Contents

Response to Referee #1 Comments

Overview:

Review of Chen et al.: Airborne eddy covariance of ocean-air VOC fluxes: distinguishing signal from noise
The manuscript uses PTR-TOF-MS-measured VOC data from an airborne campaign over the North Atlantic to calculate airborne fluxes. The data is characterized by low signal-to-noise ratios, which the authors use as an opportunity to explore different sources of errors, especially uncorrelated sensor noise. This systematic analysis will be a valuable source of information and inspiration for scientists who want to conduct similar measurements and have to decide which kind of sensor to take on board, or who are analyzing airborne flux data and are wondering which sources of error to take into account. I do not have many criticisms, my main concern being that the authors ignored vertical flux divergence as a potential source of systematic uncertainty and as a potential reason for underestimation of the flux values which they report. The text is generally well written, the figures could partly be presented in a clearer way. Some more specific points are below. They should be addressed before the manuscript can be accepted for publishing in ACP.

Thank you for the very thoughtful review and constructive suggestions. We have added a discussion of the potential role of vertical flux divergence to the revised manuscript (see response to l. 467 ff). Specific point-by-point responses are listed below in blue, with the corresponding updates to the manuscript text indicated in green/quoted.

We have also made some additional minor corrections as reflected in the updated manuscript with tracked revisions. Furthermore, during the revision process we identified a few additional flight legs that were suitable for flux analysis. Since their inclusion improved our data coverage we have added them to our data ensemble used for analysis (with corresponding changes tracked). None of the main discussion points or conclusions are altered in any substantive way.

Specific comments:

l. 106: How low were the low-level legs, and how long were they? The actual flight altitudes should be given, since this influences the potential importance of vertical flux divergence on the reported fluxes. I found the information in Table S1 (which is not even properly referenced in the text), but a summary of the range should at least be given in the main text. Or just put the table in the main manuscript.

The altitudes of low-level legs are summarized in Sect. 4.1 with details listed in Table S1. We have added text at the location mentioned by the reviewer to direct the interested reader to these specifics:

"The C-130 flights included multiple low-level MBL legs (detailed in Sect. 4.1) ~~within the MBL~~, satisfying one of the prerequisites for deriving air-sea fluxes by airborne EC."

We have also made some corrections to the content in Sect. 4.1, accounting for the additional flux legs noted above:

"**Table S1** lists the 44~~34~~ flux legs from 13~~11~~ NAAMES-2 and NAAMES-3 flights identified in this way, which average 9~~8~~ min (~72~~64~~ km) in duration and range from 3-25 min (~20~~22~~-210~~194~~

km) at mean sampling altitudes of 126-219 m. Mean airspeeds range from 111~~102~~-157 m·s^{-1} across the flux legs, corresponding to approximately one 5 Hz data point per 22~~20~~-31 m."

l. 108: What type of PTR-TOF-MS? Brand name/mass resolution of the TOF should be given (basics should not require looking up the Müller et al. paper)?

We have now updated the manuscript (Sect. 2) to include more details on the deployed instrument:

"VOC mole fractions were measured on-board the C-130 by proton-transfer-reaction time-of-flight mass spectrometry (PTR-ToF-MS) during NAAMES-1 (for 4 of the 7 flights), NAAMES-2 (9 of 11 flights), and NAAMES-3 (9 of 12 flights). The employed instrument is an upgraded version of the prototype PTR-ToF-MS 4000 described by Müller et al. (2014) and manufactured by IONICON Analytic GmBH (Innsbruck, Austria). VOCs were sampled from outside the aircraft boundary layer using a winglet and a heated (50°C) ¼" surface-treated stainless-steel line (Sulfinert®, Restek Corporation, Bellefonte, PA, USA). During NAAMES the winglet was positioned in a downward-facing orientation at the fuselage of the C-130. The inlet flow varied with ambient pressure from 15 standard liters per minute (L·min^{-1}) at high altitudes to 30 L·min^{-1} at sea level. Reported analytes were calibrated via dynamic dilution of a certified compressed-gas standard (Apel-Riemer Environmental, Inc.), and sensitivities during NAAMES ranged from 320 cps·ppb^{-1} for methanol to 2206 cps·ppb^{-1} for methylethyl ketone. One-second concentration limits of detection were in the low 10s of ppt for most compounds, with an overall range of 9 ppt for toluene to 229 ppt for methanol. Zeroing was performed by passing ambient air through a Pt/Pd catalyst heated to 350°C (Müller et al., 2014). The instrument achieved a mass resolution ($m/\Delta m$, full-width half-maximum) of approximately 4000, with the mass axis calibrated through continuous addition of diiodobenzene (PerMassCal; IONICON Analytik GmbH, Innsbruck, Austria). The expected e-folding time (after accounting for inlet lag) for this instrument is approximately 0.1 s for most species targeted here but may be longer for sticky compounds (Müller et al., 2016).~~The instrument is described in detail by Müller et al. (2014); t~~The NAAMES measurements were conducted at 5 Hz sampling frequency with an estimated accuracy of 10 % + 5 ppt. **Table 1** lists the compounds examined in this study, which include…"

l. 110: Why were the measurements not conducted at 10 Hz?

Eddy covariance was not the original objective of the NAAMES study, and 5 Hz measurements were performed for a combination of practical reasons (e.g., file size, data load) and based on the measurement PI's prior scientific experience.

l. 110-continued: What influence on the results might this have, if any?

Changing the VOC measurement frequency to 10 Hz would not significantly affect our results, for the following reasons:

1) Based on typical eddy sizes at the flight altitudes, 5 Hz is expected to be more than sufficient to capture the relevant tracer fluctuations. For example, aircraft-based air-sea DMS and ozone flux measurements by Conley et al. (2009) (their Fig. 3) reveal negligible contributions from frequencies above 5 Hz. Similarly, Blomquist et al. (2010) found based on ship-based DMS flux measurements that there was little flux at frequencies above 3 Hz, with associated corrections of < 5 %. This expectation from prior literature is confirmed for our dataset by the well-behaved VOC-w cospectra (Fig. 3, Supplementary Spectra), which show that the dominant flux carrying scale is centered around 0.1-0.2 Hz. Increasing from 5 Hz to 10 Hz would thus not help to resolve the scales that carry the most flux energy.
2) Furthermore, the 5 Hz cospectra and ogives do not reveal any systematic high-frequency loss (Fig. 3, Supplementary Spectra), indicating that the flux contribution from additional smaller eddies observable at 10 Hz would be negligible in terms of the total flux. Specifically, across all 44 flux windows we identify only one instance where there is any indication of high-frequency flux loss (flight 20170908 L2, now noted in the manuscript).

3) 5 Hz is the highest usable frequency for the 2016 NAAMES data, when winds were also measured at this cadence.

l. 107 ff: Please report the sensitivities (cps/ppb or ncps/ppb) and limits of detection for the VOCs. This seems an important piece of information since the sensitivity and detection limit are linked to instrument signal/noise ratio, which in turn is later discussed as the most important source of uncertainty here. With this piece of information, readers may be able to estimate how well their type of PTR instrument would fare in conducting similar airborne flux measurements.

Done. Please see our response to comment l. 108 above.

l. 107 ff-continued: How were the compounds calibrated?

This information has been added to the instrument description (see response to comment l. 108).

l. 153: Are the authors sure that there is no fragment of monoterpenes on the toluene (and potentially also benzene) mass in the PTR? See Kari et al. (2018)

With the soft ionization conditions used during NAAMES, there is no significant monoterpene fragmentation to m/z 93 (used to quantify toluene) or 79 (used to quantify benzene). There is still fragmentation to m/z 81 (~50 %) and m/z 95 (minor), but this is corrected.

Fig. 2: The choice of colors is not colorblind-friendly. To help distinguish the different traces, please use different markers.

