Comment ESSD-2022-03

Thank you for this careful read of our paper and for these very constructive comments. Answers to each criticism are addressed below (in italics).

Need to measure water vapor accurately - a hard-enough problem - then determine isotopic composition which adds substantial additional challenges. Authors apparently have skipped first step, going directly to the second step. But, the first step provides very necessary constraints on instrument performance? E.g laser-based instruments - according to manufacturers - lose precision at high humidities? These authors tend to focus on low humidities more typical of natural environments but still minimize those effects? Dismissed here as so-called 'baseline effects' (line 156).

To communicate instrument performance specifications and describe more clearly our confidence in the water vapor measurements, we have added a new paragraph, at the top of Section 3.1, that conveys the following information. First, previous lab and field-based studies have demonstrated that the water vapor concentration measurements from the types of laser analyzers deployed during EUREC^{4}A are precise and stable over a large concentration range (e.g. 200 - 30,000 ppmv) and over long periods (e.g. years). Second, the latest version analyzers are now designed to perform optimally in the range 1,000 - 50,000 ppmv. The new paragraph also points readers to other ESSD special issue data papers that provide more detailed comparisons between humidity measurements from the airborne isotopic analyzers and other aircraft sensors, which demonstrate the accuracy of these systems.

Furthermore, the effects of any inaccuracies in water vapor concentration on the isotopic measurements are removed when correcting the isotopic data for known biases associated with low water vapor concentrations and normalizing them to the VSMOW-SLAP scale. Section 3 provides extensive details on these corrections and refers readers to early papers on this topic such as Aemisegger et al. (2012) and Bailey et al. (2015).

Authors present data compiled from ground-, air-, ship-, and lab-based systems using platform-specific inlets, standard or customized instruments of varying reliability, response time and sensitivity, calibrated (or, not) through independently-determined procedures. None of which they control or even influence! Remarkable effort to even present such an assemblage; good on them for the effort. Strongly agree with one of their summary sentences, e.g. from line 793 "challenge of accurately characterizing and correcting for all relevant biases in field deployable water vapor isotopic instruments". This reader thinks they provide very skillful assessments; who if anyone could have handled and described such a variety of data. ESSD should publish careful data compilation efforts regardless of 'success' or 'failure' which mostly derive from needs and interests of users. Authors could/should make better presentation!

Thank you for this encouragement. We have reworked the Abstract and Introduction so that they better describe the needs met with this collection of data (namely, closing water isotopic budgets, characterizing fluxes, and evaluating isotopically enabled numerical simulations) while also giving readers more context on data uncertainties and quality flags. Specifically, by providing detailed reports of uncertainties and flagging (rather than masking) data, we hope to promote more accurate cross-platform comparisons, raise the bar for uncertainty reporting within the nascent water vapor isotope measurement community, and facilitate an open dialogue around improving instrument performance, sampling installation, collection and calibration protocols, and, ultimately, data quality. There should be no doubt, however, that EUREC^{4}A's isotopic measurements were an enormous success, which our new framing aims to convey more clearly.

Basically, we need cautions and overall disclaimers upfront, e.g. in Abstract. Even to finish with a sentence or two about necessary cautions in many uses of these data? As now presented, the abstract offers only deployment summaries, e.g numbers of instruments on which platform. Substantial uncertainties and cautions only emerge in section 4. Give readers an earlier hint?

We have reworked the Abstract so that it lists data of quality concern and refers readers to Table 3, Section 5 Data availability for a comprehensive list both of dataset DOIs and dataset flags. If one excludes these flagged data from comparisons, as we have done in Section 4 (except when wanting to highlight discrepancies), data users should be able to merge individual datasets as needed for greater spatiotemporal coverage or for closing water isotopic budgets. The cross-platform comparisons in Section 4, especially the strong coherence illustrated by Figures 10, 11, and 13, further bolster our confidence in this statement.

Line 47 conflicts with line 45. E.g. if ships collected seawater (line 45, also at line 64) then vertical extent of samples can not start "a few meters above sea level". Perhaps for vapor phase but not for all isotopic samples?

We have clarified this point by stating that vapor measurements extended from a few meters above sea level to the mid-free troposphere and that seawater samples spanned the ocean surface to several km depth.

