

Response to Anonymous Referee #1

We thank the Referee for their insightful comments. We have implemented a number of changes as outlined below.

General Comment: The paper by Wolfe et al. describes airborne eddy covariance measurements on a C-23B Sherpa aircraft. It summarizes results from flights in the eastern US. The general topic is suitable for AMT, but there are a couple of issues that need to be addressed before publication. In particular, the discussion on errors needs revision.

Specific Comments:

It is not clear whether a Webb correction was necessary for CO₂ and H₂O fluxes, and how this was incorporated in the flux analysis code. The 10Hz humidity correction for the LGR instrument mentioned on page 5 (line 31) seems tricky – since there was a redundancy of humidity measurements, a better experimental setup would have been to use a Nafion dryer for the EC system and just focus on CO₂ and CH₄ to avoid this problem all together. As the data are treated it is not clear to what extent the water vapor flux influences CH₄ and CO₂ fluxes, or how the correction procedure would degrade the precision of the flux calculation, given the large random errors of 10 Hz concentration datasets.

We do not perform a Webb correction, as the 10 Hz observations of CO₂, CH₄, and H₂O (both LGR and DLH) are corrected to dry mixing ratios prior to calculation of fluxes. The interpolation of LGR H₂O to the CO₂ time base is not especially tricky, as 1) the gas sampling systems and native sampling rates are identical, and 2) the CO₂-H₂O correlation is sufficiently strong to provide good lag-correlation. We have added the following statement in the first paragraph of section 3.1:

Raw gas concentrations are provided as dry mixing ratios, eliminating the need for density corrections (Webb et al., 1980) to derived fluxes.

In our standard procedure, we calculate latent heat fluxes from DLH by first converting from mole fraction (moles per mole moist air, the native DLH measurement) to mixing ratio (moles per mole dry air), which theoretically negates the need for a Webb correction. As a test, we recalculated LE using mole fraction and applied a density correction following Eq. (23) of Webb et al. (1980). The full range of difference between the two methods is $\pm 1\%$, and the normalized mean bias is 0.12%. The LGR instrument water corrections (Fig. S3) include both dilution and spectroscopic effects, thus it is somewhat more difficult to separate out the dilution component in our analysis code, but we expect a similar result. We further note that other researchers also avoid density corrections through similar approaches (Desjardins et al., 2018).

For the initial CARAFE deployments, we decided that not drying the GHG sample gas was preferable as the comparison of LGR H₂O mixing ratios and fluxes against DLH provides a valuable performance

cross-check. Also, redundancy is insurance against instrument failure. We will consider drying the sample for future deployments.

According to eq. 12 the turbulent random error should always be smaller than the combined error which includes instrument noise. Inspecting figure 7 actually shows the opposite for most tracers; the relative turbulent error is larger than REFS01 for T, H2O and CO2; this contradicts the theory. An explanation is needed – could there be a calculation error in the analysis code?

The turbulence random error defined in Eq. (11) represents an upper limit (note the \leq sign). Thus, we expect that the empirical total random errors, RE_{FS01} and RE_{wave} , should be generally smaller than the root-sum-square of RE_{turb} and RE_{noise} . Fig. S9 illustrates this point for RE_{wave} . We have added a plot to Fig. S9 to show a similar correlation for RE_{FS01} and modified Sect. 3.4.2 as follows:

The maximum lag for the summation is set to 10 seconds based on comparison with the root-sum-square of RE_{noise} and RE_{turb} , the latter representing a theoretical upper limit for total random error (Fig. S9a).

Eq. 12 is cast in the time domain. For aircraft measurements the time domain is not really meaningful. The discussion of errors should be handled in the spatial domain. For example, a cut off frequency of 0.02 Hz corresponds to a distance of 3.75 km at the aircraft speed of the C-23B Sherpa. The same criterion would correspond to a 12 km distance on a G5-aircraft. The issue of spatial vs. temporal scale should be treated consistently throughout the manuscript. While the error discussion is treated in the time domain, some figures show a spatial, others a temporal scale. Figures 5 and 9 should be modified to show a spatial scale as well.

We agree that it is more appropriate to cast discussion in the spatial domain when referring to turbulence scales or wavelet-derived fluxes. For certain aspects, however, the time domain is meaningful – specifically, with regards to instrumentation. For example, the characteristic response times described in Sect. 3.4.1 are inherent to each instrument and independent of platform speed. A similar argument holds for the influence of instrument noise on spectra shown in Fig. 9(b). With specific regard to Eqs. (12) and (13), we would obtain the same results (for this particular set of measurements) regardless of whether we use the temporal or spatial domain, as the Sherpa cruise speed is fairly constant for most flux legs (81 ± 9 m/s). The use of temporal domain is mostly a matter of convenience.

We have added the following statement to the beginning of Sect. 3:

The following discussion references both the time and spatial domains as appropriate, the two coordinates being linked by leg-average aircraft speed.

In addition, we have added/modified text throughout Sections 3 and 4 to better address spatial vs temporal scales (see esp. Sect. 3.3), and we have added spatial scales to Figs. 4, 5, and 9.

Total error: Systematic errors inherent to unresolved scales always lead to an underestimation of fluxes and should be used to correct the data rather than adding these to a total error. Adding

systematic errors to the total error is generally only admissible, if they are not separable from other errors or if their sign cannot be defined. Neither is true for SE_{rt} and SE_{turb}.

Based on current literature, there seems to be no consensus in the flux community on how to handle systematic errors. Some groups lump systematic errors into total error as we have done (Misztal et al., 2014; Vaughan et al., 2016). Gioli et al. (2004) applies only a high-frequency correction, while Mauder et al. (2013) derives long-wavelength errors based on energy closure but advocates against using these to correct fluxes. Still other groups seem to ignore uncertainties entirely in their analysis (Desjardins et al., 2018; Sayres et al., 2017). Furthermore, SE_{rt} is sometimes unreasonably large when spectra are noisy (P. 11, Line 29), and SE_{turb} represents an upper limit. For these reasons, we believe it best to report SE as separate data columns and allow data users to decide how to treat these errors.

We have added the following discussion to the end of Sect. 3.4.1:

Systematic errors can be applied as a correction factor to fluxes (if of known sign) or be included as part of the total uncertainty. Both practices are common among the airborne flux community (Gioli et al., 2004; Misztal et al., 2014). For the errors discussed above, SE_{acc} is of unknown sign, while SE_{turb} and SE_{RT} should both increase the flux. We are, however, reluctant to employ the latter two as correction factors. SE_{turb} represents an upper limit and thus may slightly “over-correct” the fluxes, while SE_{RT} can become unrealistically large when fluxes are small due to the amplification of high-frequency noise by Eq. (9). Thus, we elect to include all systematic errors in the total flux error and assume all error components are symmetric for simplicity. Total systematic error (SE_{tot}), given as a fraction of the flux over any interval, is then the root-sum-square of SE_{turb}, SE_{rt}, and SE_{acc}. Total systematic error is reported as a separate variable in flux archive files and may be used as part of the total error or as a correction factor (after removing the accuracy contribution) at the discretion of data end-users.

Additional systematic errors for surface fluxes arising from flux divergence are discussed separately but should probably be part of section 3.4.

We have moved this discussion to Sect. 3.4 and added sections on total error and error averaging.

Repeatability: it is mathematically not sound to simply average second moments as presented in Figure 11 (see for example: <https://www.eol.ucar.edu/content/combining-short-term-moments-longer-time-periods>). Within the uncertainty of the presented data it might not make a large difference for Figure 11, but it would be worth double checking using the correct averaging formula.

From the referenced link, the relevant formula here is

$$\overline{x'y'}^N = \frac{1}{N} \sum_{j=1}^m N_j (\overline{x'y'}^j + \overline{X}^j \overline{Y}^j) - \frac{1}{N} \sum_{j=1}^m N_j \overline{X}^j \frac{1}{N} \sum_{j=1}^m N_j \overline{Y}^j$$

Here, x and y correspond to scalar and vertical wind measurements and there are m sub-intervals of length N_j included in the average over a total interval of length N. For any flux-relevant sub-interval,

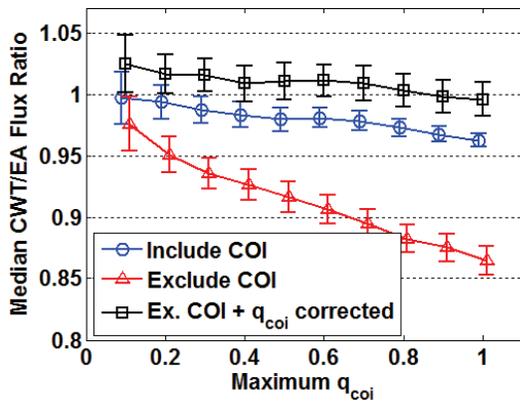
however, the mean vertical wind should be sufficiently close to 0 that the terms on the right are negligibly small. Also, all N_j are roughly equal in our averaging routines. In this case, the formula simplifies to an unweighted average, as used in this and many other studies.

We did check anyway, as the reviewer suggested. We find no appreciable difference in average fluxes using either averaging method.

Figure 6: the plotted differences are likely caused by a dramatic increase of systematic errors (eq. 7) towards the edges of the CWT – could the calculated flux ratios improve when accounting for these SE ? (e.g. by introducing a weighted SE along the CWT). To be more specific, the COI cuts off a substantial part of the frequency domain towards the edge of the CWT which should result in a systematic flux underestimation according to eq 7.

We assume that the reviewer is referring to the difference between the “exclude COI” and “include COI” cases. Exclusion of the COI from scale-averaging necessarily leads to a systematic underestimate of the true flux for the reasons the reviewer describes, and this is discussed in Sect. 3.3.1. Figure 6 provides an ensemble estimate for the resulting systematic error; in the case where we do not filter with the q_{coi} flag (rightmost points), the error (taken as the difference between the “include” and “exclude” cases) is $\sim 10\%$ of the flux. We have added some text to this section to clarify this point.

It is difficult to develop a robust (time-dependent along the CWT) estimate of the systematic error resulting from the COI, as this area by definition represents a region where the wavelet transform suffers from limited information. In theory the q_{coi} flag provides a rough means of doing this calculation: by first scale-averaging the CWT while excluding the COI, and then dividing by $(1-q_{coi})$ to correct for the fraction of cospectral power that was lost in the COI. The below plot shows the results of this calculation (black squares). The correction does indeed mitigate the systematic errors caused by excluding the COI for the ensemble of all fluxes. We hesitate to recommend such a correction for time-resolved CWT fluxes, however, as it inherently assumes that the globally-averaged ogive/cospectrum is representative of the local ogives/cospectra.



The purpose of the q_{coi} flag is to allow filtering of fluxes that may suffer from large COI-related systematic errors. We believe this conservative strategy is preferable to attempting to recover cospectral power within the COI via a correction factor.

Minor Comments: Figure 8: How high was zi?

1070 m. We have added this info to the figure caption.

Figure 9a: A label for the CO2 and CH4 instrument should be added (e.g. LGR)

The “LGR” is meant to distinguish between the two water measurements. Since there is only one CO2 and CH4 flux measurement, we feel this change is unneeded. Actually the label should be GHG as this is the name of the system; we have modified Figs. 7, 9 and 10 accordingly.

References

- Desjardins, R. L., Worth, D. E., Pattey, E., VanderZaag, A., Srinivasan, R., Mauder, M., Worthy, D., Sweeney, C., and Metzger, S.: The challenge of reconciling bottom-up agricultural methane emissions inventories with top-down measurements, *Agr. Forest Met.*, 248, 48-59, 2018.
- Gioli, B., Miglietta, F., De Martino, B., Hutjes, R. W. A., Dolman, H. A. J., Lindroth, A., Schumacher, M., Sanz, M. J., Manca, G., Peressotti, A., and Dumas, E. J.: Comparison between tower and aircraft-based eddy covariance fluxes in five European regions, *Agr. Forest Met.*, 127, 1-16, 2004.
- Mauder, M., Cuntz, M., Drue, C., Graf, A., Rebmann, C., Schmid, H., Schmidt, M., and Steinbrecher, R.: A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, *Agr. Forest Met.*, 169, 122-135, 2013.
- Misztal, P. K., Karl, T., Weber, R., Jonsson, H. H., Guenther, A. B., and Goldstein, A. H.: Airborne flux measurements of biogenic volatile organic compounds over California, *Atmos. Chem. Phys.*, 14, 10631-10647, 2014.
- Sayres, D. S., Dobosy, R., Healy, C., Dumas, E., Kochendorfer, J., Munster, J., Wilkerson, J., Baker, B., and Anderson, J. G.: Arctic regional methane fluxes by ecotope as derived using eddy covariance from a low-flying aircraft, *Atmos. Chem. Phys.*, 17, 8619-8633, 2017.
- Vaughan, A. R., Lee, J. D., Misztal, P. K., Metzger, S., Shaw, M. D., Lewis, A. C., Purvis, R. M., Carslaw, D. C., Goldstein, A. H., Hewitt, C. N., Davison, B., Beevers, S. D., and Karl, T. G.: Spatially resolved flux measurements of NOx from London suggest significantly higher emissions than predicted by inventories, *Faraday Disc.*, 189, 455-472, 2016.
- Webb, E. K., Pearman, G. I., and Leuning, R.: Correction of flux measurements for density effects due to heat and water vapour transfer, *Quarterly Journal of the Royal Meteorological Society*, 106, 85-100, 1980.