Thank you for the suggestion. We have updated Fig. 2 as requested:

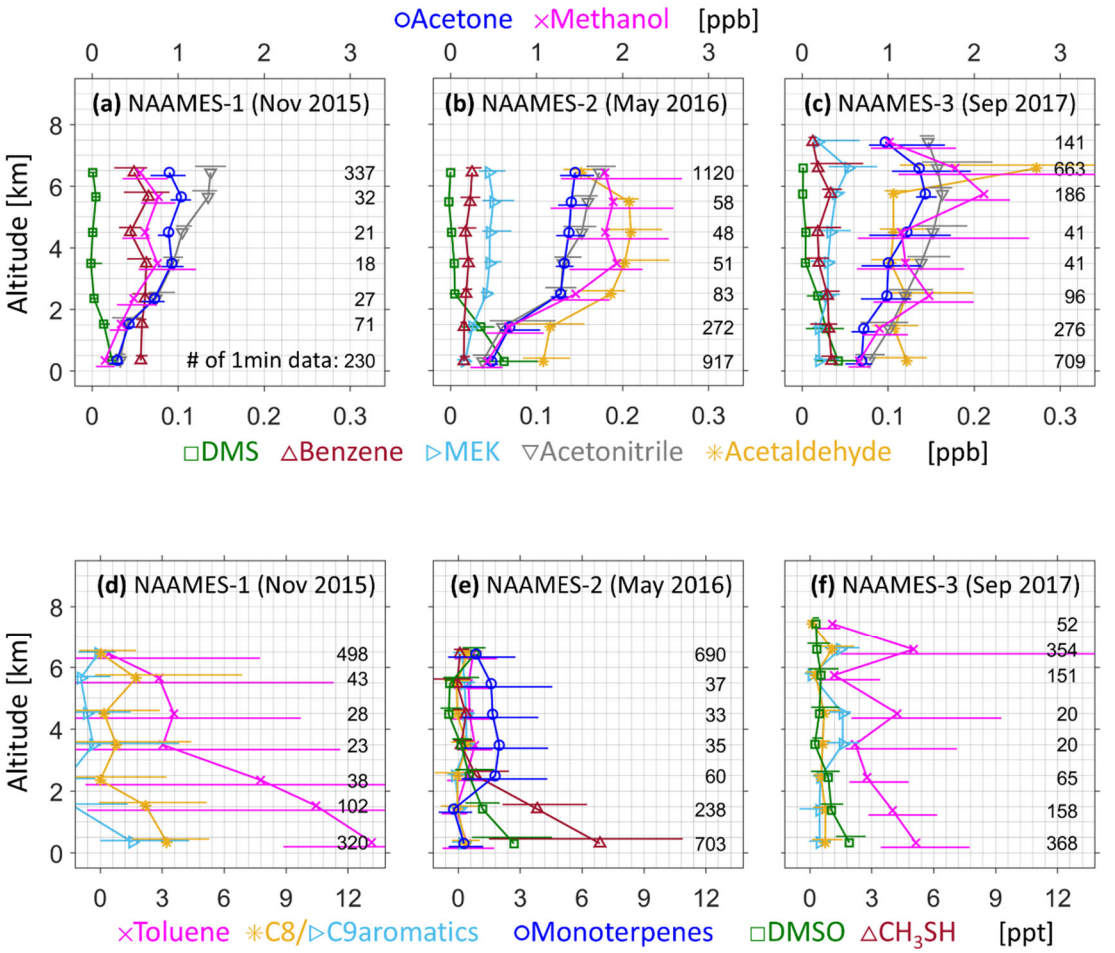

Figure 2. (updated)

Fig. 3: It took me a while to find the tiny panel labels a-j. Please consider enhancing their size.

Thank you for the suggestion. Fig. 3 has been updated as requested.

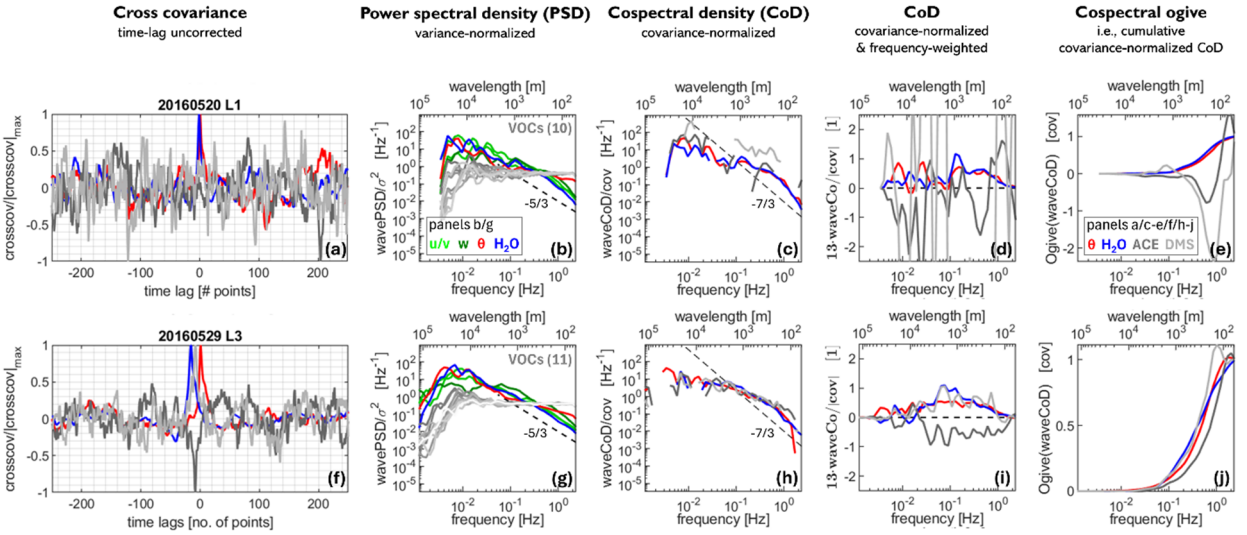

Figure 3. (updated)

l. 260: It is not clear to me how the low-frequency loss is computed / recognized. Please explain more clearly.

This point has now been clarified as follows:

"C. *Low-frequency losses*. A sampling duration that is too short to capture the largest-scale eddies will cause a systematic flux underestimate. The ogives (plateauing at low frequencies) and 95 % eddy scales (on average < 10 % of the leg length) seen during NAAMES indicate that such losses are unimportant for this study ~~VOC-w cospectra do not indicate that such impacts are important for NAAMES, and the 95 % eddy scales are substantially shorter than the leg lengths~~ (**Fig. 3**, **Supplementary Spectra, Table S1**)."

l. 287 ff: Is the method the one that is called "noise covariance method" in the Supplement of Wolfe et al. (2015)?

The Supplement of Wolfe et al. (2015) computed *total* flux random error via two methods:
- white noise covariance method
- covariance standard error method

In this work, Sect. 4.5 (starting from l. 287) involves multiple approaches that compute either *total* or *USN-induced* flux random error. Methodologies described in the Supplement of Wolfe et al. (2015) are involved in two places in this section:

- In the discussion of *total* flux random error (lines 288-294): the covariance standard error method from the Supplement of Wolfe et al. (2015) is listed as approach (iii) on line 291, with the paper properly cited in the manuscript as below.

287 **4.5 Flux random error: quantification, partitioning, and impact on signal-to-noise ratio**
288 We tested four empirical approaches for computing the total flux random error ($\sigma_{f,RE}$), based on: (i) the
289 standard deviation of the scalar-wind covariance away from the true time lag (Spirig et al., 2005;
290 Wienhold et al., 1995); (ii) a modified version of (i) that also accounts for cross-covariance offsets
291 (Langford et al., 2015); (iii) the $\overline{w's'}$ standard error (Wolfe et al., 2015); and (iv) the variance of the
292 scalar-wind covariance (Bendat and Piersol, 2010; Finkelstein and Sims, 2001). The results from these
293 approaches agree closely (R > 0.83) across NAAMES VOCs, with approach (iii) giving generally lower
294 error estimates. In what follows we report flux random errors based on the Spirig-Wienhold approach (i),

- In the discussion of *USN-induced* flux random error (lines 298-315): the white noise covariance method in the Supplement of Wolfe et al. (2015) inspires the LLW approach (lines 311-312) applied in this work; this has also been properly referenced in the manuscript as below.

311 The empirical approach outlined above combines ideas from Lenschow et al. (2000), Langford et al.
312 (2015), and Wolfe et al. (2015), and is referred to as LLW hereafter. When we compare the resulting

l. 344: Was the search window for the lag time in the non-prescribed runs limited in any way?

