Response to Reviewers

Helene Gloeckner on behalf of the authors

April 21, 2026

1 General comments

We thank the reviewers for sharing their time and expertise in the evaluation of our mansuscript. In response to the reviews and our own evaluation, the following changes were made to the datasets

- w from ASPEN was added to Level 2
- if $\sigma_{uv} > 3 \cdot \overline{\sigma_{uv}}$, `alt` is used as the altitude dimension and u and v are masked in Level 3. An additional qc flag `gps_valid` is introduced for these cases.
- changed the structure of the Level 2 data to follow the ragged array CF convention. This change was made to be able to store Level 2 data in a single dataset as well.
- DOIs were created for all data levels.
- dataset attributes for Levels 2, 3, and 4 were updated to follow the ORCESTRA attribute conventions (`https://orcestra-campaign.org/attribute_convention.html`).
- `omega_std_error` in Level 4 was corrected (sign error)
- add `sonde_id` to Level 3 QC dataset

Where required, the descriptions in the manuscript were adapted accordingly (Sections 3.2.1, 3.2.2, 3.3.1, 3.3.2, and the data availability section). The references to the datasets, Stevens et al 2026 and Winkler et al 2026 were updated. In addition, the introduction was reformulated slightly.

2 Reviewer 1

line 32 The citations in this sentence feel mismatched. The text states that Yanai (1961) applied the methods and produced some of the first estimates of W, but the supporting citations given are Reed and Recker (1971) and Yanai et al. (1973). Please clarify which study provides the first estimates and align the citations accordingly.

This paragraph in the introduction was reformulated.

Old Yanai (1961) applied these methods to sounding measurements over the west pacific to provide some of the first estimates of W in the tropics (Reed and Recker, 1971; Yanai et al., 1973). The utility of this approach was demonstrated during GATE, and in a great many field studies thereafter, as sounding arrays increasingly became part of the experimental design.

New Yanai (1961) applied these methods to sounding measurements over the west pacific to provide the first estimates of W in the tropics. The utility of this approach was demonstrated in subsequent analyses of the tropical atmosphere (Reed and Recker, 1971; Yanai et al., 1973), during GATE, and in a great many field studies thereafter, as sounding arrays increasingly became incorporated in the design of field campaigns.

line 77–78 The notation "ca 40 min" is unclear. Please clarify what "ca" means and what the 40 minutes refers to. Also explain how 40 min maps to $\approx$ 140 km diameter (state the assumed speed or conversion).

We reformulated the sentence to avoid unnecessary confusion.

Old A larger variation in circle diameter is associated with measurements in the East, where additional, ca 40min (ca. 140km diameter) circles were flown at lower altitudes to coordinate with MAESTRO measurements by the SAFIRE ATR-42 research aircraft near the Cape Verde island Sal.

New Additional circles with a smaller diameter of $\approx$ 140km were flown at lower altitudes in approximately 40min to coordinate with MAESTRO measurements by the SAFIRE ATR-42 research aircraft near the Cape Verde island Sal, leading to a larger variation in circle diameter in the East.

line 175–180 The QC step that removes measurements where `gpsalt` exceeds the aircraft altitude is described, but does not specify what aircraft altitude reference is being used or how it is made consistent with the dropsonde `gpsalt`.

We added a reference to the BAHAMAS measurement system that provides the aircraft altitude. We don't aim to make those measurements consistent with the sonde gpsalt, because it is only used here and only to remove GPS-altitude data that is obviously wrong (i.e. we had a sonde that claimed a GPS-altitude at 30 km). For all other purposes this is not relevant because, by the time the sondes calibrate, they are so far below that small discrepancies in the altitude measurements do not matter.

New [...] values for any measurement with a gpsalt above the aircraft altitude as measured by BAHAMAS (Konow et al., 2021).

line 15, 78, 83, 270 The repeated use of phrasing such as Windmiller and authors and "Bony and authors" is nonstandard and reads informal.

We removed these references as they were referring to papers that are not yet published or submitted. At the time of the first submission of this paper it was not clear whether they would be finished before the final submission. We also updated the reference to the ORCESTRA overview paper, which is now in press, and to the Radiosonde overview (Winkler et al 2026), which is now published.

line 182-183 In Section 3.5, the manuscript explains that JOANNE's profile fullness (sat-test) was reformulated as a profile sparsity metric based on missing values relative to a hypothetical perfect sonde. However, there is no clear connection of this reformulation to the selected threshold used in Section 3.2.1 (passing if $> 20\,\%$ of theoretically available data are present). As written, it is difficult to interpret what level of completeness this corresponds to. Citation: `https://doi.org/10.5194/essd-2025-647-RC1`

That is a mistake in the manuscript—thank for pointing it out. It should say that the test is passed if less than 20 % of the data is missing (or more than 80 % is present). We reformulated this sentence.