ESSD will require DOI information (referred here to Section 5) repeated at end of abstract. Section 5 unfortunately reports URLs (unreliable), not DOIs (reliable). Table 3, with individual data sets references by DOI seems, again unfortunately, incomplete. Or, incompletely documented. User needs easy access to full set of products. Either convert one of the AERIS or NCEI links to a DOI labelled product or put all products together under a third-party archive service (e.g. Zenodo?) which will provide top-level encompassing DOI, reliable off-site storage, plus very good version control. Not acceptable in current form or format.

We have followed ESSD special issue papers Quinn et al. (2021) and Bony et al. (2022) in referring to the Data availability section of the manuscript rather than listing individual DOIs in the Abstract. However, we now make reference to Section 5 Data availability at the end of the Abstract (rather than in the middle). If this is insufficient, please let us know, and we will make additional changes.

Our decision to provide individual DOIs in Table 3 (Section 5 Data availability) was motivated by ESSD special issue papers Pincus et al. (2021), Quinn et al. (2021), and Bony et al. (2022). Our table provides searchable DOIs for every dataset except the Brown sea samples, which are provided in the Supplemental. We have opted to continue providing data users with individual dataset DOIs while also requesting an umbrella DOI, as recommended. The new umbrella DOI will be furnished as soon as possible. For now, we have included a placeholder statement in Section 5.

One additional note: it appears there was a PDF conversion issue, which truncated the DOI hyperlinks in Table 3, sending reviewers to an error landing page. That has been corrected.

Line 95: “last” I think you mean ‘most recent’?

Indeed. We have removed this particular sentence and shifted much of the associated discussion to Section 6, where we discuss future scientific uses for the data collection.

Line. 118: “new” I think you mean additional, especially because you have just expended several sentences to justify isotopic measurements based on past data.

Substitution made.

Section 2: not clearly specified, perhaps will clarify later, but I suspect:
7 vapor-phase measurements: 2 at BCO, 1 on ATR, 1 on P-3, 3 on ships, sum = 7? 5(?)
liquid-phase (precip) measurements: 1 at BCO, 3 on ships above, 1 on additional ship = 5? 3 seawater samplers on 3 ships = 3?

Yes; we have rephrased here and in the Introduction to make this more clear.

From Fig 2, rainwater samples occurred mostly along longitude approx 57W, with relatively few exceptions.

The eye is drawn to that meridian because the Meteor, represented by magenta symbols, sampled almost exclusively there (over a relatively short N-S transect). However, as seen in Fig. 3, rain samples from the Meteor comprise but a small portion of the total number of rain samples.

Line 145: very strong statement here: "… no island effects …". No upwind island effects? This needs referencing to back it up?

The appropriate reference (Stevens et al., 2016) was cited in the previous sentence and has now been moved (one sentence down) for clarity.

Line 148, "regionally representative": I think you mean representative across tropical trade wind environments globally but as written readers could interpret statement as referring only to local BCO environment? Needs some revision.

Indeed. Revised as suggested.

Line 198, "stored at room temperature": this reader doubts that authors could keep samples at steady temperature during long-range (BCO to Freiburg lab) long-duration transport and storage. Perhaps not important for isotopes? Need some justification here?

We have modified the sentence so that it begins: "except when in transit". Shipping samples for laboratory analysis is common practice and not expected to fractionate the samples so long as they are capped tightly and/or wrapped in a sealant like parafilm. If caps and/or seals are inadequate in preventing evaporation, colder temperatures can be advantageous because they minimize isotopic fractionation. However, true leaks are likely to be apparent in liquid samples, regardless of the storage temperature. We have reported storage temperatures mainly for consistency and because there were some differences from one platform to the next. However, as now emphasized in Section 4.1.1, these differences have had no detectable effect on cross-platform comparisons.

Line 219: that the Picarro instrument worked (acceptably?) during deployment nearly a decade earlier provides little help here about reliability, precision, accuracy?

As described in both Sections 2 and 3, the ATR analyzer was calibrated extensively in the field and after the EUREC^{4}A deployment. All post-processing corrections were based on these calibration checks.

To avoid confusion, near Line 219 we have rephrased as follows: "Isotopic measurements aboard the ATR were made with a customized, fast-response version of Picarro's L2130-i cavity

ring-down spectrometer (with nominal sampling frequency of 1 Hz). The analyzer had been deployed in previous field campaigns, both for near-surface (Thurnherr et al., 2020) and airborne (Sodemann et al., 2017) measurements."

Line 222: please give temperatures consistently in K or C, or explain why one seems appropriate in some cases but not others? Here readers confront inlet reference temperatures in K followed - within same sentence - by tubing heating temperatures in C! One or the other? Not both unless justified for valid reasons.