Yes: 0 ± 10 s. This is now clarified in the Fig. 5 caption:

"**Figure 5**. Flux histograms for the noise perturbation ensemble. H_2O fluxes from NAAMES flight 20160601 L1 are subjected to five different noise levels ($\sigma^2_{imposed}$) according to $log_{10}\left(\frac{\overline{[H_2O]}^2}{\sigma^2_{imposed}}\right) = 4.5, 3.5, 2.5, 1.5, 0.5$~~as described in text~~. Results are shown for scenarios in which the scalar-wind time lag is known (-2.2 s) and prescribed (a-e) and in which the time lag is determined independently (0 ± 10 s search window) for each flux derivation (f-j)."

l. 344-continued: Is there any risk in choosing an arbitrary window for the lag "far away" from the true lag in the Spirig/Wienhold approach? What window was chosen as the "far away" window

There are two primary considerations for this:

1) The window needs to be sufficiently far away from the true lag that the two signals are no longer correlated and the covariance solely reflects random noise. In practice this requires a greater distance for airborne EC than for surface EC, given the higher altitudes, larger eddy sizes, longer integral length scales, and broader covariance peak widths. We find that 30-50 s satisfies this requirement for our dataset given the time lags determined here (see update to time lag information below) and the peak widths in the cross-covariance plots (Fig. 3, Supplementary Spectra).

   "we therefore employ two alternative treatments as follows. (1) Flight-specific: here the time lag is defined based on the strongest cross-covariance peak for any species across the entire flight. The 13 lags calculated in this manner range from 0.4-10.8 s. (2) Species- and leg- specific: here the time lag is allowed to be species- and flight-leg specific."

2) The second consideration is that with increasing distance from the true lag, the size of the overlapping dataset used to calculate the scalar-wind covariance decreases. In an extreme case this would decrease the statistical reliability of the results.

For our case, we find that 30-50 s is an appropriate balance between considerations 1 and 2: it is far enough away from the truth to ensure that random noise dominates, while still providing a statistically robust covariance calculation across the remainder of the ~3-25 min duration (Table S1) flux legs.

Fig. 5: At least in the caption it would be helpful to mention how much noise was added in each panel, it is not reader-friendly to have to look it up in the text.

Thank you for the suggestion. We have updated the Fig. 5 caption as requested. Please see our response to l. 344.

l. 467 ff: In the discussion of the magnitudes of the fluxes, the authors are not mentioning the fact that vertical flux divergence may have caused a bias in their measurements causing underestimation of surface fluxes. Especially over the ocean with strong horizontal winds, this may be important even at the relatively low altitudes of ~140 m. In another airborne flux study of DMS over an ocean, Conley et al. (2009) showed a clear vertical divergence of the DMS flux. The authors should try to use their vertically resolved data to get some indication on whether or not vertical divergence played a role here and discuss systematic uncertainties of their reported fluxes in that regard.

We have now added a paragraph discussing the potential role of vertical flux divergence:

"5.3 Potential impact of flux divergence
Prior studies have examined the role of vertical flux divergence for VOCs in the MBL, focusing on DMS (e.g., Bandy et al., 2002; Stevens et al., 2003; Faloona et al., 2005; Conley et al., 2009) and its oxidation products (Novak et al., 2021). Processes that can lead to a change in EC-measured fluxes with height include: i) photochemical oxidation or production of a given VOC; ii) entrainment of VOC-depleted or VOC-enriched air from aloft; iii) horizontal advection / changing source footprint with height; or iv) cloud processing. The sign of the expected effect varies depending on the dominant process and on the source-sink profile for a given VOC. Flux legs analyzed here range in elevation from 126-219 m, and based on the gradients reported in the above-cited studies we expect air-sea exchange to be the predominant influence within that range. However, NAAMES did not include sufficient multi-level sampling for us to quantify the impact of vertical flux divergence across the study, and we cannot rule out its role entirely."

l. 540 ff: Are these really benzene/toluene? Please consider fragmentation as discussed by Kari et al. (2018).

See response to l. 153.

Sect. 6: Please discuss the potential impact of vertical flux divergence.

See response to l. 467 ff.

Response to Referee #2 Comments

Overview:

This is a highly technical paper that reports aircraft VOC fluxes measured by a PTR-tof-MS. The paper looks in detail at sources of random and systematic uncertainties in the flux measurements. Due to the small magnitude of the VOC fluxes, the authors found that only DMS (emission) and acetone (deposition) fluxes were unambiguously detected on a consistent basis. Overall, I think this is a useful methods paper. Though it doesn't seem that the flight plans were designed to optimize flux measurements, and the numbers of flux segments were very small, thus it's hard to say how representative the measured fluxes are of air-sea exchange in the North Atlantic.

Thank you for recognizing the uncertainty discussion of the work and for pointing out the limitations in the reported fluxes.

The reviewer is correct that the flight sampling was not designed to specifically optimize flux measurements, but it nevertheless includes numerous MBL legs that are 1) suitable for EC, and 2) of sufficient duration to capture the relevant eddy scales. As a result, NAAMES provides a useful real-world dataset for assessing measurement requirements for airborne EC in the context of air-sea VOC interactions.

We have modified Sect. 5 to make the caveat about limited representativeness more clear.

Thank you also for the thoughtful comments and constructive suggestions that follow. We have included our point-by-point responses below in blue, with the corresponding updates to the manuscript text indicated in green/quoted.

We have also made some additional minor corrections as reflected in the updated manuscript with tracked revisions. Furthermore, during the revision process we identified a few additional flight legs that were suitable for flux analysis. Since their inclusion improved our data coverage we have added them to our data ensemble used for analysis (with corresponding changes tracked). None of the main discussion points or conclusions are altered in any substantive way.

Detailed comments:

The introduction had a very comprehensive review of previous techniques.

> Thank you for this comment.

Sensitivity cps/ppb of PTR-tof-MS for main compounds of interest? What's the mass resolution?

> We have updated the manuscript (Sect. 2) to include these and other details on the deployed instrument:
>
> "VOC mole fractions were measured on-board the C-130 by proton-transfer-reaction time-of-flight mass spectrometry (PTR-ToF-MS) during NAAMES-1 (for 4 of the 7 flights), NAAMES-2 (9 of 11 flights), and NAAMES-3 (9 of 12 flights). The employed instrument is an upgraded version of the prototype PTR-ToF-MS 4000 described by Müller et al. (2014) and manufactured by IONICON Analytic GmBH (Innsbruck, Austria). VOCs were sampled from outside the aircraft boundary layer using a winglet and a heated (50°C) ¼" surface-treated stainless-steel line (Sulfinert®, Restek Corporation, Bellefonte, PA, USA). During NAAMES the winglet was positioned in a downward-facing orientation at the fuselage of the C-130. The inlet flow varied with ambient pressure from 15 standard liters per minute ($L \cdot min^{-1}$) at high altitudes to 30 $L \cdot min^{-1}$ at sea level. Reported analytes were calibrated via dynamic dilution of a certified compressed-gas standard (Apel-Riemer Environmental, Inc.), and sensitivities during NAAMES ranged from 320 $cps \cdot ppb^{-1}$ for methanol to 2206 $cps \cdot ppb^{-1}$ for methylethyl ketone. One-second concentration limits of detection were in the low 10s of ppt for most compounds, with an overall range of 9 ppt for toluene to 229 ppt for methanol. Zeroing was performed by passing ambient air through a Pt/Pd catalyst heated to

350°C (Müller et al., 2014). The instrument achieved a mass resolution ($m/\Delta m$, full-width half-maximum) of approximately 4000, with the mass axis calibrated through continuous addition of diiodobenzene (PerMassCal; IONICON Analytik GmbH, Innsbruck, Austria). The expected e-folding time (after accounting for inlet lag) for this instrument is approximately 0.1 s for most species targeted here but may be longer for sticky compounds (Müller et al., 2016).~~The instrument is described in detail by Müller et al. (2014); t~~The NAAMES measurements were conducted at 5 Hz sampling frequency with an estimated accuracy of 10 % + 5 ppt. **Table 1** lists the compounds examined in this study, which include…"

Line 105. Altitude of low level legs?

