section could include a brief outlook on future observational plans related to this project and discuss whether further intercomparisons with other TORUS observing platforms are planned to further evaluate the RAAVEN dataset.

We've added the following to the Summary:

> Ongoing and planned efforts aim to 1) better characterize the 3D wind fields derived from the multi-hole probe used on the RAAVEN through comparison to reference sensors, 2) improve the reliability of the multi-hole probe in precipitation, and 3) improve the overall reliability of the iPTH sensor suite, particularly the dedicated temperature sensor. None of these efforts are expected to improve the quality of RAAVEN observations collected during TORUS, but they are expected to improve data quality for future operations of the RAAVEN within supercells.

6. Please ensure that the Data Availability statement is presented as a separate section before the summary section, following the ESSD format requirements.

According to the Copernicus Word Template, this section should be “an unnumbered section…at the end of the manuscript”.

7. Please carefully check the formatting of the references for consistency with the journal style.

Checked. Hopefully we caught any inconsistencies.

Community Comment (Weiteng Qiu)

The manuscript provides a thorough description of quality control procedures but offers limited discussion of the physical insights that can be derived from the dataset. While differences between flight regions are documented—such as stronger vertical motions in the left- and right-flank compared to the near-inflow region—these results are not sufficiently connected to underlying supercell dynamics. A brief interpretative section linking these observed patterns to processes such as convective updrafts, outflow boundaries, and mesoscale circulation (e.g., rear-flank downdrafts or boundary-induced vorticity) would substantially enhance the scientific value of the paper.

> We’ve included text to the Introduction that adds specificity to the motivation behind using UAS for this work and, as such, offers a better idea of how these data could be used to yield physical insights into supercell structure:
>
> > To satisfy TORUS objectives, it was necessary to expose the 4D character of storm-generated airmass boundaries and coherent features and relate their 4D character to the thermodynamics and kinematics of supercell outflow. Since their cross-boundary scale is < 1 km, direct sampling of outflow thermodynamics and vertical motion is required. Neither remote platforms nor unguided in situ platforms are capable of consistently collecting thermodynamic and vertical motion observations at such small scales. Fixed-wing UAS like the RAAVEN are ideally suited to collect these data.
>
> Few if any new insights are likely enabled by the analysis presented in this manuscript. Instead, analysis presented herein, e.g., median vertical motion in the near-inflow tends to be slightly negative, is offered to support the consistency of these observations with expected conditions in and around supercells.

The manuscript applies a number of quality control thresholds (e.g., wind speeds exceeding 50 m s^{-1}, flow angles greater than 45°, or uncertainties above 30 m s^{-1}), but the rationale for selecting these values is not always clearly articulated. It would strengthen the methodological transparency to clarify whether these thresholds are derived from instrument limitations (e.g., sensor operating ranges), physical constraints (e.g., realistic bounds in supercell environments), or empirical tuning based on prior studies or dataset-specific characteristics. Providing this context—ideally with references or sensitivity tests—would help readers better assess the robustness and transferability of the QC procedures.

For thermodynamic data, although no set thresholds are used, the approach to assessing the occurrence of sensor wetting has been described in more detail in the modified manuscript:

> The approach was manual and iterative (i.e., cases were examined multiple times as author judgement was refined) and involved concurrent assessment of the radar reflectivity at the location of the aircraft and the character of the temperature and RH time series. Both WSR-88D and/or mobile radars were considered for coincident reflectivity, but set thresholds were not used since values of reflectivity associated with precipitation and not, for instance, biological scatterers, depend on radar characteristics, geography, and time of day.

Wind quality control thresholds were chosen based on reasonable bounds unlikely to be exceeded in flight; for instance, wind angles (which are actually the angle of attack and sideslip angles, a clarification we've made in the modified text) greater than 45 degrees are physically unfeasible for the platform in controlled flight. Uncertainty thresholds are similarly based on reasonable (manual) bounds. This has been added to Section 5.2.