Thank you for catching this. We have converted K to degrees C. Degrees C are already used exclusively in the Supplemental Information.

Line 292: text here, e.g. "limit particle debris" bears (too) remarkable similarity to earlier text describing aircraft inlets (e.g. line 224). One suspects authors adopted text from ship and aircraft sources - fair enough - and no doubt particulate contamination proves troublesome in both cases but authors either need to acknowledge similarity of text (and sampling challenge) or take more care about repeating identical phrases.

The same filter was used in both cases for the same purpose. However, we have modified the text in the first instance to avoid using the same exact turn of phrase: "Ambient air was pumped at a rate of 13 SLPM…through the gooseneck, past a particle filter, and down a 1.5 m long, 10 mm ID PTFE tube…"

Line 294: "sniff tests"? Sloppy, at best. Sniff for RH changes? For odiferous tracers? Certainly not for isotopic composition! Again one suspects authors repeat text from ship-board operators / data providers but seems seriously out-of-place in a careful presentation of isotope data.

We have replaced "sniff tests" with "empirical time-response tests".

Line 427, Section 3 on data processing expends several sentences on water vapor quantification, both as mole fraction and as humidity. Surprised because not addressed in measurement sections?

We have revised the top of Section 2 so that it is clear that all analyzers deployed during EUREC^{4}A measure water vapor as a mole fraction. The statements at the top of Section 3 are meant to explain that some datasets have converted mole fraction to another expression of humidity, such as specific humidity, for ease of comparison with other humidity sensors on the same platform.

Line 439: “water vapor isotopic measurements varied widely”. Okay, not really a surprise, but have authors given us tools to adjust expectations and make use? For this reader, no.

To clarify, the statement here makes the case that the post-processing of measurements varied widely. In other words, bias corrections were tailored to each individual analyzer, as is best practice in the water vapor isotope measurement field. We have added new material to the Introduction to clarify this need upfront: “Uncertainties in these data reflect the diverse operating conditions and constraints associated with each platform and the need to tailor post-processing corrections to individual instrument performance, as is considered best practice (Aemisegger et al., 2012; Bailey et al., 2015)”.

We have also made minor modifications to the top of Section 3.1, adding, for example, the following statements: “post-processing corrections ensure that measurements from any one platform are not only self-consistent but also comparable with measurements from other EUREC^{4}A platforms or with isotopic data from previous field deployments…Accounting for…uncertainties and excluding data flagged for quality concerns will ensure that cross-platform comparisons are as accurate as possible”. By applying the appropriate, tailored corrections, and flagging data of quality concern, we have made sure that the remaining high-quality data are comparable.

Side-by-side systems at BCO should have provided “reference dataset with 1 minute time resolution”. Failed. Authors conclude (line 793): “unexpected discrepancy between the BCO analyzers”. Comparable systems on ATR and P-3 also failed intercomparison critera. Data not bad from ATR (albeit with uncertain response times and corrections) but not - unfortunately - comparable to P-3 data which, apparently, only reference to prior P-3 data. Comparable systems on ships? Only barely useful. Authors know these discrepancies better than any reader but fail to provide a coherent summary. If they can’t extract useful summary, who can?

These comments make us realize that “reference” is probably too loaded a term. We have replaced “reference” with “continuous fixed-point”. Also, although one of the BCO analyzers exhibited problems with its spectroscopy and unexpected isotopic biases, the other analyzer’s dataset serves as a downwind anchor point for ship-based and airborne measurements made to the east, which is already discussed both in Sections 4 (time series comparisons between the Meteor and BCO) and 6 (opportunities to carry out Lagrangian analyses).

Perhaps we have downplayed the very strong coherence in the EUREC^{4}A isotopic data to a fault. For example, Fig. 13 demonstrates that the two airborne analyzers show very strong coherence in the boundary layer, where most of the ATR data are from. Similarly, Figs. 10-11 show very strong coherence in campaign-mean water vapor, rainwater, and seawater values. We have revised much of Section 4.1 so that the data consistency emerges as the key take-home message. We have not removed any discussion around cross-platform discrepancies; however,

we have attempted to communicate more clearly that these discrepancies are either the result of unique platform issues (e.g. spectroscopic oscillation in the BCO OA-ICOS time series), which we have identified and flagged, or the result of real environmental variability (e.g. lower isotope ratios associated with frontal rain as compared to shallow convective precipitation).

