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, without influence on any discussion points or conclusions.)

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 minor corrections to the content in Sect. 4.1 to match the details in Table S1:

"**Table S1** lists the 34 flux legs from 11 NAAMES-2 and NAAMES-3 flights identified in this way, which average 9~~8~~ min (~70~~64~~ km) in duration and range from 3-25 min (~20~~2~~-210~~194~~ km) at mean sampling altitudes of 126-219 m. Mean airspeeds range from 111~~02~~-157 m·s^{-1} across the flux legs, corresponding to approximately one 5 Hz data point per 22~~0~~-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 significant 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.
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 or 79. 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:

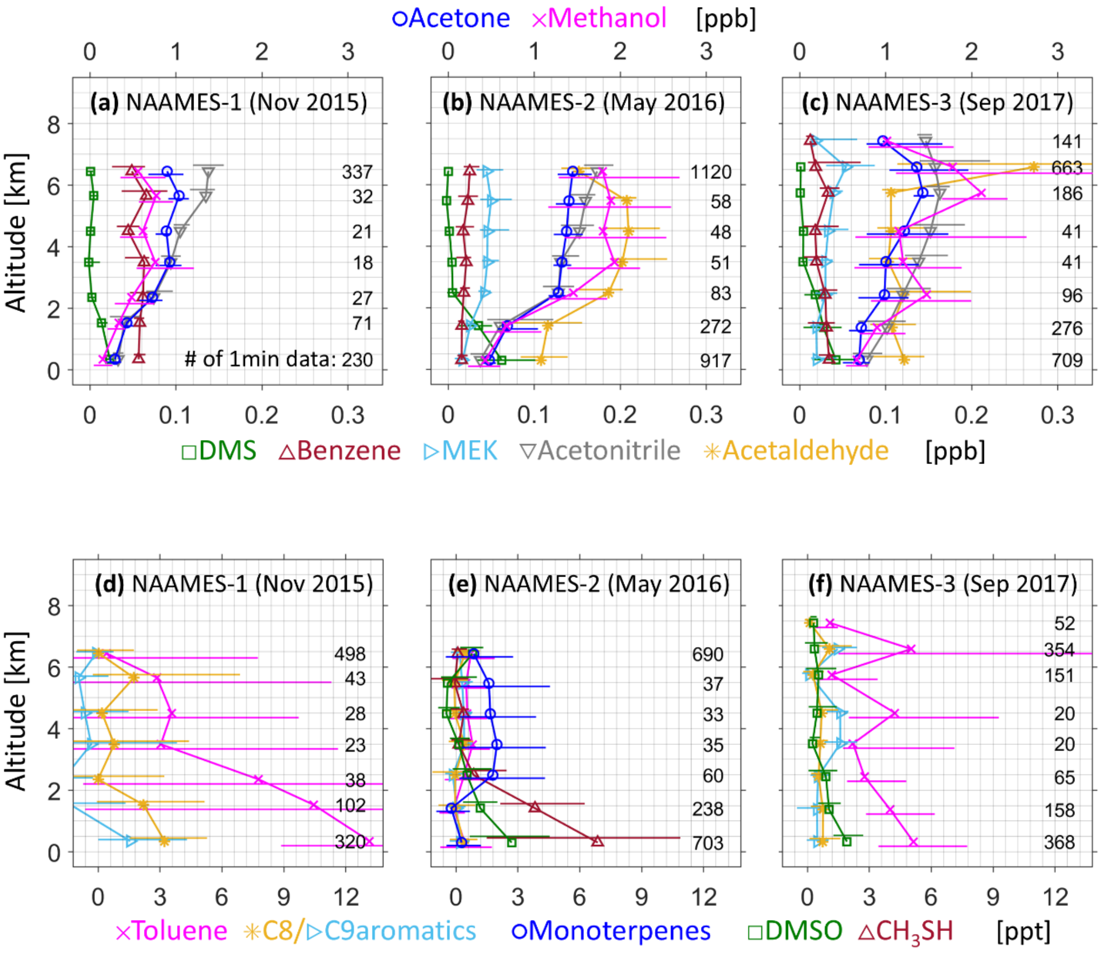


Figure 2. (updated)

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.