Response to Anonymous Referee #2

We thank the Referee for their insightful comments. We have implemented a number of changes as outlined below.

Summary/General comments: Wolfe et al. describe the CARAFE aircraft, payload, and measurement methodology including flight data from campaigns in 2016 and 2017. Much of the manuscript focuses on the airborne eddy covariance method, how it is applied, and uncertainty analysis. The manuscript is well written and well placed in AMT. The authors have done a commendable job attempting to investigate the many challenges and sources of uncertainty in performing airborne eddy covariance. I do have some reservations and questions that need to be addressed. Once appropriate changes are made I would recommend publication.

Presentation/conceptual concern: The manuscript presents the CARAFE payload and eddy covariance technique as a useful new tool for improving our understanding of carbon gas exchange. This tone underlies much of the manuscript, but the authors fall short of actually justifying, and this should be rectified. There is a cursory review of other airborne approaches that misses many techniques (such as the mass balance method), and the relative strengths/weaknesses are not really clearly highlighted. This isn't a problem if the manuscript focused on the CARAFE payload, but this would need to be addressed to assert the added value of airborne EC for CO₂ & CH₄. Even more so, the authors don't actually link observed eddy covariance to surface fluxes and provide added science value; it is made clear it is not known how to best link to horizontal spatial flux scales on the surface. More so, the error analysis suggests flux errors when considering the surface that can easily exceed 100%. I finished the manuscript wondering whether this approach was a wise usage of the aircraft and payload. Making high accuracy GHG airborne measurements from aircraft can be used with mass balance and different inversion systems to quantify fluxes with errors of 20%. With such an approach, larger areas can be covered with the aircraft as repeat legs are not needed and there are far less stringent requirements on level flight and surface characteristics, enabling the usage of far more data. Further, problems like the 2016 wind measurement error reported render all those flights of no scientific value because of the stringent requirements for EC. I commend the authors for their efforts and rigorous analysis, but at this point they cannot assert that airborne EC for CO₂ & CH₄ as presented in the manuscript provides added science value over more conventional accurate airborne sampling. I would actually think given the gaps in linking to horizontal surface domains, the tight restrictions on where the approach is useful, the limits imposed on flight area coverage, and the high fractional uncertainty, it is worth questioning if for surface carbon exchange this technique will add to addressing current science questions or whether accurate flight measurements for usage in inversions and mass balance approaches would be more scientifically fruitful. My suggestion is that the authors make changes in the abstract, introduction, and conclusion to more accurately capture this reality. The emphasis should be on the presentation of the CARAFE payload. The extensive discussion of EC and uncertainty should remain, but a clear discussion of the limitations and that added science value is yet to be shown should be made clear.

Most of the above comments are handled more specifically below. We have substantially revised the introduction and error discussion following the comments of both reviewers.

Regarding the utility of airborne EC, the wealth of work cited in the introduction is, we hope, sufficient evidence. We do not claim that airborne EC is a replacement for other methods. All methods have unique strengths and weaknesses, and all are needed for scientific progress on our understanding of surface exchange, which is a very broad umbrella.

Detailed comments:

Page 1 Line 16: not accurate as stated – exchange between surface and atmosphere only drives atmospheric abundance of some gases – not all atmospheric composition.

We have changed “drives” to “strongly influences.”

Page 1 Line 17: should modify to “potentially helping”. Also, what are you defining as regional? Need spatial scale. Traditional airborne measurements can cover similar scales so would need to be specific and distinguish.

Wording changed. We define local to regional scales as 1 – 1000 km.

Page 1 Line 25-26: It does not follow from the paper that this system will further our understanding of ecosystem exchange – this has not been established.

We have changed the last sentence as follows:

Results from these campaigns highlight the performance of this system and its potential to provide fresh observational constraints on ecosystem exchange.

Page 2 Line 3: the Dlugokencky reference is a very incomplete citation for such a broad statement.

We have added several additional references.

Page 2 Line 4: The above described global approaches can also be defined as topdown and bottom-up. Need more specificity referencing spatiotemporal scales.

We have changed the wording to clarify here. The first paragraph refers to abundance-based budgets, and the second to flux distributions/budgets.

Page 2 Line 4-14: This is a very cursory coverage of other approaches that does not address many airborne approaches (mass balance, point source circling, eulerian/lagrangian inversion) that have been well established to evaluate fluxes at 10-100 km scales. NB those approaches are more flexible than EC and can deal with point sources that can be important for CO₂ (power plants) and CH₄ (lots of point sources). Addressing point sources is important for Carbon gases, and EC is ill-equipped for this. This point needs to be addressed.

We have added several paragraphs to the intro mentioning these other methods and discussing the strengths/weaknesses of airborne EC.

Page 2, line 15: EC does not directly quantify surface-atmosphere exchange – it quantifies exchange between two atmospheric levels. An important distinction, as surface exchange is inferred, which large errors induced due to flux divergence.

We have changed the wording as follows:

Eddy covariance (EC) directly quantifies vertical turbulent fluxes in the atmospheric boundary layer.

Page 2 line 24-26: As stated above, other airborne approaches are more flexible and have similar spatial capabilities.

This paragraph focuses on the distinction between ground and airborne EC, but has been deleted in revision.

Page 3 lines 1-2: This strong statement needs citation support.

We have deleted this sentence as part of the intro revisions. This is an inherent feature of the wavelet transform due to its localized nature.

Page 3 lines 28-31: This is illustrative of the very limited spatial domain that can be covered.

The total 2017 dataset includes over 3000 km of surface fluxes. We would not call this limited.

Section 2.2: I need to see more validation of winds. We should see the results from box patterns and other maneuvers done to test/validate winds, and thus be able to determine accuracy.

We have added a new figure for wind calibrations (Fig. S1) and reference to this figure in the text.

Section 2.2: The problem with the 2016 data is buried here. Based on this large, systematic problem, the authors decide not to use 2016 data. The authors should follow through and only show 2017 data (there is 2016 data still in many places). Further, this point should be made up front in the manuscript – small mistakes led to wind problems that rendered a whole deployment not useful for EC. This is illustrative of a major weakness to the CARAFE approach.

We call out this problem explicitly in the text and provide a figure in the SI to illustrate the effects. The upshot is an unfortunate but well-defined systematic error that is partially correctable. While it reduces the utility of the 2016 data, it does not render the whole deployment useless. We have added the following text to this section to better clarify/quantify:

Comparison of the two wind systems for 2017 indicates that the 858 probe/Honeywell system misses $28 \pm 3\%$ of vertical wind variance, resulting in a systematic flux underestimate of $\sim 24\%$ (Fig. S2b). Division of all 2016 fluxes by a factor of 0.76 rectifies this bias in the mean sense, but additional random error arises from point-to-point variability. Based on differences in 2017 fluxes derived from

the two wind datasets, we estimate 1σ random errors in 2016 1Hz fluxes of sensible heat, latent heat, CO_2 and CH_4 of 50 W m^{-2} , 110 W m^{-2} , $7 \mu\text{mol m}^{-2} \text{ s}^{-1}$, and $50 \text{ nmol m}^{-2} \text{ s}^{-1}$, respectively. These errors are in addition to those discussed in Sect. 3 and reduce with averaging (note the tighter correlation for leg-average fluxes in Fig. S2b).

We have also added several notes about this in Sect. 3.4, and the archive data files will be updated accordingly.

The 2016 flux data appears only one time, in Fig. 8. We prefer to leave this as is; it is our best example of the effects of diel cycles on divergence.

We have added a sentence in the intro stating that the measurement requirements make this a challenging technique, but we might also note here that such problems are not isolated to airborne EC. For example, a recent mass balance analysis on CH_4 emissions was retracted following discovery of an issue in another aircraft wind system (Ren et al., 2018).

Page 5 line 4: Not sure where this comes from. I'd like to see more on this.

We have added several references to support DLH accuracy. Also, we note the excellent agreement with independently-calibrated GHG water vapor shown in Fig. 10.

Section 2.4: I need to see more on the in-flight performance of the GHG analyzers. What is the accuracy in flight? Can the authors show the LGR analyzers show no vertical dependency compared with the Picarro?

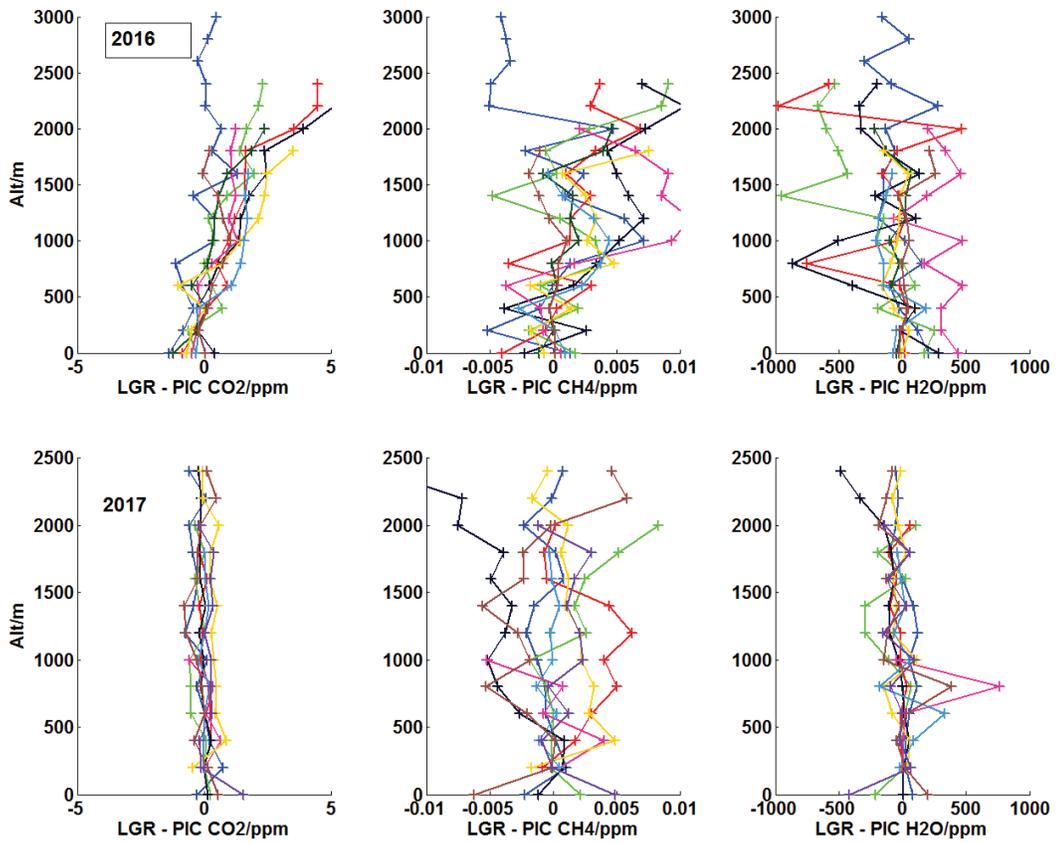
We do not currently calibrate the GHG system in-flight. Chen. et al. (2010) have demonstrated the stability of the Picarro G1301-m under flight conditions and suggest that ground-only calibrations are sufficient to obtain an accuracy of 0.05 ppm for CO_2 . Karion et al. (2013) suggest a total uncertainty (including precision and other error terms) of 0.23 ppm for CO_2 and 2.3 ppb for CH_4 from another airborne G1301-m system operating over a broader range of altitudes than those covered by CARAFE. Based on these values and considering that we do not calibrate in-flight, we have revised our accuracy estimates in Table 1 to be more conservative.

The LGR is necessarily less accurate than the Picarro because of our fitting procedure, by roughly a factor of 3-4 based on comparisons between the final 1 Hz data products. This information is included in our data files but was accidentally omitted in Table 1, and we have added it now.

High accuracy is not critical for airborne EC, as evidenced by our error discussion; accuracy is generally a small term in the total flux error budget (Fig. 7). Indeed, in the case of CO_2 and CH_4 , the accuracy error is dominated by the 5% winds uncertainty.

As shown in the plots below (colored lines denote different flights), there is some vertical dependence in the difference between the LGR and Picarro for CO_2 measurements in 2016, the reasons for which are unclear at this time (possibly improved pressure control). The bulk of surface flux observations

come from measurements below 500 m, where the difference is generally within our stated uncertainty. This issue has little bearing on derived fluxes.



Line 11: pressure fluctuations may impact accuracy however.

In laboratory calibrations, we do not observe any significant correlation of measured concentrations with sample pressure over this small range.