Old The profile-sparsity test is passed if more than 20% of the theoretically available data is present.

New The profile-sparsity test is passed if less than 20% of the theoretically available data is missing.

3 Reviewer 2

First, the output from Aspen, both netCDF and FRD, contain vertical velocity measurements. I think this is important since the study seeks to calculate large-scale vertical velocity. The dropwindsonde measures vertical velocity at the exact location of the instrument, whereas the study seeks to calculate averages within the aircraft circles, so the measurements are different. Still, they may provide some additional information for the study. Vertical velocity is mentioned starting on line 162, but is not mentioned after that.

We originally decided to not include vertical velocity in higher level BEACH products, because it needs assumptions on the idealized fall velocity of the sondes. Assuming hydrostatic balance, the fall speed of a sonde w_{sonde} should be the difference between an idealized fall and the actual air velocity w_{air}:

$$w_{\mathrm{sonde}} = w_{\mathrm{idealized\ fall}} - w_{\mathrm{air}} \tag{1}$$

$$= -\frac{1}{\rho g}\frac{\mathrm{d}p}{\mathrm{d}t} - w_{\mathrm{air}} \tag{2}$$

By our reckoning w_{sonde} should only be dependent on air density if it is perfectly constrained and equal for each individual sonde. Fig. 1 shows a spread of $\approx 2\,\mathrm{m\,s^{-1}}$ for any given air density, which means that w_{air} is only accurate up to $1\,\mathrm{m\,s^{-1}}$ to $2\,\mathrm{m\,s^{-1}}$. In our context this uncertainty is substantial.

In addition, the experiment was conducted to derive mesoscale vertical velocities and the vertical velocities of individual sondes do not hold information on them. However, the reviewer comment helped us appreciate that for some applications w as output by ASPEN might be interesting. For this reason we added it to the Level 2 data. However, we do not include it in Level 3 so as to not cause confusion between area averaged vertical velocities and point measurements.

New Sec 3.2.2 Variables in Level 2 Only the measurements for temperature (ta), relative humidity (rh), pressure (p), and the wind components u (u), ~~and~~ v (v), and w (w) are transferred from the Level 1 output to the Level 2 dataset.

New Sec 3.3.2 Variables in Level 3 Vertical velocity w of individual sondes as estimated by ASPEN is removed between Level 2 and Level 3.

In addition, the vertical velocity can be used to assess the validity of other observations from the dropwindsondes. Some instruments are fast falls, or partial fast falls, and some of the observations may be questionable due to the parachute not properly being deployed. It is not mentioned in the text whether this is used, though Aspen flags fast falls. However, not all fast falls are found by Aspen, so

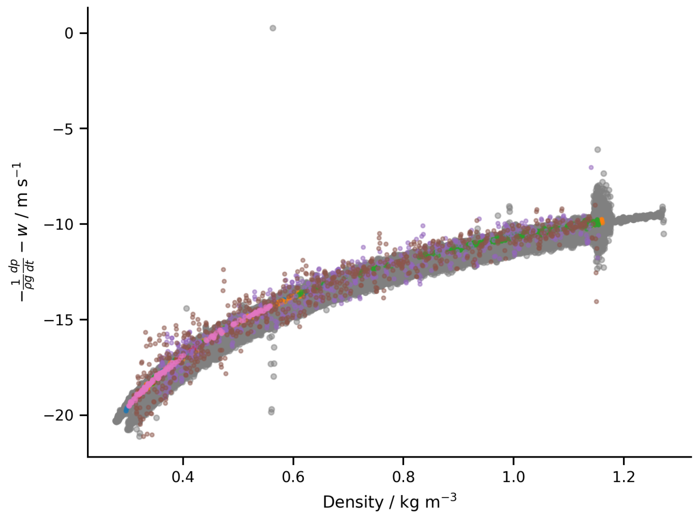

Figure 1: RHS of Eq. (2) vs density for all sondes. Colored dots are the same as in Fig. 2.

research-quality datasets require some manual checks. Line 143 suggests that a launch detect means that the parachute opened properly, but it only means that the parachute opened.

We did not think about the consequences of fast falls before, but realize now that they substantially impact the GPS measurements. To automatically determine fast fall sondes, as well as other sondes with an unusually large vertical variation in u and v, we calculate

$$\sigma_{uv} := \sqrt{\sigma^2_{\Delta u} + \sigma^2_{\Delta v})}, \tag{3}$$

where $\Delta *$ is the difference of $*$ between two consecutive measurements.