Authors informed assessments of data utility (e.g. Section 4):
For vapor phase, perhaps a “subtle” (or, later, “very subtle) latitudinal pattern emerges. This reader can neither see nor credit such spatial patterns but accepts that authors and other users might. These cautions need to move earlier, e.g. in abstract?

Perhaps we’ve misunderstood the criticism, but, in our opinion, the subtle patterns indicate strong coherence, which (as pointed out near the top of Section 4.1.1, lends “confidence to the measurement accuracy”). We have revised Section 4.1.1 such that the take home messages are, in order of importance, as follows:

1. *Data coherence is strong, indicating high quality data*
2. *Accounting for quality flags is important for accurate cross-platform comparisons*
3. *Differences in high-quality data represent real environmental variability and are consistent with theoretical expectations.*

The specific reference to the “very subtle” pattern in rainwater samples has been removed, as we choose to emphasize, instead, how very close in mean value the rainwater samples are.

Storage effects, which vary from ‘frozen’ to even ‘poisoned then frozen’ to room temperature and even ‘subject to drastic heating’, impose evaporation and biological effects on isotope ratios. Authors recognize such issues and provide (where possible) summaries of storage protocols but only rarely deal with the larger issue? E.g. Brown samples should differ substantially from Meteor samples due to differences in storage? Systematic or erratic? Not clear and not well addressed?

As explained in a previous response, in general, we do not expect basic storage differences among platforms to affect sample quality. We have also added a statement to Sect. 4.1.1 that suggests that the high cross-platform consistency provides compelling evidence that differences in sample storage did not influence sample isotope ratios.

As now more clearly emphasized in the Introduction and Sect. 4, we do not recommend liquid water samples flagged for data quality concerns for most scientific analyses.

Line 1425: Given the scale of aircraft tracks, land mass (grey) in Figure 2 represents South America with Trinidad/Tobago clearly visible. Barbados (13N, 59W), apparently in black, fails to

appear to these old eyes even under extreme zoom. All land masses should stay dark grey, particularly if you want color matching to Fig 3. Show location of Deebles Pt BCO? As readers find later, this scale driven by satellite products while less helpful for immediate locale.

Actually, the maps are scaled to fit the in-situ data. The EUREC^{4}A satellite datasets cover a much larger region than that shown.

As per the suggestion, we have remade the Figures so that Barbados appear in gray, rather than black.

Authors make extensive reference to and use of column-integrated satellite products, on two or more spatial scales. Not surprising given EUREC4A relevance and motivation, and authors have offered good access and reasonable interpretation. Summary (in my words): not bad, nothing in remotely sensed data proves or disproves EUREC4A in situ data but subsequent users should take great care in any such intercomparisons for a variety of reasons. Some such caution should emerge as a clearer outcome?

The idea of including the EUREC^{4}A remotely sensed isotope ratios was to provide broader spatial and temporal context for the in-situ measurements with the full understanding that such comparisons require care. Indeed, Section 4.2 contains several strong cautionary statements about comparing remotely sensed and in-situ measurements. Note, for example, the following: "As demonstrated in Fig. 14, the satellites provide rich spatial context for the in-situ data. Nevertheless, when using the two in tandem, care must be taken to consider differences in what each type of measurement represents." The next several paragraphs all discuss the different sensitivities of the measurements and the need to consider averaging kernels.