Page 6 line 5 & Figure 2: I am unclear on this linear transforming one instrument to the other. More clarity is needed. I also am concerned that this may not be appropriate for CO2 and CH4. When I look at figure 2 I get greatly concerned as the variation from flight to flight is actually very significant for gases that we care about fractions of a ppm (ppb for CH4). I also worry about inflight variations. We need more information on the validation of the GHG obs.

We have added the following text to Sect. 2.4.

The transformation requires several operations, including 1) averaging dry mixing ratios to a common 1-Hz time-base, 2) smoothing LGR data to match the slower cell throughput of the Picarro, 3) time-lagging the LGR (typically < 2 s) to optimize correlation with the Picarro, 4) calculation of fit coefficients for an ordinary least-squares fit ($LGR = m \cdot PIC + b$), and 5) correction of LGR dry mixing ratios using the

fit parameters. This procedure, akin to performing a flight-by-flight span and intercept calibration correction, rectifies calibration errors and flight-to-flight drift that may occur in the LGRs under different operating conditions.

As mentioned above, accuracy is not critical for airborne EC. We have also added the following to the end of Sect. 2.4.

It is not currently feasible to calibrate the LGR systems in-flight due to high gas flow rates. As a result of our correction procedure, we estimate that the accuracy of LGR CO₂ and CH₄ is degraded by a factor of 3 and 4, respectively, compared to the Picarro (Table 1). It may be possible to expand the above correction method to account for potential in-flight variability in LGR accuracy; however, as discussed later, measurement accuracy is a negligible contributor to total flux uncertainty for greenhouse gases.

Section 2.5 (and figure 3): This section is somewhat of an aside. There is not other usage or discussion of this system.

This system was part of the CARAFE payload and deserves mention. At present we have not been able to devote resources to making full use of this data, but we feel it is important to let readers know of its availability and discuss how it was acquired.

Page 7 line8-9: This +/-20 m requirement is very tight (as needed)- and will restrict the ability to use this data. Also, it should be made clear this is above ground and not asl. This makes it even harder to meet this requirement over terrain with any variability exceeding 20m. Further the 5 degree restriction is tough, but might it need to be tighter?

This sentence specifically states that the altitude range is with reference to ground level. The limit of +/- 20 m was chosen for this particular set of flights based on observed variability in the data. It is not a hard limit for airborne EC. We have added the following statement to clarify:

We note that this altitude window is based on observed variability in the 2016 and 2017 data-sets and is not a hard limit.

We do not require flat terrain; for example, (Wolfe et al., 2015) reported fluxes over rolling terrain with ground altitudes of 200 – 350 m asl. Airborne EC is of course not possible over extremely rough terrain (mountains). We discuss this point in the revised introduction.

The 5-degree roll restriction is to ensure quality in the vertical wind data. The limit is based on inspection of the data and expert judgment from operators who have flown similar systems on multiple NASA aircraft. These data filtering procedures are standard among airborne EC groups (Misztal et al., 2014).

Page 7 Line 22: The problem is undersampling of the PBL depth can lead to systematic biases, and this is a major problem.

PBL depth variations have little bearing on derived fluxes and do not, to our knowledge, lead to systematic biases. We have added some text to clarify this point. Misztal et al. (2014) estimate a maximum uncertainty of 3% in isoprene surface fluxes resulting from +/- 100 m uncertainty in PBL depth, but they also arrive at their divergence correction differently from us.

Page 11 Line 19: I'm confused, I had thought the authors earlier asserted 10Hz wasn't necessary, but here it seems this is an important error term.

Median values for SE_{RT} are 4-5%, which is roughly double that of SE_{turb} . We have modified the text here to be more specific. Though non-negligible, this error term is smaller than typical random errors. Moreover, it is systematic and thus correctable, as we discuss in revisions to Sect. 3.4.1 following Reviewer 1's suggestions.

Page 14 line 18: This is disappointing. If all these flight hours are being used there should be more planning for multiple altitude legs for this type of validation.

We have deleted this sentence as it is irrelevant to the discussion. Divergence corrections are a well-established aspect of airborne EC. We elected to allot most flight time to collecting near-surface observations to maximize statistical robustness in derived surface fluxes.

Page 15 lines 7-8: And these large uncertainties are a major problem for the approach. With relative uncertainties that push to 100% the utility of the technique is degraded.

The quoted uncertainties represent a full range across a wide distribution with a long tail. We have changed the error discussion throughout to consistently use interquartile ranges, and this is stated at the beginning of Sect. 3.4. We have also modified Fig. 7 and added text (Sect. 3.4.4 and 3.4.5) to clarify total errors in surface fluxes and the effects of averaging. For example, the interquartile range for uncertainty in CO₂ fluxes is 18 – 30% for leg averages and 40 – 90% for 2 km averages. This is comparable to uncertainties reported other airborne flux studies (Vaughan et al., 2016).

Page 16 line 15-16: This is fair, but it means this manuscript has not established the utility of this approach.

As discussed above, we believe that numerous previous studies have established the utility of airborne EC.

Conclusions: I take issue with much of how the conclusion is written (see major comment above). This would be better served to summarize the aircraft system and payload, and then highlight the challenges in the EC approach and the resulting expected uncertainties.

We have added several sentences to the opening paragraph discussing the payload and uncertainties. Regarding the remainder of this section, this seems to be the appropriate place to discuss potential uses of this dataset and future potential applications of the technique in general.

Page 17 Lines 29-30: It has not been established that this approach should be a part of a standard toolbox.

Respectfully, we disagree. Numerous previous studies have highlighted the value of such measurements, as detailed in the introduction. The atmospheric chemistry community has made significant strides with this technique over the last few years, and the carbon/biosphere community has benefited from it in the past. As we state, acquisition of the data is only the first step. We are now developing tools to analyze this data in a scientific context.

At any rate, we have modified this sentence following the reviewer's previous comment, as follows:

The NASA CARAFE project aims to incorporate eddy covariance fluxes as a standard component of the airborne science toolbox

Page 18 Lines 7-8: This isn't a new vector – as stated by authors the approach is old, and has been applied to Carbon before. The utility has never really been established, particularly given the limitations, and that is why it has hardly been adopted.

We have changed the wording here. Please also see our response to the comment above.

Page 18 Line 17: This does not clearly follow.

We have changed "would" to "may" and added several references.

Table 1: Methane and CO2 should be shown in ppb and ppm respectively (not fractional uncertainty).

Fixed.

Figure 1: The authors should indicate which flight legs are actually of use for EC on this plot – showing all the flight legs is misleading. 2016 data was deemed not useable, so should not be shown.

We have changed the map to include only flux legs. As discussed above we have not deemed the 2016 data "unusable," thus we have left this information on the map.

Figure 2: These are worryingly large to me. Also, this should show 2017 data as the authors don't use the 2016 flights.

The quality of concentration measurements for 2016 and 2017 are comparable, it is only the fluxes that bear the systematic error from turbulence sampling. We have added the 2017 version of this figure to the appendix and added a reference to this figure in the text.

Figure 7: This plot is sobering, and the log scale relative error brings into question the utility for CO2 and CH4.

Please see our response to the comment on Page 15 above. The bulk of the distribution for total errors is not unreasonable.

References

- Chen., H., Winderlich, J., Gerbig, C., Hofer, A., Rella, C. W., Crosson, E. R., Van Pelt, A. D., Steinbach, J., Kolle, O., Beck, V., Daube, B. C., Bottlieb, E. W., Chow, V. Y., Gantoni, G. W., and Wofsy, S. C.: High-accuracy continuous airborne measurements of greenhouse gases (CO₂ and CH₄) using the cavity ring-down spectroscopy (CRDS) technique, *Atmos. Meas. Tech.*, 3, 375-386, 2010.
- Karion, A., Sweeney, C., Wolter, S., Newberger, T., Chen, H., Andrews, A., Kofler, J., Neff, D., and Tans, P.: Long-term greenhouse gas measurements from aircraft, *Atmos. Meas. Tech.*, 6, 511-526, 2013.
- Misztal, P. K., Karl, T., Weber, R., Jonsson, H. H., Guenther, A. B., and Goldstein, A. H.: Airborne flux measurements of biogenic volatile organic compounds over California, *Atmos. Chem. Phys.*, 14, 10631-10647, 2014.
- Ren, X., Hall, D. L., Vinciguerra, T., Benish, S. E., Stratton, P. R., Ahn, D., Hansford, J. R., Cohen, M. D., Sahu, S., He, H., Grimes, C., Salawitch, R. J., Ehrman, S. H., and Dickerson, R. R.: Retraction: Methane Emissions From the Marcellus Shale in Southwestern Pennsylvania and Northern West Virginia Based on Airborne Measurements, *J. Geophys. Res. Atmos.*, doi: 10.1002/jgrd.54397, 2018. n/a-n/a, 2018.
- Vaughan, A. R., Lee, J. D., Misztal, P. K., Metzger, S., Shaw, M. D., Lewis, A. C., Purvis, R. M., Carslaw, D. C., Goldstein, A. H., Hewitt, C. N., Davison, B., Beevers, S. D., and Karl, T. G.: Spatially resolved flux measurements of NO_x from London suggest significantly higher emissions than predicted by inventories, *Faraday Disc.*, 189, 455-472, 2016.
- Wolfe, G. M., Hanisco, T. F., Arkinson, H. L., Bui, T. P., Crouse, 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, 2015.

The NASA Carbon Airborne Flux Experiment (CARAFE): Instrumentation and Methodology

5 Glenn M. Wolfe^{1,2}, S. Randy Kawa¹, Thomas F. Hanisco¹, Reem A. Hannun^{1,2}, Paul A. Newman¹,
Andrew Swanson^{1,3}, Steve Bailey¹, John Barrick⁴, K. Lee Thornhill⁴, Glenn Diskin⁴, Josh DiGangi⁴,
John B. Nowak⁴, Carl Sorenson⁵, Geoffrey Bland⁶, James K. Yungel⁷, and Craig A. Swenson⁸

¹Atmospheric Chemistry and Dynamics Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA

²Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, MD, USA

³Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, MD, USA

10 ⁴NASA Langley Research Center, Hampton, VA, USA

⁵NASA Ames Research Center, Moffett Field, CA, USA

⁶Wallops Flight Facility, NASA Goddard Space Flight Center, Wallops Island, VA, USA

⁷AECOM, ATM Project, Wallops Flight Facility, NASA Goddard Space Flight Center, Wallops Island, VA, USA

⁸SSAI, ATM Project, Wallops Flight Facility, NASA Goddard Space Flight Center, Wallops Island, VA, USA

15 *Correspondence to:* Glenn M. Wolfe (glenn.m.wolfe@nasa.gov)

Abstract. The exchange of trace gases between the Earth’s surface and ~~its~~ atmosphere ~~drives strongly influences~~ atmospheric composition. Airborne eddy covariance ~~can provide observational constraints on~~ can quantify surface fluxes at local to regional scales (1 – 1000 km), potentially helping to bridge ~~the gaps~~ between top-down and bottom-up flux estimates and offering novel insights into biophysical and biogeochemical processes. The NASA Carbon Airborne Flux Experiment (CARAFE) utilizes the NASA C-23 Sherpa aircraft with a suite of commercial and custom instrumentation to acquire fluxes of carbon dioxide, methane, sensible heat, and latent heat at high spatial resolution. Key components of the CARAFE payload are described, including the meteorological, greenhouse gas, water vapor, and surface imaging systems. Continuous wavelet transforms deliver spatially-resolved fluxes along aircraft flight tracks. Flux analysis methodology is discussed in depth, with special emphasis on ~~evaluation-quantification~~ of uncertainties ~~and vertical flux divergence~~. Typical uncertainties in derived surface fluxes are 40—90% for a nominal resolution of 2 km or 16—35% when averaged over a full leg (typically 30 – 40 km). CARAFE has successfully flown two missions in the Eastern U.S. in 2016 and 2017, quantifying fluxes over forest, cropland, wetlands, and water. Preliminary Results from these campaigns are presented to highlight the performance of this system ~~and its potential to further our understanding of ecosystem exchange~~.

20
25

1 Introduction

30 Accurate, quantitative, process-based understanding of current greenhouse gas (GHG) budgets (principally carbon dioxide (CO₂) and methane (CH₄)) is essential to projecting carbon-climate feedbacks and, hence, future climate (Friedlingstein et al., 2006; Schimel et al., 2015). While the atmospheric concentrations of these gases can be readily measured (Andrews et

al., 2014), it is the sources and sinks at the Earth's surface and chemical conversion in the atmosphere that drive their changing abundances. ~~On a global basis, the typical~~

~~Global~~ CO₂ budgets ~~is~~ are typically constructed in a bottom-up sense from fossil fuel use inventories, estimates of ocean flux from solubility calculations, the measured time rate of change of the atmospheric CO₂ burden, perhaps a land use change/biomass burning emission term, and a land vegetation uptake flux inferred as the residual (Ciais et al., 2013; Le Quere et al., 2016). This construct provides very little information on the nature and distribution of the land flux or its potential variations. Global CH₄ budgets are similarly under-constrained in detail (e.g., Bousquet et al., 2011; Dlugokencky et al., 2011; Worden et al., 2017).