The altitudes of low-level legs are summarized in Sect. 4.1 with details listed in Table S1. We have added text at the location mentioned by the reviewer to direct the interested reader to these specifics:

"The C-130 flights included multiple low-level MBL legs (detailed in Sect. 4.1) ~~within the MBL~~, satisfying one of the prerequisites for deriving air-sea fluxes by airborne EC."

We have also made some corrections to the content in Sect. 4.1, accounting for the additional flux legs noted above:

"**Table S1** lists the 44~~34~~ flux legs from 13~~11~~ NAAMES-2 and NAAMES-3 flights identified in this way, which average 9~~8~~ min (~72~~64~~ km) in duration and range from 3-25 min (~20~~22~~-210~~194~~ km) at mean sampling altitudes of 126-219 m. Mean airspeeds range from 111~~102~~-157 m·s^{-1} across the flux legs, corresponding to approximately one 5 Hz data point per 22~~20~~-31 m."

Line 110. This might have been described briefly, but is worth repeating.

Was the PTR calibrated against standard gases?

VOCs are calibrated using a multiple component calibration gas standard (Apel-Riemer Environmental, Inc.) that was dynamically diluted. We have now added this information to the instrument description (see our reply to this reviewer's prior question about instrument specifications).

What was used as the background or blank measurement?

Zeroing was performed by passing ambient air through a heated Pt/Pd catalyst held at 350°C. We have now added this information to the instrument description (see our reply to this reviewer's prior question about instrument specifications).

Type/length of inlet?

VOCs were sampled with a winglet and a heated surface-treated stainless steel line. Inlet details have now been added to the instrument description (see our reply to this reviewer's prior question about instrument specifications).

Flow rate?

The inlet flows varied with ambient pressure 15 to 30 L·min^{-1}. We have now added flow information to the instrument description (see our reply to this reviewer's prior question about instrument specifications).

What's the expected e-folding time (response time) of the system?

The e-folding response time of the complete inlet/PTR-ToF-MS system was not specifically measured for this study. It was reported previously by Müller et al. (2016) to be approximately 0.1 s for most species targeted here but longer for sticky compounds. The

Müller et al. (2016) value of 0.1 s is in-line with our expectations for NAAMES based on the volume exchange time of the inlet and PTR-ToF-MS (both ~0.1 s). We have added this additional context to the instrument description in Sect. 2:

"The expected e-folding time (after accounting for inlet lag) for this instrument is approximately 0.1 s for most species targeted here but may be longer for sticky compounds (Müller et al., 2016)."

And to the discussion of results for DMSO and C9-aromatics, which are the analyzed species most likely to be impacted by such effects:

"• DMSO/C9 aromatics. To our knowledge there have not been any reported air-sea exchange rates for these species. Their fluxes were undetectable during NAAMES, and the uncertainty range for the campaign-mean results suggests that any air-sea exchange that did occur had a magnitude less than approximately 0.1 μmol·m^{-2}·d^{-1} (Table 1). However, it is possible that wall interactions may have inhibited flux detectability for these lower-volatility species."

Line 135. Was a correction for the aircraft motion on the measured relative winds necessary for these flights?

All corrections from calibration maneuvers that are described in detail in Thornhill et al. (2003) are applied to the full Lenschow (1986) equations and those values are based on statistics collected from multiple calibration maneuvers that were performed throughout the multi-year NAAMES EVS campaign. No additional corrections were required for the aircraft wind to force any segments of the data to be ideally 0.0 m·s^{-1}. We have updated the text to provide more detail on this point:

"Three-dimensional winds are computed from the full air motion equations (Lenschow, 1986) -and corrected for aircraft motion as described by Thornhill et al. (2003) based on statistics derived from calibration maneuvers performed throughout the multi-year NAAMES EVS campaign. No additional corrections were required, and the uncertainty in the resulting wind fields is estimated at 10 %~~with uncertainty estimated at 10%~~."

Figure 2. please use different marker shapes for people who might struggle with the colors.

Thank you for the suggestion. We have updated Fig. 2 as requested:

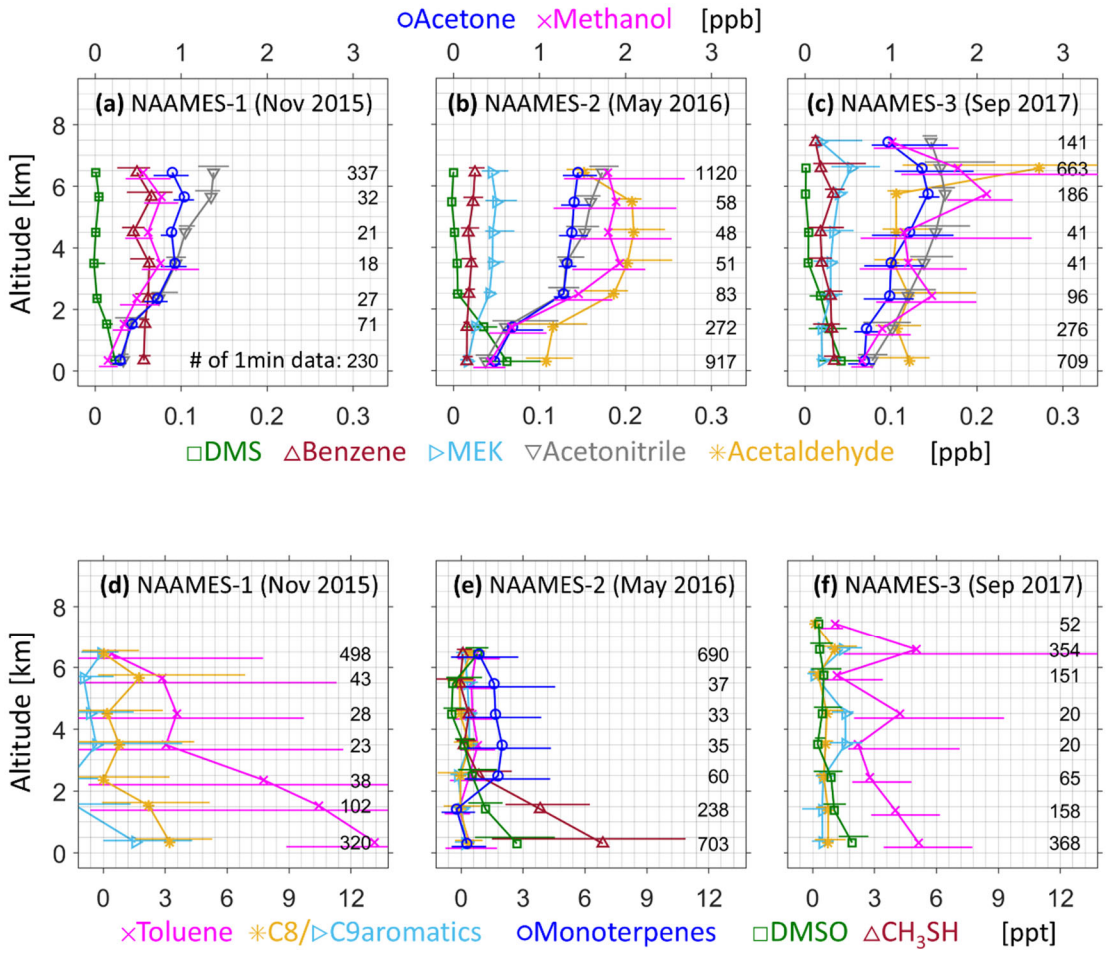

Figure 2. (updated)

Line 176. Random uncertainty in flux can usually be reduced by more averaging (scales with 1/sqrt(N)). Here a rather large range in averaging time (and so space) was chosen: 3-25 km, which by itself preassembly leads to quite different noise levels. Why not just using 3 km averaging period through out? And if needed, further average the 3-km fluxes afterwards?

Sampling lengths need to be long enough to capture all relevant eddy scales. The 95 % eddy scales summarized in Table S1 range from 1-30 km, and show that the leg lengths as used here meet this requirement but a uniform 3 km length would not.

Line 186. How different are flux results using this approach vs a simple linear detrend?