We then define a threshold of $3 \cdot \overline{\sigma_{uv}}$ above which `alt` is used as the altitude coordinate and u and v are masked in Level 3, because we assume the GPS measurements to be unreliable. A large σ_{uv} can be caused by either fast falls or large gaps in the GPS measurements. A factor of three is chosen to not exclude extreme, but valid u and v measurements. Inspection of u and v showed that the factor of 3 only excludes profiles with large gaps or noise (Fig. 2 right), and that a factor of 2 would exclude valid u and v profiles.

Of the excluded sondes, only two (`f472b50c` and `b75ffee3`) have had an unusually fast fall (Fig. 3), and they do not bias the estimates for w_{sonde} in Fig. 1 (see colors).

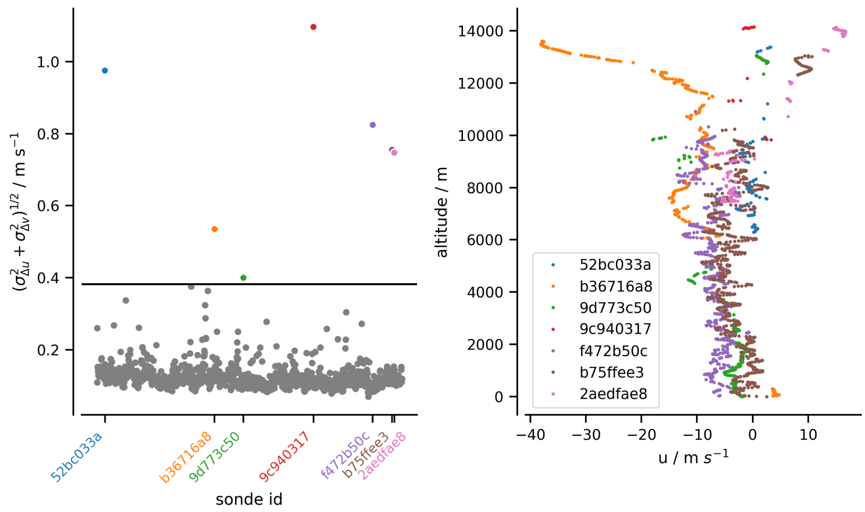

Figure 2: σ_{uv} for all sondes in grey and colored for sondes with $\sigma_{uv} > 3 \cdot \overline{\sigma_{uv}}$ whose winds are now masked from Level 3.

New Sec 3.2.1 Quality control tests QC:gps-valid: A second QC checks whether the u and v measurements have an unusually large variation. It is failed if $\sigma_{uv} := \sqrt{\sigma^2_{\Delta u} + \sigma^2_{\Delta v}}$, exceeds three times the mean σ_{uv} of all sondes: $\sigma_{uv} > 3 \cdot \sigma_{uv}$, because large σ_{uv} values indicate that the GPS measurements are faulty, either due to fast falls or large gaps in the GPS measurements.

New Sec 3.3.1 Defining a common altitude For sondes that do not have a valid gpsalt profile; i. e. if the gps-valid QC failed, the alt variable is used for the altitude if a valid surface-p measurement was taken. In addition, if the gpsalt profile is incomplete, [...]

line 150 Line 150 starts a description of Aspen. It isn't clear whether the default settings of the Aspen configuration were used or not. The default settings were designed for operational quality control so that data can be assimilated into models and used for situational awareness. Different settings can be used to create research-quality datasets. It isn't mentioned whether any changes were made.

We used the default settings (following the JOANNE processing). It is now specified in the paragraph.

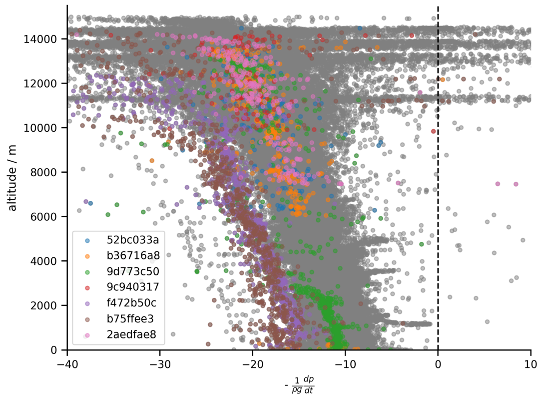

Figure 3: $\frac{\delta z}{\delta t}$ for all sondes in grey. Colored sondes have a σ_{uv} larger than the threshold (Fig. 2).