~~More~~ Somewhat more specific ~~flux~~ information on source and sink distributions is commonly inferred from so-called top-down and bottom-up flux estimates. The former involves measuring atmospheric gradients in GHG mixing ratios and combining them with some estimate of volume transport to infer flux to/from the surface (flux inversion) (Chevallier et al., 2010; Gurney et al., 2002). This method can be applied across a wide range of scales (global to 10's of km, depending on model resolution) but requires accurate transport characterization and intensive, high-accuracy GHG sampling. More recently, satellite measurements have been employed to provide increased sample density and coverage beyond that available from in situ measurements (Basu et al., 2013; Houweling et al., 2015). Bottom-up GHG flux estimates for terrestrial vegetation can be obtained from biogeophysical process models (Schaefer et al., 2008) and/or analysis of GHG flux observations, the latter typically from large tower networks such as AmeriFlux (Boden et al., 2013; Jung et al., 2012). Top-down and bottom-up flux estimates often compare poorly and models disagree among themselves in regional flux estimation, with uncertainties typically exceeding 100% for continental-scale flux estimates (Hayes et al., 2012; Huntzinger et al., 2012). Validation of inferred fluxes is challenging due to both their coarse spatiotemporal resolution and a paucity of suitable observations. ~~This situation suggests the need for a complimentary measurement approach, such as airborne eddy covariance.~~

Observations of surface exchange at scales typically accessible to aircraft (1 – 100's of km) permit validation of high-resolution emission inventories, probing of detailed biogeochemical interactions (e.g. drought stress), and general characterization of spatiotemporal gradients not resolvable from large-scale flux estimates. Several airborne methods have been developed for flux quantification, including mass balance approaches (Cambaliza et al., 2017; Karion et al., 2013a; Trousdell et al., 2016), inversions (Chang et al., 2014; Commane et al., 2017), and eddy covariance (Dabberdt et al., 1993; Desjardins et al., 1982). Each of these methods possess unique strengths and weaknesses in terms of their measurement requirements, spatiotemporal resolution, and applicability to various processes (e.g. point source vs areal emissions, emission vs deposition/uptake). An exhaustive review is outside the scope of this work, which focuses on airborne eddy covariance.

Eddy covariance (EC) directly quantifies ~~surface-atmosphere exchange~~ vertical turbulent fluxes in the atmospheric boundary layer. ~~The technique relies on acquiring fast, simultaneous measurements of vertical wind speed and a scalar quantity (e.g. temperature, gas concentration).~~ When measuring near the surface, fluctuations in vertical wind speed and

scalar magnitude (e.g. temperature, gas concentration) correlate positively/negatively if the surface is a net source/sink for that scalar ~~and negatively if the surface is a net sink~~. The time- or spatial-average product of vertical wind and scalar fluctuations (their covariance) thus yields a direct measurement of the flux at the measurement altitude. Extrapolation to the surface is possible with knowledge of the vertical flux divergence (change in flux with altitude), which is typically linear and can be obtained from flux measurements at multiple altitudes or independent constraints on the continuity equation (Conley et al., 2011; Lenschow et al., 1980). EC requires high-precision measurements of scalar and vertical wind fluctuations throughout the range of turbulent timescales (up to several Hz for sampling mixed-layer eddies), which can be technically challenging. Uncertainties, dominated by the stochastic nature of turbulence, typically range from 20–80% for horizontal averaging scales of 1 – 30 km but can exceed 100% when fluxes are small. Similar error ranges are reported for other surface exchange quantification methods (Cambaliza et al., 2014; Chang et al., 2014; Heimburger et al., 2017).

The main advantage offered by airborne EC is the ability to map gradients in surface exchange at relatively fine scales (~1 km) and over relatively broad regions (~100's of km). As with ground-based EC, airborne EC is not feasible over rough terrain (e.g. mountains), but moderate terrain (rolling hills) is acceptable (Misztal et al., 2014; Wolfe et al., 2015). The technique is especially well suited for disperse sources and sinks, such as vegetation and open water. Though not designed for single point sources, airborne EC can quantify aggregate fluxes over multiple small emitters, such as oil and shale gas production regions (Yuan et al., 2015). Airborne fluxes do not provide the long-term temporal information afforded by tower networks, but they can help characterize tower representativeness and/or extend tower observations to larger ecosystems (Chen et al., 1999; Kustas et al., 2006). In combination with spatially-resolved surface information (e.g. remotely-sensed vegetation properties), airborne fluxes can also help to refine surface exchange parameterizations (Anderson et al., 2008; Zulueta et al., 2013).

~~Airborne and ground-based flux techniques each offer unique advantages. Tower networks can provide sparse but continuous data over long periods of time; however, fixed towers typically have an effective “flux footprint” of a few km or less and may not be fully representative of an ecosystem. Airborne observations provide superior spatial coverage but at the cost of reduced temporal information. Upscaling to diurnal cycles and long-term trends may require coupling with co-located tower observations and/or biophysical models. The main advantage offered by airborne EC, especially when combined with wavelet transforms, is the ability to characterize a wide range of ecosystems at model-relevant scales (10–100 km) while retaining the fine scale information (~1 km) needed to test process-based biophysical parameterizations.~~

Airborne EC has elucidated surface-atmosphere exchanges processes for more than three decades (Dabberdt et al., 1993; Desjardins et al., 1982; Lenschow et al., 1981; Ritter et al., 1992; Ritter et al., 1994; Ritter et al., 1990; Sellers et al., 1997). Recent GHG applications include evaluations of NEE over forests and cropland complex ecosystems (Miglietta et al., 2007; Zulueta et al., 2013) and, quantification of CH₄ emissions from shale gas production regions (Yuan et al., 2015), and agricultural areas (Desjardins et al., 2018; Hiller et al., 2014), and Arctic biomes (Sayres et al., 2017). The technique has also proven valuable for measurements of emissions, deposition and chemistry of reactive gases ~~over forests~~ (Gu et al., 2017; Karl et al., 2009; Karl et al., 2013; Misztal et al., 2016; Misztal et al., 2014; Wolfe et al., 2015; Yu et al., 2017).

Traditionally, airborne EC has been limited to small, low-flying aircraft (Gioli et al., 2004), but recent work has demonstrated successful flux observations from larger platforms (Wolfe et al., 2015; Yuan et al., 2015) that offer increased payloads for more complete [atmosphere and ecosystem](#) characterization. ~~Furthermore, the application of wavelet transforms relaxes experimental requirements for stationarity and homogeneous land cover~~(e.g., Desjardins et al., 2018; Misztal et al., 2016).

The NASA Carbon Airborne Flux Experiment (CARAFE) is a new system engineered specifically for acquisition of airborne fluxes and related properties. To date, CARAFE has flown two 40-hour missions, one in September 2016 and another in May 2017. Table S1 in the supplementary information (SI) details the times and locations of each flight, and flight tracks are shown in Fig. 1f. Based out of NASA Goddard Space Flight Center’s (GSFC) Wallops Flight Facility (WFF), flights targeted forest, farmland, wetlands and open water along the central U.S. East Coast. Here we describe the key components of the CARAFE payload and the methodology for deriving surface fluxes of CO₂, CH₄, sensible heat and latent heat. We utilize selected observations from both missions to demonstrate capabilities and performance. Future publications will present flux results for specific process representations in greater detail.

2 Platform and instrumentation

Here we describe the aircraft and core measurements included on both CARAFE deployments. Table 1 summarizes relevant specifications for each system. Key components include the aircraft, 3-D winds and associated meteorology, fast water vapor and greenhouse gas measurements, and a surface imaging system.

2.1 NASA C-23B Sherpa

The NASA C-23B Sherpa (Fig. 1a) is a high-wing, twin turboprop aircraft operated by NASA WFF. Modifications to support airborne science include addition of a variety of instrument ports and dedicated experimenter power (Fig. 1b—e). The Sherpa was first deployed during NASA’s Carbon in Arctic Reservoirs Vulnerability Experiment (CARVE) to survey the abundance of CO₂, CH₄, and related gases in Canada and Alaska (Chang et al., 2014). Subsequent upgrades for CARAFE, as detailed below, now permit direct observations of surface exchange via eddy covariance.

The Sherpa is ideally suited for airborne flux measurements. ~~The nominal minimum~~Typical flight speeds of ~~75–80 ±~~ 10 m s⁻¹ ~~is-are~~ sufficient for sampling turbulence statistics in the mixed layer. For example, a 10 Hz measurement corresponds to a data point every ~~7.58~~ m, and the scale of peak turbulence is roughly proportional to altitude (~100 m for a typical low-level leg). The nominal altitude envelope of 0.1—3 km facilitates both near-surface sampling and boundary layer profiling, and a range of ~1000 km (duration of 4—5 h) permits regional sampling from a single deployment location. With a payload weight and power capacity of 7000 lb and 6 kW, the Sherpa can support a full measurement suite for detailed in situ and surface observations. The payload described below uses roughly half of available weight and power.

Typical flux flight patterns consist of stacked level legs and vertical soundings. Level legs are mostly at an altitude of 90—150 m above ground level and range in length from 20—100 km. Occasional level legs higher in the boundary layer (200—400 m) provide a constraint for vertical flux divergence. Vertical soundings are required to assess boundary-layer depth and ideally occur in the middle of the target area at both the beginning and end of a sortie.

5 2.2 Meteorology and telemetry

The Sherpa Turbulent Air Motion and Meteorology System (TAMMS) is a suite of sensors for high-frequency measurement of horizontal and vertical wind vectors, pressure, and temperature. Winds are derived from five pitot static-pressure ports mounted on the radome (nose) of the aircraft (Brown et al., 1983; Thornhill et al., 2003). In essence, these sensors provide the velocity of air with respect to the aircraft. Combining this information with a high-quality GPS and Inertial Navigation System (Applanix 510) yields the velocity of air with respect to the Earth's surface. Calibration via standard aircraft maneuvers (Barrick et al., 1996) corrects for aircraft motion and specific features of the pressure field around the aircraft (Fig. S1). A hatch on the forward left side supports a Rosemount model 858 angle-of-attack probe for redundant vertical wind measurements and a Rosemount model 102 non-deiced total air temperature sensor housing coupled with a platinum sensing element (E102E4AL) (Stickney et al., 1990) for fast (~8 Hz) air temperature (Fig. 1c). A NASA Airborne Science Data and Telemetry (NASDAT) system (<https://asapdata.arc.nasa.gov/asf/sensors/nasdat.html>) records data from these sensors at 20 Hz. The NASDAT also serves as a hub for GPS and network connections to other instruments.

The quality of the 3-D wind measurement hinges on the performance of the differential pressure measurements. For the 2016 campaign, both the radome system and the 858 probe were equipped with Honeywell PPT2 transducers. Afterward, spectral analysis of vertical wind speeds revealed anomalies at frequencies above 0.1 Hz. For the 2017 campaign the radome system was equipped with pressure sensors employing a higher sampling rate (Rosemount Model 1221), giving wind spectra more consistent with theoretical expectations (Fig. S1aS2a). Comparison of the two wind systems for 2017 indicates that the 858 probe/Honeywell system ~~misses under-samples~~ $28 \pm 3\%$ of vertical wind variance, resulting in a systematic ~~flux bias flux underestimate~~ of $18 \pm 2\% \sim 24\%$ ~~for leg average fluxes and larger potential bias at higher spatial resolution~~ (Fig. S1bS2b). ~~Division of all 2016 fluxes by a factor of 0.76 rectifies this bias in the mean, but additional random error arises from point-to-point variability (discussed further in Sect. 3.4).~~ Due to this issue, fluxes and related quantities presented here will primarily utilize results from the 2017 mission. ~~Work is ongoing to develop a spectral correction for the 2016 observations.~~

An upward-looking photosynthetic photon flux density (PPFD) sensor (LI-COR LI-190R) is mounted on the wing and sampled at 1 Hz via the greenhouse gas system. This sensor is designed for level and stationary ground applications but performed well on initial flights. Section S1 in the SI describes post-processing of PPFD data to correct for aircraft attitude and sun position.

2.3 Fast water vapor

The NASA Langley Diode Laser Hygrometer, or DLH (Diskin et al., 2002), is an open-path infrared absorption spectrometer that uses a variation of wavelength modulation spectroscopy (Silver, 1992) to measure water vapor ~~mixing ratios~~mole fraction. The DLH uses a laser locked to a water vapor absorption feature at $\sim 1.395 \mu\text{m}$ and directs the beam from a transceiver mounted on the fuselage onto a retroreflector fixed to the upper surface of the Sherpa landing gear fairing (Fig. 1d). The returning light is collected and detected in the transceiver with a total roundtrip light path of $\sim 2.5 \text{ m}$. Modulated signals are demodulated at twice the driving frequency (2F detection) and are converted to water vapor ~~mixing ratios~~mole fraction using laboratory-determined laser characteristics, spectral parameters taken from the HITRAN 2012 database (Rothman et al., 2013), and the aircraft static pressure and temperature measurements. Raw data are processed at the instrument's native $\sim 100 \text{ Hz}$ acquisition rate. For CARAFE, data are averaged to 20 Hz with a typical precision (1σ) of 0.3% or better. Overall measurement accuracy is within 5% , based on ~~numerous~~ field inter-comparisons on other airborne platforms (Jensen et al., 2017; Rollins et al., 2014).