Although the two detrending approaches result in highly consistent flux estimates (Fig. R2-1 panels a, b), the 100 s moving average approach is better at filtering out extremely large-scale motions (panel d) than is linear detrending (panel c). But in any event, the overall flux agreement indicates that these large scale motions at the low-frequency tails are not contributing appreciably to the derived values.

We have clarified this in the manuscript:
"The resulting data subsets undergo mean-removal and detrending to isolate the scalar concentration fluctuations that are driven by turbulence. We employ 100 s (11-16 km) moving-average mean removal; as shown in the **Supplementary Spectra** this acts as a high-pass filter to remove low-frequency, non-turbulent components (e.g., caused by air

mass variability) that might not be excluded with a simple linear detrend (Novak et al., 2021; Moncrieff et al., 2005)."

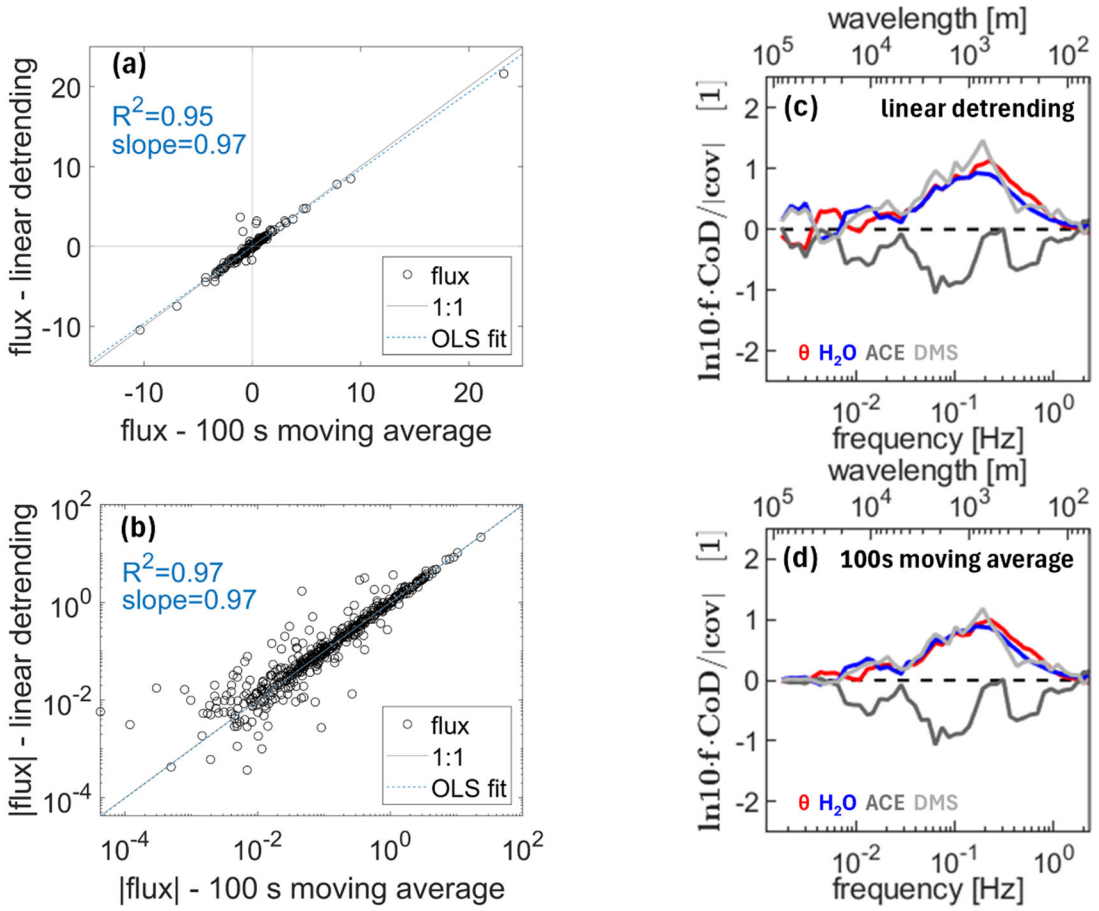

Figure R2-1. Comparison of VOC fluxes derived using 100 s moving average detrending vs. linear detrending on linear (a) and log (b) scales. Example covariance-normalized and frequency-weighted cospectra from flight 20160528 L2 with linear detrending (c) and 100 s moving average detrending (d).

Line 190. What is roughly the expected lag time based on the inlet dimensions and flow rates?

The total inlet length was approximately 3.3 m, corresponding to an internal volume of 26 cm^3. With flow rates of 15-30 $L \cdot min^{-1}$ this corresponds to an expected physical lag time of 0.1 s or less. The lag times derived via cross-covariance analysis are frequently greater than this (up to 11.8 s, depending on the flight) and that is because this derivation is also accounting for clock offsets between the PTR-ToF-MS and wind measurement systems. There was not a common time server used on these flights and therefore there were minor differences between the instrument time stamps that varied by flight. We have clarified this point in the revised manuscript:

"In multiple cases the cross-covariance for an individual flux leg lacks a sufficiently clear peak for time-lag diagnosis (**Fig. 3**), and we therefore employ two alternative treatments as follows. (1) Flight-specific: here the time lag is defined based on the strongest cross-covariance peak for any species across the entire flight. The 13 lags calculated in this manner range from 0.4-10.8 s. (2) Species- and leg- specific: here the time lag is allowed to be species- and flight-leg specific. For both treatments, the derived ~~The latter treatment~~

~~yields an ensemble of time lags that range from 0-2.8 s or from 6.2-11.8 s, depending on the flight; for cases without a clear and physically reasonable peak the lag is set to the mean of the within-flight ensemble.~~Derived ~~-~~lag times ~~of up to 10.8 s~~ primarily reflect clock offsets between the PTR-ToF-MS and wind measurement systems, which varied by flight; the physical lag time expected for the 3.3 m (0.3175 cm ID) inlet with flows exceeding 15 $L \cdot min^{-1}$ is < 0.1 s.

Treatment 1 above assumes that the targeted VOCs interact similarly with the sampling inlet and instrument, and that inter-species lag differences mainly reflect statistical artifacts. Fluxes derived in this way are conservative relative to those obtained with treatment 2, which (in addition to clock synchronization) allows for species-specific sampling effects and tends to derive the largest possible flux magnitude. In the following analysis we use treatment 1 (a single flight-specific lag applied across species) as default ~~unless otherwise noted~~."

Line 190-continued. Have authors tested with injections of VOC standards to see whether different VOC show comparable or different responses within their inlet?

We addressed this point in our reply to the reviewer's related question above("What's the expected e-folding time (response time) of the system?"), please see that response for details.

Figure 3. on my screen there looks to be breaks in the cospectral density plots?

Thank you for capturing this. This is now clarified in the caption:

"**Figure 3.** Example *s*-*w* cross-covariances (panels a, f) and wavelet-based spectral analyses (b-e, g-j). The top row shows an example flux leg in which clear cross-covariance peaks and well-behaved cospectra/ogives are obtained for water vapor (H_2O) and potential temperature (θ) but not for VOCs. The bottom row shows an example leg with clear cross-covariance peaks and reasonably well behaved cospectra/ogives for all plotted scalars. The impact of white noise on the VOC measurements is evident in the high-frequency range of the power spectra in panels (b) and (g). The wavelength scale is computed as the aircraft ground speed divided by frequency, without any height scaling. Line breaks in panel (c) occur where the flux power is in the opposite direction to that of the overall (net) flux value."

Line 254. Again, what's the response time of the PTR/inlet system?

This has been added. Please see prior responses.

Line 254-continued. It might be that the cospectra are too noisy to say with confidence that high frequency attenuation is negligible.

We agree and have now clarified this point:

"*B. High-frequency attenuation*. Loss of high-frequency flux contributions can arise from an insufficient sampling frequency, inadequate instrument time response, or inlet-line damping. If uncorrected, the result is a systematic flux underestimate. The NAAMES VOC-*w* cospectra~~, while noisy,~~ do not show ~~clear~~ systematic evidence of such an effect when compared to those for the non-VOC scalars (**Fig. 3**, **Supplementary Spectra**). Specifically, across all 44 flux windows we identify only one instance where there is any indication of high-frequency flux loss (flight 20170908 L2). It is possible that the instrument noise is obscuring ~~some~~ attenuation in some of the other cases, but since there is no direct evidence of a systematic impact w~~W~~e ~~therefore do not~~ perform ~~any~~ no high-frequency correction ~~for NAAMES, with the expectation~~and expect that other error sources are more important."