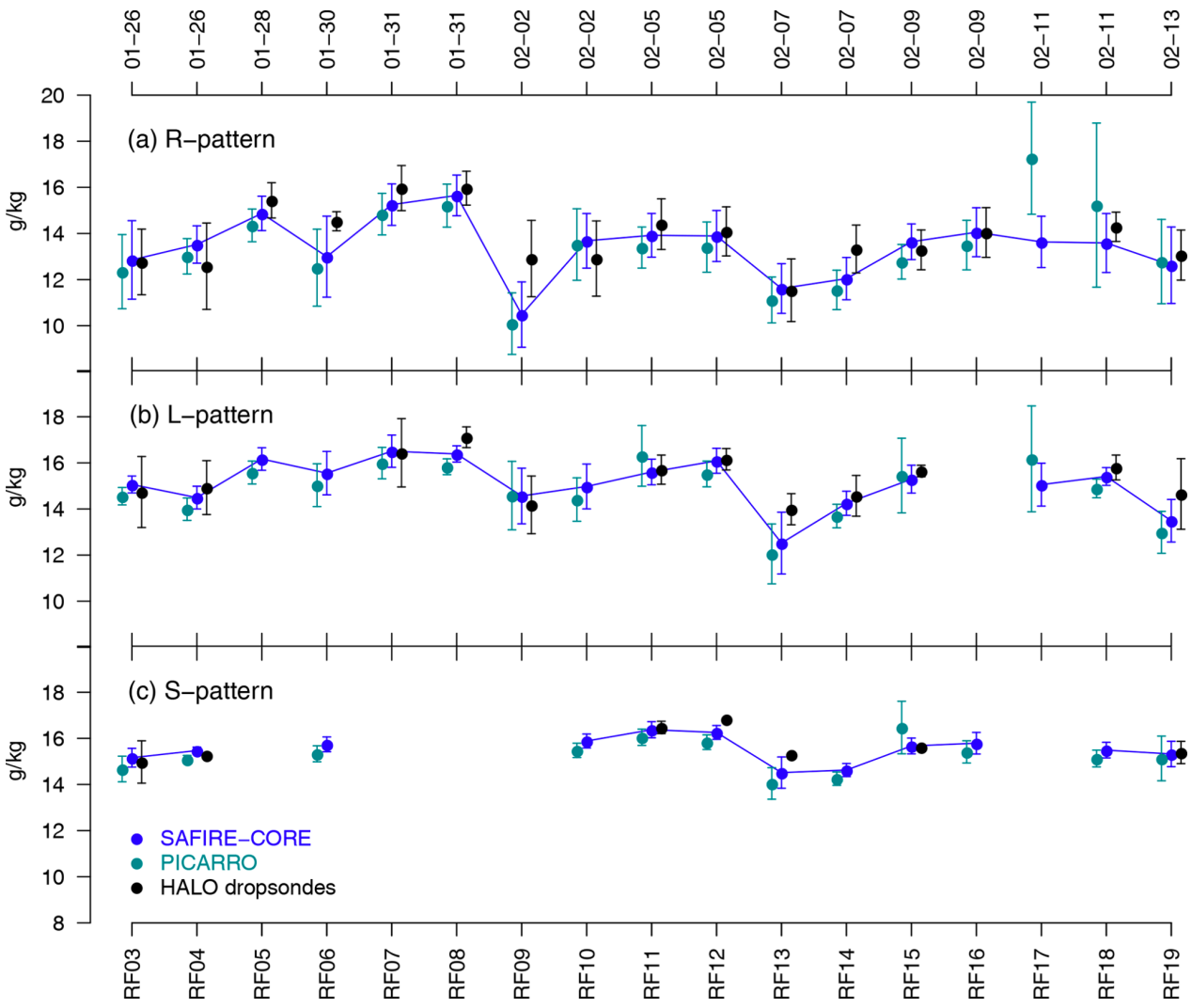
To provide more caution up front, we have added statements to the end of the Introduction and to the beginning of Section 4.2 that emphasize the vastly different sensitivities to the atmosphere between the remote sensors and the water vapor isotopic analyzers deployed in situ.

Because of focus on vertical profiles driven by remote sensing, this reader notes again absence of radiosonde profiles (e.g. BCO must launch sondes daily) and of dropsonde profiles from HALO and P-3. Because EUREC4A expended efforts in track planning and resources in sondes themselves, and because correlation with water vapor / humidity turns out such an important factor in isotope measurements, why have authors not at least mentioned sonde humidity profiles? Even to say 'not useful'. To this readers, seems a strange omission.

Based on this recommendation, we now encourage data users to seek out ESSD special issue papers on EUREC^{4}A radiosonde and dropsonde measurements at the top of Section 4 and explain at the top of Section 3 that multiple expressions of water vapor concentration are often

included in the data files for ease of comparison with other sensors on the same platform, dropsondes, or radiosondes.

Sensor intercomparisons have also been performed in other already published ESSD papers; therefore, we decided not to repeat these results in this manuscript. To satisfy the editor's well-appreciated curiosity, we have reproduced Fig. 16 from Bony et al. (2022) below, which compares specific humidity data (in g/kg) from different sensors onboard the ATR, including the Picarro isotopic analyzer (teal), and the HALO dropsonde data for different flight patterns ("R", "L", "S").



Authors use term 'diurnal' when in fact they mean 'diel'. Strictly, diurnal refers to daylight, nocturnal refers to night, diel refers to full 24-hour cycle.

Fixed.