2.4 Greenhouse Gas Suite (GHG)

The GHG system consists of several modified commercial analyzers coupled with custom hardware for fast gas flow and centralized data acquisition. Two Los Gatos Research (LGR) analyzers, one for CO_2 (model # 907-0020-1000) and the other for CH_4 and water vapor (model # 913-0014-0001), acquire mixing ratios at 10 Hz . The gas sampling system of each LGR is modified with a proportioning valve (IQ Valves, $0.234''$ orifice) coupled to a PID controller (Omega) to maintain a sample cell pressure of $140.0 \pm 0.1 \text{ Torr}$ throughout the Sherpa altitude range. Small pressure fluctuations in this range do not noticeably impact measurement precision, and instrument pressure fluctuations are uncorrelated with wind speed fluctuations. Dry scroll pumps (Edwards nXDS15i) maintain a gas flow of $\sim 35 \text{ SLM}$ through each system. Laboratory evaluation of the LGR time response (Fig. S4) gives an e-fold flush time of $90 \pm 16 \text{ ms}$, or an effective cutoff frequency (following the definition of Aubinet et al. (2016)) of 3.8 Hz . A Picarro G1301-m analyzer supplies an additional set of $\text{CO}_2/\text{CH}_4/\text{H}_2\text{O}$ mixing ratios. Compared to the LGR analyzers, the Picarro provides a reduced duty cycle ($\sim 5 \text{ s}$ for CO_2 and CH_4 , $\sim 15 \text{ s}$ for H_2O) but greater precision and stability. Thus, the LGR systems provide fast measurements needed for EC flux calculations while the Picarro serves as an accuracy standard. For the 2016 deployment, a Picarro G2401-m replaced the G1301-m for the last four flights following a power supply failure on the latter. The specifications and accuracy are very similar for these two instruments, though the (newer) G2401-m offers a faster data rate ($\sim 0.5 \text{ Hz}$) and better precision (based on in-flight comparison with the LGRs, Fig. 2d). Data streams from all analyzers, along with a GPS feed, are recorded via RS232 using a National Instruments CompactRIO controller.

The external gas inlet is a $6''$ length of $0.5''$ OD stainless steel tube (Fig. 1e). The tip of the inlet features a 15° rear-facing bevel to reject large particles. The inlet is mounted on an access hatch on the starboard side roughly 3 m aft of the nose. Directly behind the mounting plate, a tee connects to two identical lengths of Teflon PFA tubing ($0.375''$ ID, 5.2 m

length). Each tube terminates at a high-flow inline Teflon particle filter (Entegris WGMXMBSS4) before connecting to one of the two LGR sample gas inputs. The Picarro analyzer sub-samples at 0.4 SLM from the LGR CH₄/H₂O line. For typical flow rates and low-altitude flight, this configuration gives a gas sampling line residence time of 0.7 s and a Reynolds number of 4600.

5 Post-processing of GHG data occurs in several steps. First, data from all instruments are roughly time-aligned (to within ~1 s for the Picarro and ~0.1 s for the LGRs) using both the internal timestamp of the data acquisition system and the GPS timestamp. Next, CO₂ and CH₄ observations are converted to dry mixing ratios using native H₂O measurements (LGR CO₂ is corrected using H₂O from the LGR CH₄/H₂O sensor, time-aligned via lag-correlation). The corrections, which account for both density and spectroscopic effects, follow the quadratic form suggested for the Picarro G1301-m (Chen. et al., 2010; Rella, 2010) but with laboratory-derived, instrument-specific coefficients (Fig. [S3S5](#)). Picarro CO₂ and CH₄ mixing ratios are calibration-corrected via small scaling factors (1.0026 for CO₂, 0.9994 for CH₄) based on comparisons to a NIST-traceable certified gas standard (NOAA ESRL) pre- and post-mission. Picarro G1301-m water vapor is calibration-corrected following the manufacturer's recommendation (Rella, 2010). Finally, for each flight, LGR observations of CO₂, CH₄ and H₂O ~~are are lag-correlated and~~ linearly transformed to optimize agreement with the Picarro data. The transformation requires several operations, including 1) averaging dry mixing ratios to a common 1-Hz time-base, 2) smoothing LGR data to match the slower cell throughput of the Picarro, 3) time-lagging the LGR (typically < 2 s) to optimize correlation with the Picarro, 4) calculation of fit coefficients for an ordinary least-squares fit (LGR = m*PIC + b), and 5) correction of LGR dry mixing ratios using the fit parameters. This procedure, akin to performing a flight-by-flight span and intercept calibration correction, corrects for/rectifies calibration errors and minor-flight-to-flight drift that may occur in the LGRs under different operating conditions. Figure 2 shows linear fits for all flights from 2016 (results ~~are similar~~ for 2017 shown in SI Fig. S6). Linear correction factors vary little from flight-to-flight. One exception to this procedure occurred on Flight 5 (16 Sept 2016), where a narrow concentration range gave an atypically poor water fit (Fig. 2d). Comparison of H₂O observations between the LGR, Picarro, and DLH instruments (Sect. 4) corroborates this result. For this flight, the campaign-average slope and intercept were used to calibration-correct LGR H₂O. Campaign-average fit parameters were also used for two of the 2017 flights due to a failure in the Picarro system.

It is not currently feasible to calibrate the LGR systems in-flight due to high gas flow rates. Due to our correction procedure, we estimate that the accuracy limit of LGR CO₂ and CH₄ is degraded by a factor of 3 and 4, respectively, compared to the Picarro (Table 1). It may be possible to expand the above correction method to account for potential in-flight variability in LGR accuracy; however, as discussed later, measurement accuracy is a negligible contributor to total flux uncertainty for greenhouse gases.

2.5 Surface Imaging System (SIS)

The nadir viewing Surface Imaging Suite (SIS) consists of three cameras: a high resolution digital visible color camera (Nikon D7000), a thermal imager (FLIR A325sc), and a multiband camera specifically intended to observe vegetative health

(MicaSense RedEdge). The Nikon D7000 is a single lens reflex camera incorporating a “DX” format 16.2 Megapixel CMOS imaging sensor (1.5X “crop factor”) and 28 mm focal length lens. The lens was chosen to maximize field of view while minimizing distortion. The shutter speed is generally set to 1/1000 second with an ASA/ISO of 640. Aperture is controlled by the camera, the manual focus is fixed at infinity, and images are captured on two SD memory cards. The FLIR A325sc is an uncooled microbolometer with sensitivity covering the range of 7.5 to 13.0 μm and a 320 x 240 imaging sensor. The lens focal length is 9.7 mm, providing a 45° x 34° field of view. Data is captured by small notebook computer. The MicaSense RedEdge is a five-band imager covering the spectral range of approximately 460 to 860 nm with blue, green, red, near-infrared, and red-edge sensors. The camera has fixed focus and fixed field of view, and images are captured internally on an SC memory card. All three cameras are set for 1-s recording intervals, and time tagging and geolocation is done by remote GPS. Aircraft power (0.6 A) eliminates battery charging requirements, and the system is essentially autonomous. All imagery is downloaded after each flight and archived by the Wallops Remote Sensing Group. Figure 3 illustrates typical products acquired from this system.

The primary purpose of the SIS is to provide real-time qualitative information on surface characteristics and features that may influence gas and energy exchange. A rigorous comparison between imagery and airborne fluxes is non-trivial, as the flux footprint extends upwind of the aircraft with a typical half-width of several km, often exceeding the (altitude-dependent) swath width of the imagers. It is possible that an assumption of local surface homogeneity may be valid in some situations. Further work is needed to fully exploit the potential of combined surface imaging and in situ observations.

3 Flux calculations

The flux methodology for CARAFE builds on previous work in airborne and ground-based EC. All calculations utilize a custom MATLAB toolbox, available upon request. Wavelet calculations utilize the framework described in Torrence and Compo (1998). [The following discussion references both the time and spatial domains as appropriate, the two coordinates being linked by leg-average aircraft speed.](#)

3.1 Pre-processing

Data from the TAMMS, DLH, and GHG systems are averaged or re-gridded to a common 10 Hz time base. De-spiking is generally not necessary as data are quality controlled prior to archiving. [Gas concentrations are provided as dry mixing ratios, eliminating the need for density corrections](#) (Webb et al., 1980) [to derived fluxes](#). Individual flux legs are identified through inspection of heading, aircraft attitude, and vertical wind speed. We require that each leg be relatively level (± 20 m altitude above ground). [This altitude window is based on observed variability in the 2016 and 2017 data-sets and is not a hard limit.](#) Data are also filtered to exclude aircraft roll exceeding 5 degrees to minimize potential artifacts in the vertical wind measurement. Wind vector rotation is not required as winds are already reported in a geodetic reference frame.

For each leg and each scalar time series, scalar data undergo mean removal and lag-correlation to the vertical wind measurement. Lag times, with a typical range of 0—0.5 s, are determined through inspection of lag-covariance plots and held constant for each flight. Following Mauder et al. (2013), lag-covariance functions are calculated using fast Fourier transforms, and frequencies below 0.02 Hz (spatial scales greater than ~4 km) are removed to limit the influence of scalar trends on (co)variances. Note that these filtered lag-covariance functions are used for lag time determination and error calculations but not for calculation of actual fluxes.

Planetary boundary layer depth (z_i), required for calculation of flux errors and footprints, is assessed through examination of gradients in potential temperature and water vapor mixing ratio during vertical soundings. Due to mission constraints, soundings are not always located in the center of the target area or only occur at the beginning or end of a flight, making it difficult to quantify spatial or temporal variability in z_i . Thus, we assume a constant value for each target area. The uncertainty in this approximation is roughly ± 100 m based on observed variability in BL depth for flights with multiple soundings. Determination of exact BL depth is not critical, as [calculated fluxes do not depend on BL depth and](#) flux errors and footprints scale as a fractional power of z_i .

3.2 Ensemble average flux

Traditional ensemble average (EA) fluxes are calculated for each leg using vertical wind speed w and scalar s with the standard formulation.

$$F_{EA} = \langle w's' \rangle \quad (1)$$

Primes denote deviation from the mean. In addition, we derive spectra and cospectra using fast Fourier transforms (FFT) for each leg. These calculations are primarily for comparison with wavelet-derived quantities and are not used for scientific analysis, as terrain is often heterogeneous and observations are often non-stationary. A stationarity quality flag, q_{stat} , is calculated using the criteria of Foken and Wichura (1996):

$$q_{stat} = |1 - \overline{F_{EA,5}}/F_{EA}| \quad (2)$$

Here, $F_{EA,5}$ represents the mean of EA fluxes calculated on five evenly-sized sub-intervals within a leg. Ideally stationary legs will give $q_{stat} = 0$.

3.3 Continuous wavelet transform flux

The continuous wavelet transform (CWT) is a powerful and popular tool for time series analyses in atmospheric science. For a time series $x(t)$, wavelet coefficients $W_x(a,b)$ are calculated as a function of location (time or distance) and scale (frequency or wave number) by convolving the time series with a wavelet function (ψ).

$$W_s(a,b) = \int_{-\infty}^{\infty} s(t)\psi_{a,b}(t)dt \quad (3)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi_0\left(\frac{t-b}{a}\right) \quad (4)$$

Here, the size and location of the wavelet are determined by the scale (a) and translation (b) parameters, respectively. Normalization of the wavelet function by $a^{-1/2}$ preserves the energy of the wavelet at different scales (Torrence and Compo, 1998). The CWT cospectrum is defined here as the cross-wavelet power of w and s , $|W_w W_s^*|$, normalized by wavelet scale to correct for bias (Liu et al., 2007). We note that the latter operation is functionally equivalent to the common practice of multiplying FFT cospectra by frequency.

The wavelet function is actually a family of functions stemming from a “mother” wavelet, ψ_0 . Mother wavelets are typically chosen based on the application, but a defining feature is localization in both the time and frequency (or distance and wave number) domains. This property, combined with scaling and translation, permits the wavelet transform to deconvolve contributions to time series variance along both the time (distance) and frequency (wave number) domains. In this work we utilize the Morlet wavelet, which is a plane wave modified by a Gaussian:

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2}, \quad \omega_0 = 6 \quad (5)$$

The Morlet wavelet is the standard choice for eddy covariance calculations, giving reasonable localization in both time and frequency domains (Schaller et al., 2017).