Line 254-continued. Were there conditions of stable atmosphere, where attenuation might be more pronounced?

Thank you for the suggestion. As examined in Fig. R2-2, for NAAMES airborne fluxes analyzed in this work, we do not see clear evidence of the level of attenuation as a function of atmospheric stability.

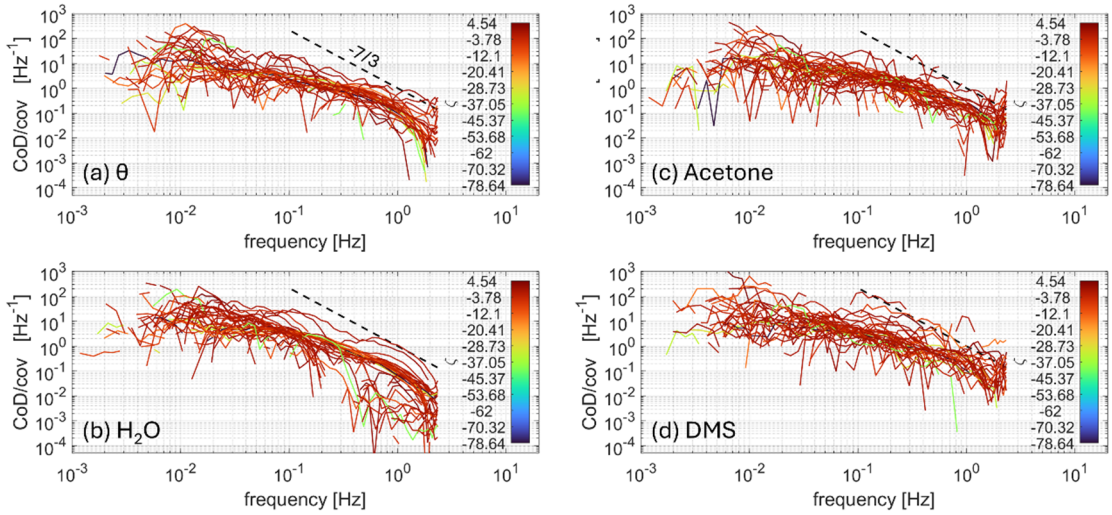

Figure R2-2. NAAMES cospectra (covariance-normalized) colored by flight altitude atmospheric stability. The stability indicator is calculated as $\zeta = \frac{z}{-\frac{g}{\kappa \overline{\theta} \overline{w'\theta'}}(\overline{w'u'}^2+\overline{w'v'}^2)^{0.75}}$, where $\kappa = 0.41$, $g = 9.81\ m \cdot s^{-2}$, and θ, $\overline{w'u'}$, $\overline{w'v'}$,$\overline{w'\theta'}$ are obtained at flux leg altitude z.

Line 260. The detrend method used implicitly assumes that there are no low frequency eddies that contribute to flux, right?

To the contrary, we use this detrending method specifically to remove low frequency, non-turbulent components. Please see our response to Line 186 for details.

Line 260-continued. Again, a simple comparison with this approach vs linear detrend would be useful.

Done. See response to Line 186.

Line 314. The use of 'VOCs' here is ambiguous. Surely the random error contribution to well resolved fluxes such as DMS and maybe acetone is smaller than for VOCs such as benzene and monoterpenes?

The main purpose of this passage is to compare behavior between VOCs and θ / H_2O. It is correct that VOCs with better-resolved fluxes have a smaller error contribution from USN.

For example, we calculate mean ± SD of $\frac{\sigma_{f,USN}^2}{\sigma_{f,RE}^2}$ values as follows:

Acetone: 66% ± 31% (total number of fluxes: 44)
DMS: 72% ± 35% (44)
Benzene: ~100% (37)
Monoterpenes: ~100% (20)

However, the overall point still stands: the VOC flux errors are in general USN-dominated whereas flux errors for θ and H_2O are in general turbulence-dominated. We have updated the text to convey this point:

"The empirical approach outlined above combines ideas from Lenschow et al. (2000), Langford et al. (2015), and Wolfe et al. (2015), and is referred to as LLW hereafter. When

we compare the resulting $\sigma_{f,USN}$ estimates to the $\sigma_{f,RE}$ values obtained earlier we find that the total flux random error is dominated by the USN contribution for VOCs (92~~89~~ % of $\frac{\sigma_{f,USN}^2}{\sigma_{f,RE}^2}$ ratios exceed 50 %, with smaller ratios for species with better-resolved fluxes) but by turbulent stochasticity for θ and H_2O. This flux-based finding aligns with the concentration-based autocovariance results above and shows that the VOC flux measurements are occurring in a noise-limited regime. The situation therefore differs from that encountered for air-sea flux measurements of CO_2, where sampling uncertainties are often the principal random error source (Dong et al., 2021)."

Figure 5. This is a neat way to illustrate the importance of getting the lag time correctly. But I don't get why such a large lag wind window (+/- 50s) is needed for the lag correlation calculation. Was the pump flow so variable (e.g. with altitude)?

In fact a search window of 0 ± 10 s is applied in the non-prescribed runs, and this is now clarified in the Fig. 5 caption.

"**Figure 5**. Flux histograms for the noise perturbation ensemble. H_2O fluxes from NAAMES flight 20160601 L1 are subjected to five different noise levels ($\sigma_{imposed}^2$) according to $log_{10}\left(\frac{\overline{[H_2O]^2}}{\sigma_{imposed}^2}\right) = 4.5, 3.5, 2.5, 1.5, 0.5$ ~~as described in text~~. Results are shown for scenarios in which the scalar-wind time lag is known (-2.2 s) and prescribed (a-e) and in which the time lag is determined independently (0 ± 10 s search window) for each flux derivation (f-j)."

Figure 7. Again, I think it'd be useful to specify which VOC here, as the S/N must be very different for say DMS vs benzene.

There is insufficient space in Fig. 7 to label all 13 VOCs due to their significant overlap. However, to provide the requested context on this point we calculate examples of mean (± SD) SNR_f: 3.35±1.77 (acetone), 2.28±2.14 (DMS), 1.45±1.31 (benzene), 0.53±0.51 (monoterpenes), and have added two of these examples to the accompanying text:

"The results in **Fig. 7** provide a framework that can be used to assess whether an instrument's noise performance is sufficient for a given flux application and desired SNR_f. That is, for an expected flux magnitude, the plotted curves can be used to identify the maximum $\sigma_{s,USN}$ that is likely to yield an appropriate SNR_f under the NAAMES sampling conditions. For the 11~~ten~~ actual flux legs examined in **Fig. 7**, 17~~15~~ % of the VOC fluxes have $SNR_f > 3$; for example, the mean (± SD) SNR_f value for acetone is 3.35±1.77 but just 0.53±0.51 for monoterpenes. W~~w~~e find that if USN was lowered by a factor of 7-23~~18~~, 75 % of these VOC fluxes would exceed SNR_f=3. A 10-20× advance in instrumental noise performance compared to the PTR-ToF-MS deployed during NAAMES would therefore provide significant benefits in our ability to quantify ocean-air VOC fluxes."