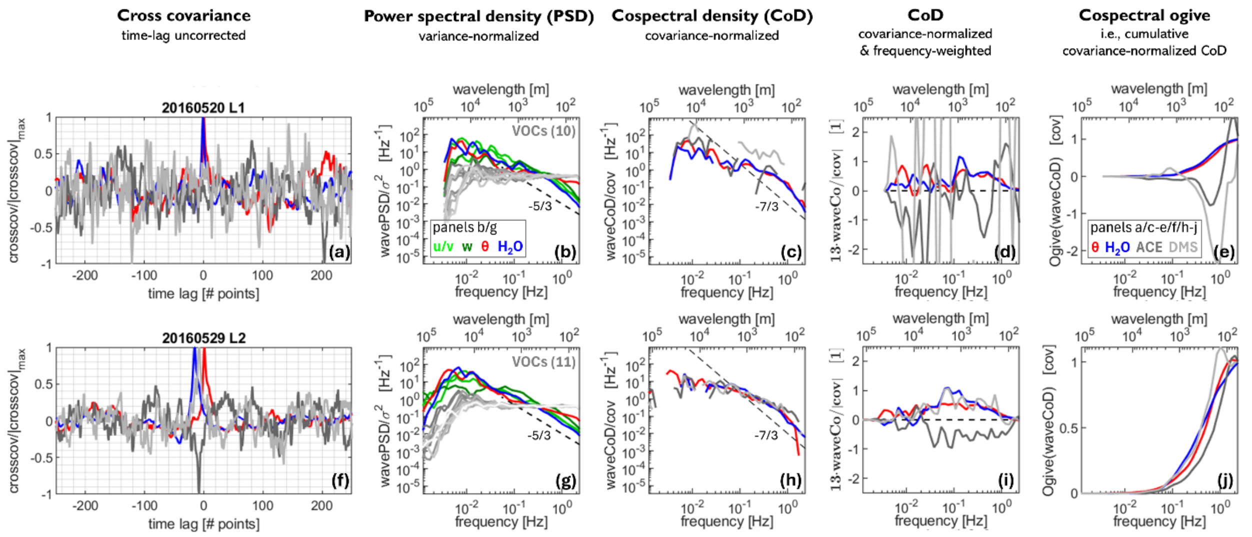


Figure 3. (updated)

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**)."

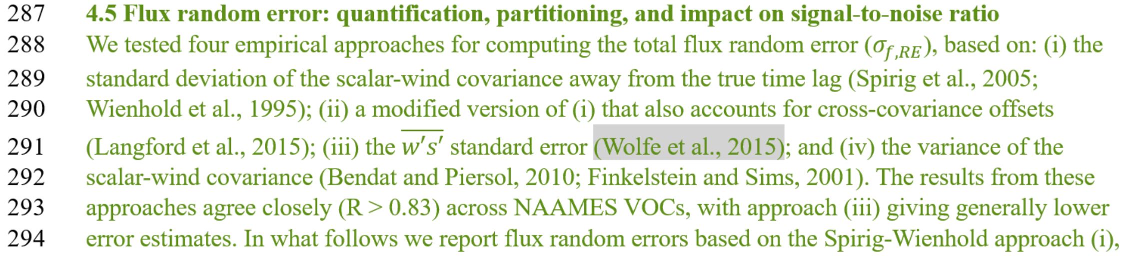
l. 287 ff: Is the method the one that is called "noise covariance method" in the Supplement of Wolfe et al. (2015)?

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 $log_{10}\left(\frac{[H_2O]^2}{\sigma^2_{imposed}}\right) = 4.5, 3.5, 2.5, 1.5, 0.5$~~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 11 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. This second 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."

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.

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.

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, International Journal of Mass Spectrometry, 430, 87-97, https://doi.org/10.1016/j.ijms.2018.05.003, 2018.

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., Mark, 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, e2110472118, 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.

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.