The CWT is well-suited for airborne fluxes, offering several ~~key~~ advantages over traditional EA and FFT. Application of the CWT does not require stationarity, a condition that may be violated during long flight legs. This property also eliminates the need for signal detrending, improving quantification of long-wavelength flux contributions (Mauder et al., 2007). The technique provides a time series of fluxes along a flight track, removing the need to block-average homogenous sub-sections of the flight and giving relatively fine surface resolution that is essential when surveying patchy terrain. Further descriptions of CWT applications to airborne fluxes can be found elsewhere (Desjardins et al., 2018; Karl et al., 2009; Kaser et al., 2015; Mauder et al., 2007; Metzger et al., 2013; Misztal et al., 2016; Misztal et al., 2014; Vaughan et al., 2016; Wolfe et al., 2015; Yuan et al., 2015).

Figures 4 and 5 illustrate a typical set of CWT flux results for CO₂ using observations over the Great Dismal Swamp, VA. The transect included sampling over forest, bog, and a small lake. The location of the lake at 15–22 km (distance relative to the start of the transect) is evident in the reduced variability of vertical wind speed (Fig. 4a). As expected, cospectral power (Fig. 4b) is mostly negative, enhanced over land (beginning and end of leg), and diminished over the lake. Integrating the local wavelet cospectrum over all scales yields the flux time series (Fig. 4c). In this example, the EA flux (also shown in Fig. 4c) is within 4% of the leg-average wavelet flux. This agreement can vary considerably from leg to leg, and in general we find that the agreement of EA and CWT fluxes is well correlated with stationarity (Fig. S75). Thus, comparison of EC and CWT fluxes may not be a useful quality metric for wavelet fluxes as suggested by some previous studies (Misztal et al., 2014). “Instantaneous” 10 Hz fluxes exhibit significant variability due to both instrument noise and the random nature of turbulence, and some averaging is required to obtain an estimate of “true” surface fluxes (discussed further below).

Averaging the local wavelet cospectrum along time/distance gives the global cospectrum (Fig. 5). Ogives (cumulative cospectra) are also shown. In this example, 90% of the cospectral power occurs at eddy scales of 2500–100 m.

in the range of 0.03–0.7 Hz, corresponding to eddy scales of 2500–100 m sample frequencies of 0.03–0.7 Hz for an aircraft speed of 8075 m s⁻¹. The comparable FFT cospectrum shows the same features but with more noise. The ogive also indicates that 99% of flux-carrying eddies occur below at wavelengths longer than 54 m (54 m frequencies faster lower than 1.4 Hz). This point serves as a reminder that fast instrument time response is less critical for sampling turbulence in the mixed layer as compared to the surface layer. Such measurement requirements scale with platform speed: on a faster-moving aircraft, an instrument must sample faster to resolve the same eddy scales. Thus, fast sampling requirements are less stringent at altitude than at the surface and a true 10-Hz measurement, while desirable, is not strictly necessary. On the other hand, eddy scales increase with altitude and legs must be long enough to capture all modes of turbulent variability. Figure 5 shows a secondary peak in cospectral power at 0.0075 Hz, or 10 km.

3.3.1 Cone of influence

The cone of influence (COI) is the spectral region where wavelet coefficients may contain artifacts due to edge effects. As shown in Fig. 4b, the COI encompasses all scales at the ends of the time series and tapers toward larger scales (larger wavelengths/lower frequencies) near the center. Following the definitions of Torrence and Compo (1998) for the Morlet wavelet with $\omega_0 = 6$, the COI threshold scales linearly with distance from the beginning or end of the time series with a scaling factor of 0.73. This is an important consideration in flight planning. For example, if the largest flux-carrying eddies are ~5 km as in Fig. 5, the COI will fall below this value at distances greater than 5 km/0.73 = 6.8 km from leg edges. Thus, flight legs should be padded at least this distance on either end to ensure fully-resolved fluxes over a target area.

Treatment of the COI can impact scale-integrated CWT fluxes. Some previous studies exclude the COI before calculating fluxes (Vaughan et al., 2016), while others include it (Misztal et al., 2014). For short-length or higher-altitude flight legs, neglecting cospectral power within the COI may create systematic errors due to exclusion of larger-scale flux contributions. On the other hand, inclusion of covariance within the COI can lead to spurious fluxes, especially near the ends of a leg and when fluxes are small. Here we develop a quality flag to quantify the potential impact of the COI and examine the effects of including or excluding the COI in scale-averaged fluxes.

The quality flag, q_{COI} , is calculated by interpolating global ogives (e.g. Fig. 5) onto the time series of COI threshold (given in wavelength or frequency units). Mathematically, for time t and frequency/wavelength λ ,

$$q_{COI}(t) = ogive_{w,s*}(\lambda f = COI(t)) \quad (6)$$

For this calculation, the ogive is calculated by integrating the absolute value of cospectral power to avoid negative values. Ranging from 0 to 1, this flag roughly represents the fraction of global cospectral power within the COI at each point in time. For example, for the data shown in Fig. 4b, the COI threshold at a distance of 5 km (time of 60 sec) occurs at a frequency/wavelength of 0.022 Hz/3.7 km. The global ogive value at this frequency/wavelength is 0.12 (Fig. 5), indicating that 12% of the global cospectral power lies at lower frequencies, and thus $q_{COI} = 0.12$ for this time/location. The q_{COI} flag can be used to

filter scale-averaged CWT fluxes; for example, crosses in Fig. 4c indicate fluxes with $q_{COI} > 0.5$. Thus, more than half of the cospectral power resides within the COI for these times.

Figure 6 compares CWT and EA fluxes to quantify the potential impacts of COI treatment. This analysis is restricted to legs that are highly stationary ($q_{stat} < 0.1$) to ensure the quality of EA fluxes. Scale-averaged CWT fluxes are calculated by either including or excluding the cospectral power within the COI (see Fig. 4b). The resulting flux time series is subsequently filtered using a specified maximum q_{coi} prior to leg-averaging. For example, a max q_{coi} of 0.2 indicates that all points in the time series with more than 20% of cospectral power within the COI are excluded from the leg-averaged flux. CWT fluxes systematically under-predict EA fluxes by as much as 14% on average when omitting the COI and agree within 5% when including the COI. Thus, the extra systematic error from exclusion of the COI in the case without a q_{coi} filter (rightmost points in Fig. 6) is 9%. Addition of the q_{coi} filter reduces the CWT-EA discrepancy by removing CWT fluxes near leg edges where the COI influence is strongest. More restrictive filters (lower maximum q_{coi}) improve agreement at the expense of data density. For example, filtering for CWT fluxes with q_{coi} less than 0.6, 0.4 and 0.2 removes 6%, 12% and 28% of 1 Hz CWT fluxes, respectively. We choose to include the COI when calculating CWT fluxes and assume fluxes are valid for $q_{coi} < 0.5$. From Fig. 6, this choice may reduce leg-average fluxes by $\sim 2\%$ on average.

3.3.2 Data gaps

Wavelet analysis is inherently designed for contiguous data, but data gaps are inevitable in field observations. Such gaps most commonly result from instrument calibrations or unfavorable aircraft attitude (e.g. evasive pitch or roll to avoid avian hazards). Since the wavelet algorithms of Torrence and Compo (1998) rely on fast Fourier transforms, these gaps must be removed or filled prior to performing the CWT. Several studies have suggested procedures for modifying wavelet basis functions to handle gaps (Frick et al., 1997; Frick et al., 1998; Mondal and Percival, 2008), but it is not clear how to implement such methods within our framework. SI Section S2 describes an empirical method that utilizes the covariance of scalar and vertical wind speed fluctuations to fill gaps with projected values. Even this method introduces some error in wavelet fluxes, especially in the immediate vicinity of a gap (Fig. S86). Transient errors are typically below 30% with a spatial extent that scales with the width of the gap. To be conservative, we discard wavelet fluxes within a gap and on either side of a gap out to a distance equal to the gap width. In practice, gaps are rare and this procedure has a minimal impact on the total flux dataset.

3.4 Uncertainties

Uncertainty in EC fluxes arises from measurement limitations, sampling strategies, and the fundamental nature of turbulence. When extrapolating airborne fluxes to the surface, uncertainties in vertical flux divergence must also be considered. Methods to quantify flux errors are reviewed elsewhere (Langford et al., 2015; Mauder et al., 2013; Rannik et al., 2016). Here we leverage a combination of these methods and suggest a new technique to quantify the total random error in wavelet fluxes. Figure 7 summarizes individual errors terms for the 2017 field campaign for all legs at altitudes below 200

m. Note that fluxes of CH₄ were negligibly small for most of the mission, thus the fractional errors are biased high. When flying over methane source regions (e.g. wetlands), CH₄ flux errors are comparable to those in CO₂ flux. In the following discussion, we adopt the convention of defining systematic errors as a fraction or percentage of the flux and random errors as an absolute value (in flux units). “Typical” error ranges quoted in the text refer to interquartile ranges (upper and lower box boundaries in Fig. 7) and are 1σ.

3.4.1 Systematic errors

Under-sampling of turbulent fluctuations at both low and high frequencies creates systematic flux errors. For airborne observations, Lenschow et al. (1994) derive an upper limit for systematic error fraction due to under-sampling of low frequencies (long wavelengths) as a function of altitude above ground level (z), boundary layer depth (z_i), and leg length (L).

$$SE_{turb} \leq 2.2 \left(\frac{z}{z_i}\right)^{0.5} \frac{z_i}{L} \quad (7)$$

Typical SE_{turb} ranges from 1.6% to 63.3% of observed fluxes (Fig. 7, light gray). The CWT utilizes data from the whole leg and thus theoretically captures all resolvable long-wavelength flux contributions at any given point in time. Thus, we assume this fractional error is constant for each point within a leg, irrespective of averaging length (e.g., a 1-km average is assigned the same SE_{turb} as the whole leg).

Limited instrument time response is the main cause of high-frequency systematic errors for the CARAFE payload. The open-path DLH instrument does not suffer from such limitations and thus serves as a time response standard. The ratios of cospectra for temperature, CO₂, CH₄ and LGR H₂O fluxes against DLH H₂O cospectra exhibit a characteristic decay at high frequencies that is well described by the transfer function (Horst, 1997)

$$H_s(f) = \frac{c_{w,s}(f)}{c_{w,DLH}(f)} = \frac{1}{1+(2\pi\tau_s f)^2} \quad (8)$$

Here, f is natural frequency and τ_s is a characteristic response time for scalar s . Fitting Eq. (8) to global wavelet cospectra for all legs gives typical response times of 0.09 s for temperature and CO₂ and 0.1 s for CH₄ and LGR H₂O, consistent with lab tests (Fig. S4). Systematic error fraction due to response time is calculated for each leg as

$$SE_{RT} = \frac{\int_{-\infty}^{\infty} H_x(f) c_{w,s}(f) df}{\int_{-\infty}^{\infty} c_{w,s}(f) df} - 1 \quad (9)$$

~~Values~~ Typical values for SE_{RT} are ~~typically less than 10%~~ 2 – 7% of observed fluxes (Fig. 7, light green), ~~and are~~ assumed constant for each leg. ~~When fluxes are small, Eq. (9) can amplify high-frequency noise and SE_{RT} may become unrealistically large. Thus, reported fluxes are not explicitly corrected for this error.~~

Accuracy in both the vertical wind and scalar measurements (Table 1) directly propagates into calculated fluxes. This systematic error (SE_{acc}), of unknown sign, adds uncertainties of 5% to CO₂ and CH₄ fluxes, 7% to sensible heat and DLH latent heat fluxes, and 8.6% to LGR latent heat flux (Fig. 7, dark green).

Systematic errors can be applied as a correction factor to fluxes (if of known sign) or be included as part of the total uncertainty. Both practices are common among the airborne flux community (Gioli et al., 2004; Misztal et al., 2014). For the

errors discussed above. SE_{acc} is of unknown sign, while SE_{turb} and SE_{RT} should both increase the flux. We are, however, reluctant to employ the latter two as correction factors. SE_{turb} represents an upper limit and thus may slightly “over-correct” the fluxes, while SE_{RT} can become unrealistically large when fluxes are small due to the amplification of high-frequency noise by Eq. (9). Furthermore, systematic errors are typically small compared to random errors. Thus, we elect to include all systematic errors in the total flux error and assume all error components are symmetric for simplicity. Total systematic error (SE_{tot}), given as a fraction of the flux over any interval, is then the root-sum-square of SE_{turb} , SE_{RT} , and SE_{acc} . Total systematic error is reported as a separate variable in flux archive files and may be used as part of the total error or as a correction factor (after removing the accuracy contribution) at the discretion of data end-users. ~~Systematic error components are assumed symmetric for simplicity.~~ One exception to this procedure is the treatment of systematic under-sampling of vertical wind variance in the 2016 observations (Sect. 2.2). For this particular issue, all 2016 fluxes are corrected upwards by a factor of 1.32 (Fig. S2) and additional random error is incorporated as discussed below.