For further clarification, in Figure 7 we now also color the VOC datapoints according to the same scheme used for the other plot elements, indicating their noise-limited, transitional, or turbulence-limited status:

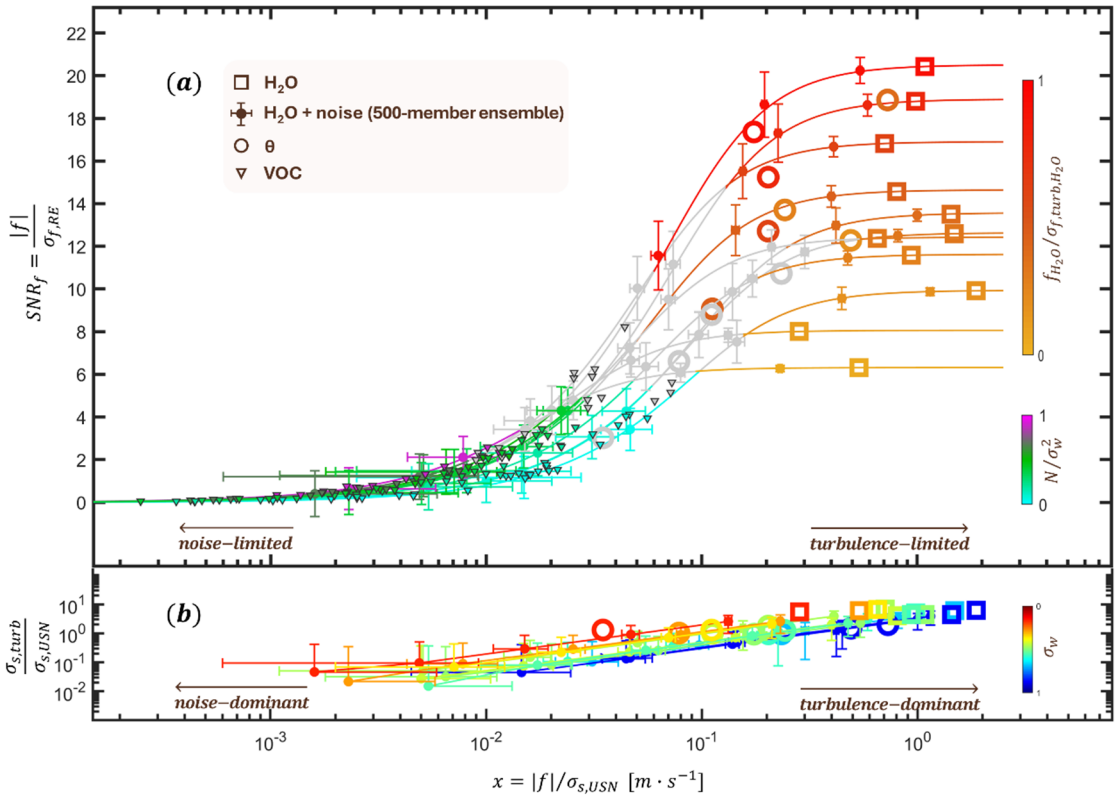

Figure 7. (updated)

Table 1. where it says 'uncertainty of the mean flux', this seems to be 3 stdev. Is this the flux noise or the flux detection limit? It seems like it might be the latter to me.

The 'Uncertainty of the mean flux (3σ)' included in Table 1 is propagated from the individual flux random errors (3σ), through $\sigma_{mean} = (1/N)\sqrt{\sigma_{f,RE,1}^2 + \sigma_{f,RE,2}^2 + \cdots}$, to give an estimate of the uncertainty in the campaign-mean flux. The flux limit-of-detection is discussed elsewhere in the paper in the context of individual (rather than averaged) flux measurements and so we do not use that term here.

Table 1-continued. How are error-weighted mean flux and its uncertainty computed?

These are computed following methodologies outlined in Taylor (1997). We have updated these citations to refer to the specific chapter in each case:

"**Table 1.** Summary of airborne VOC flux measurements during NAAMES.

VOC	Ion formula	*m/z*	Number of flux legs	Flux range	Campaign mean flux[a]	Uncertainty of the mean flux[b]	Error-weighted campaign mean flux[c]	Uncertainty of the weighted mean flux[bd]	Above-LOD flux observations[dc]
				(μmol·m⁻²·d⁻¹)					upward(↑) / downward(↓)
Acetone	$C_3H_6O{\cdot}H^+$	59.049	*31*	[−25.5, +3.7]	−4.7	*1.3*	-2.5	*0.55*	0↑ / 16↓
Dimethyl sulfide (DMS)	$C_2H_6S{\cdot}H^+$	63.026	*29*	[−1.3, +5.0]	+1.4	*0.65*	+0.52	*0.19*	7↑ / 0↓
Methanol	$CH_4O{\cdot}H^+$	33.033	*29*	[−38.1, +85.4]	+0.10	*4.3*	-2.6	*1.9*	1↑ / 2↓
Methylethyl ketone (MEK)	$C_4H_8O{\cdot}H^+$	73.065	*30*	[−2.1, +11.4]	+0.56	*0.38*	+0.11	*0.18*	5↑ / 1↓
Acetaldehyde	$C_2H_4O{\cdot}H^+$	45.033	*28*	[−5.8, +33.4]	+2.9	*1.5*	+0.51	*0.66*	4↑ / 2↓
Acetonitrile	$C_2H_3N{\cdot}H^+$	42.034	*28*	[−2.9, +5.1]	+0.41	*0.61*	+0.080	*0.21*	2↑ / 1↓

Benzene	$C_6H_6 \cdot H^+$	79.054	*23*	[−0.7, +2.0]	+0.31	*0.36*	+0.19	*0.15*	2↑ / 0↓
Methanethiol	$CH_4S \cdot H^+$	49.011	*11*	[−0.3, +1.4]	+0.26	*0.40*	+0.13	*0.12*	0↑ / 0↓
Dimethyl sulfoxide (DMSO)	$C_2H_6SO \cdot H^+$	79.021	*28*	[−0.8, +0.6]	+0.0038	*0.11*	+0.00013	*0.034*	0↑ / 0↓
Toluene	$C_7H_8 \cdot H^+$	93.070	*28*	[−0.9, +0.6]	+0.0066	*0.16*	+0.071	*0.078*	0↑ / 0↓
Monoterpenes	$C_{10}H_{16} \cdot H^+$	137.132	*12*	[−0.4, +0.5]	−0.028	*0.36*	+0.028	*0.16*	0↑ / 0↓
C8 aromatics	$C_8H_{10} \cdot H^+$	107.086	*13*	[−0.5, +0.4]	+0.0087	*0.20*	+0.028	*0.076*	0↑ / 0↓
C9 aromatics	$C_9H_{12} \cdot H^+$	121.101	*11*	[−0.1, +0.4]	+0.079	*0.12*	+0.056	*0.078*	0↑ / 0↓

[a,c]Includes below-LOD fluxes.

[b,d]3σ uncertainty.

[b]Ce~~a~~lculated via error propagation (Taylor, 1997; see Ch. 3).

[c,d]~~Includes below-LOD fluxes;~~ Ce~~o~~mputed following Taylor (1997; see Ch.7).

[d~~e~~]Number of flux measurements that exceed the corresponding observation-specific flux LOD (3σ)."

Section 5.1 This section is interpreted in the context of the expected air-sea fluxes. However the authors never mentioned the height of their 'low level legs'.

In fact this information is provided in Sect. 4.1 and in Table S1. See our response to Line 105.

Section 5.1-continued. EC fluxes are generally thought to be representative of surface fluxes if they're measured within the surface layer of the atmosphere, which is roughly lowest 10% of the marine boundary layer. Thus if the boundary layer is 500 m, the aircraft would need to fly below 50 m for the flues to be representative of surface exchange. Above the surface layer there's expected to be a vertical gradient in flux (driven by the strength of surface flux and entrainment flux at top of the marine boundary layer).

We have now added a paragraph discussing the potential role of vertical flux divergence:

"5.3 Potential impact of flux divergence
Prior studies have examined the role of vertical flux divergence for VOCs in the MBL, focusing on DMS (e.g., Bandy et al., 2002; Stevens et al., 2003; Faloona et al., 2005; Conley et al., 2009) and its oxidation products (Novak et al., 2021). Processes that can lead to a change in EC-measured fluxes with height include: i) photochemical oxidation or production of a given VOC; ii) entrainment of VOC-depleted or VOC-enriched air from aloft; iii) horizontal advection / changing source footprint with height; or iv) cloud processing. The sign of the expected effect varies depending on the dominant process and on the source-sink profile for a given VOC. Flux legs analyzed here range in elevation from 126-219 m, and based on the gradients reported in the above-cited studies we expect air-sea exchange to be the predominant influence within that range. However, NAAMES did not include sufficient multi-level sampling for us to quantify the impact of vertical flux divergence across the study, and we cannot rule out its role entirely."