New A docker image containing the command-line functionality of ASPEN is utilized within the processing pipeline with the default *editsonde* configuration.

line 151 Line 151 suggests that Aspen removes post-splash data. This is not the case; Aspen flags cases in which post-splash data may exist and asks the user to manually set the end of drop. Further, Aspen frequently misses cases that do have post-splash data. A good check is to look at the vertical velocity near the surface. If the vertical velocity in the processed data is large and positive near the surface, the product likely contains post-splash data.

This half-sentence is removed.

New The ASPEN processing includes several quality control steps, such as removal of ~~post-splash data and~~ the equilibration period, outlier checks, smoothing, and outlier checks as described in the ASPEN Manual (AspenDocs 1.0)

Lines 152-165 The characteristics of the sondes can be changed in the configuration, so that the correct vertical velocity can be recovered.

Line 175 The altitude can be removed via the configuration settings.

We did not want to change the default configuration settings to be consistent with the EUREC^{4}A datasets and to clearly separate our changes from the ASPEN default. Since we do not include the sondes vertical velocity in our Level 3 data, the mini-sonde configuration does not negatively influence our data (raw data and Level 1 ASPEN output are also available if different specifications are needed). Similarly, it is not necessary to remove the altitude automatically, because it is only used for the `profile_extent` test.

Line 179 Sondes frequently move upward, but this probably doesn't happen outside of deep convection. `https://doi.org/10.1175/MWR-D-18-0041.1`

In our measurements only very few sondes show a positive fall speed (Fig. 4), probably because the conditions we measured were relatively calm compared to hurricanes where dropsondes are frequently used. The positive values in Fig. 4 are likely measurement errors in GPS, because if $-\frac{1}{\rho g}\frac{\mathrm{d}p}{\mathrm{d}t}$ is used to calculate fall speed instead of $\frac{\mathrm{d}z}{\mathrm{d}t}$ from GPS, no sonde shows a positive fall speed. Additionally, most positive values occur close to the drop where GPS is not yet reliable. Even in the GPS fall speed, there is no sonde with a prolonged positive fall speed. For these reasons we left the processing as is.

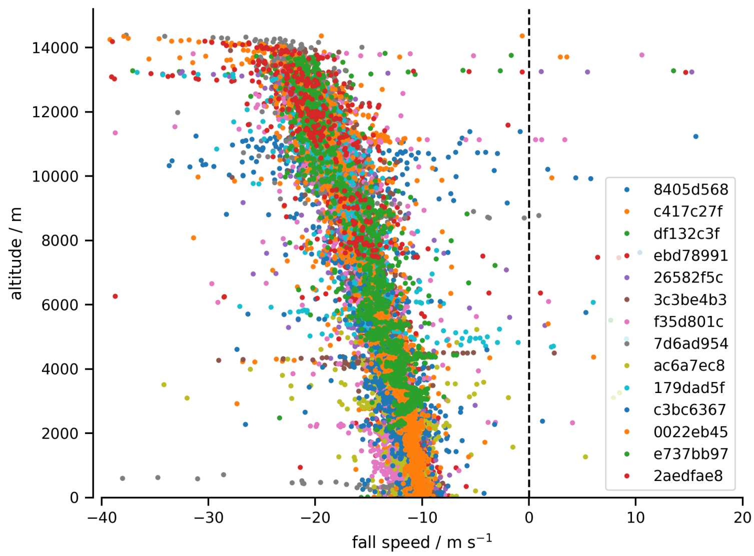

Figure 4: Positive Fallspeed in Sondes

Lines 224-224 Is there a time difference between two good points larger than which interpolation is not done?

No, currently there is no limit. In this step, only altitude is interpolated to not lose data from other sensors that were taken at different time steps, i. e. no other variable is affected by this interpolation.

Line 405 It may be useful to cite `https://doi.org/10.1175/2010JCLI3496.1`.

We appreciate the reviewer's comment, but in attempting to draw the reader's attention to the reference, or justify the citation, we found ourselves making an excursion to explain ourselves. This interrupted the flow of the narrative, and added relatively little, and so we did not in the end adopt the reviewers suggestion.

The main reference for the GPS dropwindsonde used here is `https://doi.org/10.1175/1520-0477(1999)080<0407:TNGD>2.0.CO;2`. Characteristics of the NRD41 sonde specifically are presented in `https://doi.org/10.1175/BAMS-D-22-0119.1`. Citation: `https://doi.org/10.5194/essd-2025-647-RC2`

We added a citation and a reference to NCAR in Sec. 2.1.

New The dropsondes used during PERCUSION are of the type RD41 developed by NCAR (Hock and Franklin, 1999; Aberson et al., 2023) and manufactured by Vaisala.