3.4.2 Random errors

Random flux errors arise from both uncorrelated noise in scalar measurements and the stochastic nature of turbulence. Errors due to uncorrelated instrument noise are calculated as

$$RE_{noise} = \sqrt{\sigma_{s,noise}^2 \sigma_w^2 / N} \quad (10)$$

Here, σ_w^2 is the variance of vertical wind speed over a whole leg and N is the number of points. The noise variance in scalar s , $\sigma_{s,noise}^2$, is derived from the auto-covariance function as described elsewhere (Langford et al., 2015; Lenschow et al., 2000; Mauder et al., 2013). Relative to the total flux, this error is typically less than 405% for sensible and latent heat, 4–8% for CO_2 flux, and 11–96% for CH_4 flux but can be up to 50% for CO_2 and over 100% for CH_4 (Fig. 7, indigo), particularly with higher errors occurring when fluxes are small. For turbulence sampling errors in airborne fluxes, Lenschow et al. (1994) suggest the following upper limit (as a fraction of total flux):

$$\frac{RE_{turb}}{F} \leq 1.75 \left(\frac{z}{z_i}\right)^{0.25} \left(\frac{z_i}{L}\right)^{0.5} \quad (11)$$

Values for RE_{turb} typically range from 4015% to 3021% of leg-average flux (Fig. 7, dark gray). This equation suggests several strategies for reducing random errors, including reduced altitude, averaging over distance, or averaging over repeated legs.

The total random error is the square root of the sum of squared errors from Eq. (10) and (11). Alternatively, the total error can be estimated empirically as the variance of the covariance between s and w . (Finkelstein and Sims, 2001).

$$RE_{FS01} = \sqrt{\frac{1}{N} \sum_{p=-m}^m (s' s'_p \overline{w' w'_p} + s' w'_p \overline{w' s'_p})} \quad (12)$$

Arguments $\overline{s' w'_p}$ and similar represent the unbiased cross-covariance or auto-covariance for lag p . Finkelstein and Sims (2001) suggest taking the summation over a sufficiently large m to capture the integral time scale, somewhere in the range of

10—40 seconds. Mauder et al. (2013) further note that summing over too wide a range can give unrealistic results if a time series contains trends. Following Mauder et al. (2013), we calculate the lagged cross- and auto-covariance functions using fast Fourier transforms and remove frequencies below 0.02 Hz ([spatial scales > 4 km](#)) to limit trend effects. The maximum lag for the summation is set to 10 seconds based on comparison with the root-sum-square of RE_{noise} and RE_{turb} , [the latter representing a theoretical upper limit for total random error \(Fig. S9a\)](#).

All of the above methods were originally developed for traditional EA eddy covariance, and it is not immediately obvious how to extend error calculations to time-resolved CWT fluxes. Several previous studies (Karl et al., 2009; Misztal et al., 2014; Vaughan et al., 2016) have defined the random error for distance-averaged wavelet fluxes (1—10 km means) by substituting the averaging length for leg length (L) in Eq. (11). There are several issues with this approach. First, it does not account for errors due to instrument noise, which can comprise a significant fraction of the total error when turbulence-driven scalar variability approaches measurement precision. Second, Eq. (11) assumes that the flux calculated over distance L uses only observations acquired over that same window; however, the CWT inherently integrates information from the entire leg to derive (co)variances. In other words, the CWT flux for a 1 km region is not equivalent to the EA flux calculated using only wind and scalar observations sampled within that same 1 km.

We propose a novel method to calculate total random error along a wavelet flux time series. The technique is essentially an extension of Eq. (12), with lagged cross- and auto-covariance calculated using the scale and time-dependent wavelet coefficients. In analogy to Eq. (24) of Torrence and Compo (1998), the covariance between variables s and w for time t and lag p is given by

$$\overline{s'w'_p}(t) = \frac{\delta j \delta t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{|W_s(a_j, t) W_w^*(a_j, t+p)|}{a_j} \frac{N}{N-|p|} \quad (13)$$

Here, δj is the wavelet spacing parameter (0.25 in our case), δt is the sample interval (0.1 s for 10 Hz data), C_δ is a wavelet-specific reconstruction factor (0.776 for Morlet), and a_j is the wavelet scale at index j . The last term on the right gives an unbiased covariance estimate. When taken over multiple lags, Eq. (13) defines a wavelet cross-covariance function between s and w at each point in time (Fig. [S7S10](#)). Analogous equations apply for the other cross and auto-covariance terms. Combining Eq. (12) and (13) thus gives the wavelet random error, RE_{wave} , as the variance of covariance along the time series. For this calculation the $1/N$ term in Eq. (12) is neglected, as the error is calculated for a single point in the time series and not the whole leg. As with the ensemble application of Eq. (12), the summation in Eq. (13) is restricted to scales with a corresponding Fourier frequency greater than 0.02 Hz ([spatial scale < 4 km](#)) and uses a lag range of ± 10 s. This operation is computationally expensive for 10 Hz data, thus wavelet coefficients are averaged to 1 Hz before calculating RE_{wave} . This procedure yields comparable results to calculating errors from 10 Hz wavelet coefficients and then averaging to 1 Hz.

Two internal checks validate this approach. First, leg-averaged wavelet cross- and auto-covariance functions are virtually identical to those from ensemble-based lag calculations (Fig. [S107](#)). Second, leg-averaged RE_{wave} values agree relatively well with both RE_{FS01} and the root-sum-square of RE_{turb} and RE_{noise} (Fig. [S9b-c8](#)). Typical RE_{wave} values range from 10—50% of observed fluxes (Fig. 7, cyan).

As noted in Sect. 2.2, the 2016 dataset includes an additional random error component from damping of vertical wind variance. Using 2017 fluxes derived from the two wind datasets (Fig. S2), we estimate 1σ random errors in 2016 1Hz fluxes for sensible heat, latent heat, CO_2 and CH_4 of 50 W m^{-2} , 110 W m^{-2} , $7 \mu\text{mol m}^{-2} \text{ s}^{-1}$, and $50 \text{ nmol m}^{-2} \text{ s}^{-1}$, respectively. This error is added directly to RE_{wave} rather than in quadrature, as it is not unambiguously independent of RE_{wave} (the latter being based on vertical wind and scalar (co)variance). The extra uncertainty does, however, reduce with averaging; note the tighter correlation for leg-average fluxes in Fig. S2b.

3.4.3 Total error

~~Total systematic error (SE_{tot}), given as a fraction of the flux over any interval, is the root sum square of SE_{wave} , SE_{ed} , and SE_{acc} . Systematic error components are assumed symmetric for simplicity. The total flux error is the root sum square of SE_{tot} (multiplied by flux) and the random error, RE_{wave} . When time averaging fluxes within a leg, fractional systematic errors are assumed constant while absolute random errors reduce as the mean of the root sum square RE_{wave} for each point in the average (see Eq. (1) in Langford et al. (2015)).~~

$$\overline{RE_{\text{wave}}} = \frac{1}{N} \sqrt{\sum RE_{\text{wave}}^2} \quad (14)$$

Total (random + systematic) errors in 1 Hz fluxes generally range from 10–90% (Fig. 7, black).

~~In comparing CWT fluxes with tower based or modeled surface fluxes, additional errors arise from the divergence correction and footprint calculation (discussed below). Fractional uncertainty in the divergence correction spans a wide range, from less than 2% to greater than 100% (Fig. 7, orange). This error stems from uncertainty in the linear fits and generally scales with altitude. Uncertainties due to footprint mismatch are not easily defined, though we note that a rigorous comparison to surface values must include a bound on this error term.~~

3.5.4.3 Vertical flux divergence

Extrapolation of airborne fluxes to the surface requires accounting for the change of flux with altitude, which should exhibit a near-linear profile in the boundary layer for non-reactive species. Vertical flux divergence can arise from several processes, as illustrated when considering the budget equation for a scalar s :

$$\frac{\partial F}{\partial z} = -\frac{\partial s}{\partial t} - \bar{U} \frac{\partial s}{\partial x} + Q \quad (15)$$

Terms on the right-hand side respectively represent storage, horizontal advection (the product of horizontal wind speed and concentration gradient), and net in situ production or loss, which is negligible ~~here~~ for the long-lived GHGs. We ~~have neglected the~~ generally small contributions from vertical subsidence and horizontal turbulent fluctuations (Karl et al., 2013). It is possible to constrain each of the right-hand terms with a carefully-designed flight plan (Karl et al., 2013; Kawa and Pearson, 1989). Alternatively, flux measurements at multiple altitudes in the boundary layer provide a means of directly

quantifying the flux divergence slope. Here we describe a procedure for deriving divergence corrections and discuss some of the associated challenges.

Calculation of the divergence correction begins with selection of a subset of flux observations. Figure 8 shows an example sensible heat flux profile for a series of 15 legs flown over the same 43 km forest swath (within a cross-track horizontal spread of ~2 km). ~~This is the only flight from either mission with multiple legs at 3 different levels.~~ Most legs on a typical CARAFE flight occur at low altitude (90—150 m), with only 1 to 3 legs at higher altitudes (200—400 m). Ideally the upper-level legs would be situated in the upper half of the boundary layer ($z/z_i > 0.5$), but this is not always possible due to flight restrictions and the difficulty of determining boundary layer depth in real time. Fitting and error estimation requires a minimum of 3 legs and 2 altitudes. Chosen by visual inspection, these legs must be relatively close to one another in both space and time for a reliable fit. Diurnal variability is evident at low altitude over the 3 hour afternoon flight shown in Fig. 8, and in this case we limit the fit to legs with a solar zenith angle of less than 38°. CWT fluxes are filtered prior to leg-averaging using both the COI quality flag ($q_{coi} < 0.5$) and a “proximity” filter. The latter requires that each point within a leg-average lies within 1 km of at least one point in every other leg. This proximity filter effectively trims the ends of each leg and limits the spatial average to regions of overlap. The choice of a 1 km radius is somewhat arbitrary and is a compromise between spatial overlap and data density, though we note that this is also a typical scale for a flux footprint. In some instances the proximity filter cannot be applied. For example, the fit in Fig. 8 uses legs from both the “east” and “west” tracks. Though these tracks are spaced ~2 km apart, the forest is fairly homogeneous and the fit is more robust with the inclusion of more points.

Following data selection, divergence correction factors are calculated as follows. First, an error-weighted least-squares fit of the filtered and leg-averaged flux versus altitude gives the slope, m , and intercept, b . Next, a scaling factor C_{div} is calculated as a function of altitude, with associated [random](#) uncertainty derived from fit parameter uncertainties (δm , δb) using standard error propagation.

$$C_{div}(z) = \frac{F(0)}{F(z)} = \frac{b}{mz+b} \quad (4615)$$

$$\delta REC_{div} = \left(\left(\frac{\partial C}{\partial m} \delta m \right)^2 + \left(\frac{\partial C}{\partial b} \delta b \right)^2 \right)^{0.5} = \frac{z}{(mz+b)^2} (\delta m^2 b^2 + \delta b^2 m^2)^{0.5}$$

(4716)

~~Multiplication of the CWT fluxes by the associated C_{div} then gives the surface-extrapolated flux.~~ As shown in Fig. 7, the divergence scaling factor can vary considerably. “Typical” correction factors ~~alter-rescale~~ fluxes sampled below 200 m by 10—10050%, and the additional ~~uncertainty-random error~~ from this correction ~~lies in a similar range~~ is typically 5—30%. ~~Multiplication of the CWT fluxes by the associated C_{div} then gives the surface extrapolated flux.~~ Divergence correction factors and their associated uncertainties are reported alongside CWT fluxes in archived data files.

The application of a “bulk” divergence correction inherently assumes that the correction is relatively invariant (within uncertainties) in both space and time for a given target region. It is possible to empirically test this assumption with observations. Figure [S9a-S11a](#) illustrates our tests for temporal and spatial variability in the divergence correction. Four of

the 2017 flights included two high-level legs spaced 1.5—2.2 h apart, allowing derivation of divergence corrections at two different times for the same region. For this test we separately fit two legs (one high and one low) and calculate C_{div} for each of the two sub-periods, then compare this to the C_{div} value derived from all four legs combined. Fits are done on fluxes of CO_2 , temperature, DLH H_2O , and LGR H_2O , giving a total of 16 scalar-flight pairs and 32 test cases. Based on this test, 95% of the sub-period C_{div} values differ by less than 22% from the full-flight C_{div} (Fig. S9bS11b). To assess spatial variability, we divide each set of legs into two sub-regions of equal length, calculate the divergence correction in each of the sub-regions, and compare this to C_{div} calculated for the full region. Using the same set of fluxes as described above, this test includes 4 species, 4 flights, 2 divergence pairs per flight and 2 sub-regions per pair for a total of 64 test cases. Spatial variability is larger than temporal variability, with 95% of the sub-region C_{div} values differing by less than 35% from the full region C_{div} (Fig. S9bS11b). In general, we find that the spatiotemporal variability of divergences corrections is within the calculated uncertainty for the divergence correction factor.