Section 5.2 Eddy covariance fluxes, as the authors have shown, have relatively large random uncertainties that require sufficient averaging to draw robust conclusions. I'm not saying that the conclusions in this section is wrong, but given the very low number of above LOD flux segments for these VOCs, it's very difficult to say how representative the aircraft fluxes are. I'd be tempted to minimize/remove this section. But if the authors want to keep it in, put a disclaimer at the beginning of the section about how limited the data coverage is.

Thank you for the suggestion. We have added a disclaimer about representativeness to the opening paragraphs of Sect. 5.

"Across VOCs, acetone and DMS have the greatest number of flux measurements with magnitudes exceeding the corresponding observation-specific LOD (**Table 1**). The fraction of above-detection fluxes for other species ranges from 0-21 %. However, analyses above showed that the NAAMES VOC flux errors and LODs are dominated by random USN effects, and it follows that these uncertainties are reduced through averaging. Accordingly, we see in

Table 1 that in some cases the magnitude of the campaign-mean flux (and/or the error-weighted mean flux) exceeds its uncertainty level—even with just a few above-detection flux measurements. These compound-specific results are discussed further in the next section. In all cases the limited number of flux observations and the regional sampling during NAAMES mean that results need to be interpreted with caution in terms of their broader spatial/temporal representativeness."

Note: I co-reviewed this paper with my PhD student, Irene Monreal-Campos

References

Bandy, A. R., Thornton, D. C., Tu, F. H., Blomquist, B. W., Nadler, W., Mitchell, G. M., and Lenschow, D. H.: Determination of the vertical flux of dimethyl sulfide by eddy correlation and atmospheric pressure ionization mass spectrometry (APIMS), J. Geophys. Res.: Atmos., 107, https://doi.org/10.1029/2002JD002472, 2002.

Blomquist, B. W., Huebert, B. J., Fairall, C. W., and Faloona, I. C.: Determining the sea-air flux of dimethylsulfide by eddy correlation using mass spectrometry, Atmos. Meas. Tech., 3, 1-20, https://doi.org/10.5194/amt-3-1-2010, 2010.

Conley, S. A., Faloona, I., Miller, G. H., Lenschow, D. H., Blomquist, B., and Bandy, A.: Closing the dimethyl sulfide budget in the tropical marine boundary layer during the Pacific Atmospheric Sulfur Experiment, Atmos. Chem. Phys., 9, 8745-8756, https://doi.org/10.5194/acp-9-8745-2009, 2009.

Dong, Y., Yang, M., Bakker, D. C. E., Kitidis, V., and Bell, T. G.: Uncertainties in eddy covariance air-sea CO flux measurements and implications for gas transfer velocity parameterisations, Atmos. Chem. Phys., 21, 8089-8110, https://doi.org/10.5194/acp-21-8089-2021, 2021.

Faloona, I., Lenschow, D. H., Campos, T., Stevens, B., van Zanten, M., Blomquist, B., Thornton, D., Bandy, A., and Gerber, H.: Observations of entrainment in eastern Pacific marine stratocumulus using three conserved scalars, J. Atmos. Sci., 62, 3268-3285, https://doi.org/10.1175/JAS3541.1, 2005.

Kari, E., Miettinen, P., Yli-Pirilä, P., Virtanen, A., and Faiola, C. L.: PTR-ToF-MS product ion distributions and humidity-dependence of biogenic volatile organic compounds, Int. J. Mass Spectrom., 430, 87-97, https://doi.org/10.1016/j.ijms.2018.05.003, 2018.

Langford, B., Acton, W., Ammann, C., Valach, A., and Nemitz, E.: Eddy-covariance data with low signal-to-noise ratio: time-lag determination, uncertainties and limit of detection, Atmos. Meas. Tech., 8, 4197-4213, https://doi.org/10.5194/amt-8-4197-2015, 2015.

Lenschow, D. H.: Probing the Atmospheric Boundary Layer, Am. Meteorol. Soc., Boston, Mass., edited by: Lenschow, D. H., 1986.

Lenschow, D. H., Wulfmeyer, V., and Senff, C.: Measuring second- through fourth-order moments in noisy data, J. Atmos. Ocean. Technol., 17, 1330-1347, https://doi.org/10.1175/1520-0426(2000)017<1330:MSTFOM>2.0.CO;2, 2000.

Moncrieff, J., Clement, R., Finnigan, J., and Meyers, T.: Chapter 2. Averaging, Detrending, and Filtering of Eddy Covariance Time Series, Handbook of micrometeorology: A guide for surface flux measurement and analysis, edited by: Lee, X., Massman, W., and Law, B., Springer Netherlands, Dordrecht, 7-31 pp., 2005.

Müller, M., Anderson, B. E., Beyersdorf, A. J., Crawford, J. H., Diskin, G. S., Eichler, P., Fried, A., Keutsch, F. N., Mikoviny, T., Thornhill, K. L., Walega, J. G., Weinheimer, A. J., Yang, M., Yokelson, R. J., and Wisthaler, A.: In situ measurements and modeling of reactive trace gases in a small biomass burning plume, Atmos. Chem. Phys., 16, 3813-3824, https://doi.org/10.5194/acp-16-3813-2016, 2016.

Müller, M., Mikoviny, T., Feil, S., Haidacher, S., Hanel, G., Hartungen, E., Jordan, A., Märk, L., Mutschlechner, P., Schottkowsky, R., Sulzer, P., Crawford, J. H., and Wisthaler, A.: A compact PTR-ToF-MS instrument for airborne measurements of volatile organic compounds at high spatiotemporal resolution, Atmos. Meas. Tech., 7, 3763-3772, https://doi.org/10.5194/amt-7-3763-2014, 2014.

Novak, G. A., Fite, C. H., Holmes, C. D., Veres, P. R., Neuman, J. A., Faloona, I., Thornton, J. A., Wolfe, G. M., Vermeuel, M. P., Jernigan, C. M., Peischl, J., Ryerson, T. B., Thompson, C. R., Bourgeois, I., Warneke, C., Gkatzelis, G. I., Coggon, M. M., Sekimoto, K., Bui, T. P., Dean-Day, J., Diskin, G. S., DiGangi, J. P., Nowak, J. B., Moore, R. H., Wiggins, E. B., Winstead, E. L., Robinson, C., Thornhill, K. L., Sanchez, K. J., Hall, S. R., Ullmann, K., Dollner, M., Weinzierl, B., Blake, D. R., and Bertram, T. H.: Rapid cloud removal of dimethyl sulfide oxidation products limits SO2 and cloud condensation nuclei production in the marine atmosphere, Proc. Natl. Acad. Sci. U. S. A., 118, https://doi.org/10.1073/pnas.2110472118, 2021.

Stevens, B., Lenschow, D. H., Faloona, I., Moeng, C.-H., Lilly, D. K., Blomquist, B., Vali, G., Bandy, A., Campos, T., Gerber, H., Haimov, S., Morley, B., and Thornton, D.: On entrainment rates in nocturnal marine stratocumulus, Q. J. Roy. Meteor. Soc., 129, 3469-3493, https://doi.org/10.1256/qj.02.202, 2003.

Taylor, J. R.: An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2nd edition, 1997.

Thornhill, K. L., Anderson, B. E., Barrick, J. D. W., Bagwell, D. R., Friesen, R., and Lenschow, D. H.: Air motion intercomparison flights during Transport and Chemical Evolution in the Pacific (TRACE-P)/ACE-ASIA, J. Geophys. Res.: Atmos., 108, https://doi.org/10.1029/2002JD003108, 2003.

Wolfe, G. M., Hanisco, T. F., Arkinson, H. L., Bui, T. P., Crounse, J. D., Dean-Day, J., Goldstein, A., Guenther, A., Hall, S. R., Huey, G., Jacob, D. J., Karl, T., Kim, P. S., Liu, X., Marvin, M. R., Mikoviny, T., Misztal, P. K., Nguyen, T. B., Peischl, J., Pollack, I., Ryerson, T., St. Clair, J. M., Teng, A., Travis, K. R., Ullmann, K., Wennberg, P. O., and Wisthaler, A.: Quantifying sources and sinks of reactive gases in the lower atmosphere using airborne flux observations, Geophys. Res. Lett., 42, 8231-8240, https://doi.org/10.1002/2015GL065839, 2015.