3.4.4 Total uncertainty

The total error for in situ fluxes is the root-sum-square of SE_{tot} (multiplied by flux) and the empirical random error, RE_{wave} . When extrapolating to the surface, the divergence error is also added in fractional quadrature. Thus, the total fractional error for surface fluxes is

$$\frac{E_{surf}}{F_{surf}} = \left(SE_{tot}^2 + \left(\frac{RE_{wave}}{F} \right)^2 + \left(\frac{RE_{div}}{C_{div}} \right)^2 \right)^{1/2} \quad (17)$$

For CO_2 , sensible heat and latent heat, typical E_{surf} values range from 16—35% of the leg-averaged flux (Fig. 7, black). Values for CH_4 flux are significantly higher when considering the whole campaign since most of our sampling occurred in regions with scant methane emissions. Over the Alligator River and Dismal Swamp wetlands, where CH_4 emissions were significant, leg-average E_{surf} values are 17 – 32% of the leg-average CH_4 flux.

The above values are based on the 2017 dataset. For 2016 fluxes, consideration of the additional random error due to vertical wind variance under-sampling increases E_{surf} by 1—2% of leg-averaged fluxes or 7—15% of 2 km average fluxes.

3.4.5 Error averaging

When averaging fluxes within a leg, fractional systematic errors are assumed constant while absolute random errors reduce as the mean of the root-sum-square of RE_{wave} for each point in the average (see Eq. (1) in Langford et al. (2015)).

$$\overline{RE_{wave}} = \frac{1}{N} \sqrt{\sum RE_{wave}^2} \quad (18)$$

Divergence corrections (C_{div}) are averaged directly, while RE_{div} is averaged as the root-mean-square (like Eq. (18) but without the extra factor of $N^{-1/2}$). Archived fluxes are reported at high resolution (1 Hz), but random errors due to turbulence sampling are large at these scales and some averaging is necessary to obtain statistically meaningful results. For example, the interquartile range in CO_2 surface flux uncertainty is 192—438% at the native 1 Hz (~80 m) resolution but improves to 40—

[90% when averaging to 2 km. These uncertainties are comparable to those reported in other airborne flux studies](#) (e.g., Vaughan et al., 2016). [Uncertainty reductions are also possible by averaging over repeated legs \(Sect. 4.3\).](#)

3.6.5 Footprints

5 The flux footprint defines the distribution of surface sources/sinks contributing to the net flux observed at a given point. A simple 1-D metric for footprint size is the half-width, $dx_{0.5}$, defined as the distance along the mean horizontal wind that contains 50% of the surface flux (Karl et al., 2013; Weil et al., 1992).

$$dx_{0.5} = 0.9 \frac{\bar{u} z^{2/3} z_i^{1/3}}{w^*} \quad (1819)$$

10 Here, w^* is the convective velocity scale. Footprint scales from this equation typically range from hundreds of meters to ~10 km. For the CARAFE missions, estimates of $dx_{0.5}$ are reported alongside 1 Hz fluxes.

Robust comparison with ground observations or gridded model output may require an estimate of the full 2-D footprint. In this case, the “footprint” is effectively a spatial weighting function that can be applied to spatially-resolved quantities prior to integration and comparison with observed fluxes. For tower comparisons, fluxes can be filtered for times/locations where both platforms sampled the same footprint area. Applications of footprint models for airborne flux analysis vary in complexity. For example, Misztal et al. (2016) define a series of circles with radii equal to $dx_{0.5}$ and used these areas to integrate model-derived surface fluxes (with equal weighting within the circle). Sayres et al. (2017) utilize the 1-D parameterization of Kljun et al. (2004), while others have augmented the latter with a cross-wind distribution function (Metzger et al., 2012; Metzger et al., 2013; Vaughan et al., 2016). The recent 2-D parameterization of Kljun et al. (2015) (hereafter K15) is an attractive next step, both because it is based on the same Lagrangian framework as its 1-D predecessor and because the Matlab code is freely available. All required inputs for this parameterization are available from the CARAFE flux system, and footprints are theoretically calculable at any resolution up to the native 10 Hz resolution of the data stream. Equation (198) and K15 give comparable footprint half-widths for typical low-level legs below 200 m altitude (results not shown). Agreement is primarily a function of the stability parameter, z/L_{ob} (L_{ob} being the Obukhov length). For higher-altitude legs in strong convective conditions ($z/L_{ob} < -2$), footprint half-widths from K15 can be 2 to 4 times larger than $dx_{0.5}$; K15 note that this regime approaches the limits of applicability for the parameterization. In future work with the CARAFE dataset, we will evaluate what level of footprint complexity is required for comparison with surface fluxes.

4 Performance

Here we present a subset of results that illustrate the quality and performance characteristics of CARAFE observations. This evaluation is not exhaustive, and future studies will continue to assess the quality of fluxes through both internal quality controls and, when possible, comparison to other observations.

4.1 Spectral analysis

Figure 9 exemplifies quality metrics for a low-level leg with appreciable fluxes in all measured species. All cross-covariance functions (Fig. 9a) display strong peaks with similar integral time scales of ~ 3 s (defined as the time at the first zero crossing). Power spectra (Fig. 9b) for temperature, DLH-H₂O and vertical wind speed measurements exhibit the $f^{2/3}$ power law ($f^{5/3}$ when not frequency-multiplied) in the inertial subrange, consistent with theory (Kaimal et al., 1972; Kaimal et al., 1976). In contrast, power spectra for CO₂, CH₄ and LGR-H₂O show a shallower decay and exhibit a slope of ~ 1 above 0.4 Hz, indicative of white noise. The effects of instrument noise are also reflected in the increased variability in the cross-covariance functions at longer lag times and in the higher values of RE_{noise} for these fluxes (Fig. 7). Despite the limited precision of the closed-path analyzers at higher frequencies, cospectra with vertical wind generally agree for all scalars (Fig. 9c).

4.2 Water comparison

The CARAFE payload includes two independent water vapor measurements, providing a unique inter-comparison opportunity. The DLH system is open-path, fast-response, and field-proven through numerous prior missions. The LGR system is closed-path, displays reduced precision at turbulence-relevant frequencies, and had not flown prior to the 2016 mission. Figure 10 compares 1 Hz water mixing ratios and latent fluxes derived from both instruments. LGR water mixing ratios exhibit a small positive bias (slope = 1.05, $r^2 = 0.995$) throughout the mixing ratio range sampled during the 2017 CARAFE mission (0.3—1.5% by volume). This difference is well within the uncertainties of both instruments. LGR latent heat shows a somewhat larger bias relative to DLH latent heat (slope = 1.13, $r^2 = 0.87$). The source of this extra $\sim 8\%$ bias in fluxes is unclear; it may be related to the contact of sample air with surfaces (such as the 5.2 m sample line), though we would generally expect gas-surface interactions to dampen concentration fluctuations and thus reduce the flux. The bias is small compared to typical flux uncertainties, thus we will not explore the issue further here. Similar results were obtained for the 2016 mission.

4.3 Repeatability

The stochastic nature of turbulence imparts substantial random errors into [individual-small-scale](#) flux measurements. For a typical 1 Hz (~ 75 m) flux, random errors are on the order of hundreds of percent. Averaging rapidly reduces this error; for example, random errors in a 2 km average flux are typically 30—50%. Precision is also improved by repeated sampling over the same swath of land, as long as changes in flux over the averaging period are small relative to random errors. [For example](#), Figure 11 shows CO₂ fluxes observed over five consecutive legs covering mixed farmland and forest. All legs were flown at an altitude of 100 m with a cross-track separation of less than 500 m. Typical $dx_{0.5}$ values are 800 ± 300 m, and leg-average fluxes vary by less than 12% with no discernable time trend. Random variability is evident in the 2 km average flux values. Nonetheless, all profiles exhibit the same general trend with lower fluxes (0 to $-15 \mu\text{mol m}^{-2} \text{s}^{-1}$) at the beginning of

the track, higher values (-10 to $-30 \mu\text{mol m}^{-2} \text{s}^{-1}$) near the midpoint, and a sharp decline at the eastern edge (near the Atlantic coast and more urban areas). Averaging all legs together in each 2 km bin further reveals this trend. The 1σ random error for the multi-leg average is $\sim 22\%$ of the flux. Roughly 47% of the individual 2 km averages are contained within 1σ of the mean, and 72% within 2σ . This is somewhat less than the 68%/95% expected for a normal distribution; however, we do not necessarily expect a Gaussian distribution. Furthermore, some leg-to-leg variability is expected due to changes in wind speed and direction and thus the flux footprint. Overall, this result provides some additional confidence in our random error estimates.

5 Conclusions

The NASA CARAFE project ~~represents an important step towards the adoption of~~ aims to incorporate eddy covariance fluxes as a standard component of the airborne science toolbox. The C-23 Sherpa is well-suited for EC due to its particular balance of range, speed, and payload, though any aircraft that is amenable to fast 3-D wind measurements is a viable platform. The instrumentation deployed on the 2016 and 2017 missions provided observations of sufficient quality to calculate fluxes of sensible and latent heat, CO₂, and CH₄. Continuous wavelet transforms are key to unlocking the full potential of airborne fluxes, but only if utilized within a framework that properly accounts for all sources of uncertainty and the peculiarities of the technique (notably, the cone of influence and vertical flux, divergence, etc.). Typical uncertainties in derived surface fluxes are 40–90% for a resolution of 2 km and 16–35% when averaged over an entire leg (typically 30–40 km). Initial results demonstrate sound spectral features of all measurements (with the exception of 2016 vertical winds), excellent agreement between closed and open-path water vapor observations, and reproducibility of horizontal flux gradients within random variability. ~~Despite decades of research, airborne flux methodology is, in some respects, less developed than its ground-based counterpart.~~ Future efforts must continue to refine measurement and analysis techniques by both leveraging earlier work and acquiring new observations over a variety of conditions and surfaces. Inter-comparison with other methodologies for quantifying surface exchange, where possible, would also be valuable for both performance diagnosis and evaluation of multi-scale flux variability.

Direct observations of carbon and energy fluxes at regional scales, ~~such as those provided by CARAFE,~~ offer a new vector offer unique opportunities for probing earth-atmosphere-biosphere interactions. This type of dataset is rare, and more work is needed to understand how such measurements can be applied to improve biophysical parameterizations and model or satellite-derived flux ~~inventories~~ estimates. In particular, the spatiotemporal scales of airborne flux measurements – snapshot pictures in time over regional areas – are very different from the long-term, but spatially-sparse, tower flux observations typically available to the GHG community. Future efforts with the CARAFE dataset will include detailed comparisons to both ground observations and high-resolution earth system models, with the dual goals of developing techniques to upscale flux observations and furthering process-level understanding of biosphere carbon exchange.

Potential applications of airborne fluxes extend far beyond the topics outlined here. Augmentation of the CARAFE payload with additional observations, such as mixing ratios and/or fluxes of carbonyl sulfide (Blonquist et al., 2011), ~~or CO₂/H₂O isotopes, would may synergistically~~ enhance the value of future datasets for diagnosing plant physiological responses. ~~Additional surface remote sensing capabilities~~ Co-located observations of surface properties, such as ~~mapping of vegetation characteristics or~~ solar-induced fluorescence and other markers of vegetation state/health, ~~could also prove synergistic for data interpretation~~ are also highly desirable. The methodology developed here is equally applicable to fluxes of reactive gases, including ozone, volatile organic compounds and oxidized nitrogen compounds. The process-level drivers of emission, deposition and transformation of these gases remain highly uncertain, and observational constraints on surface-atmosphere exchange are needed to challenge and improve air quality and chemistry-climate models. The combination of energy, carbon, and reactive gas fluxes may even provide new insights into the linkages between the biosphere, the atmosphere, and anthropogenic ~~activities impact on earth-atmosphere interactions.~~

6 Data and code availability

All observations, 1 Hz fluxes and related quantities are publicly available through the CARAFE mission page at <https://www-air.larc.nasa.gov/missions/carafe/index.html>. Flux analysis code is available upon request from the corresponding author and will eventually be refined into a publicly-available Matlab toolbox.

Acknowledgements

We are deeply indebted to Piers Sellers, whose vision and boundless enthusiasm enabled this effort. We thank the management, pilots, crew, engineers, and mission support staff of WFF and the C-23 Sherpa for their selfless support and flexibility. We gratefully acknowledge the data archiving services provided by Gao Chen and Ali Aknan (NASA LaRC). We thank Ray Desjardins (Environment Canada) for helpful discussions, Sally Pusede and Laura Barry (U. VA) for their assistance in operating the GHG instrument suite during the 2017 mission, and Dennis Gearhardt and Sean Kirby for their assistance with aircraft data acquisition. SIS team contributions from Ted Miles and Richard Mitchel are appreciated. Support for the CARAFE 2016 and 2017 missions was provided by GSFC Internal Research and Development, the NASA Carbon Monitoring System Program (NNH15ZDA001N-CMS), and the NASA HQ Earth Science Division. Wavelet software was provided by C. Torrence and G. Compo, and is available at <http://atoc.colorado.edu/research/wavelets/>. Finally, we thank the two anonymous referees for their insightful and thorough comments.

References

- Anderson, M., Norman, J., Kustas, W., Houborg, R., Starks, P., and Agam, N.: A thermal-based remote sensing technique for routine mapping of land-surface carbon, water and energy fluxes from field to regional scales, *Remote Sensing of Environment*, 112, 4227-4241, 2008.
- 5 Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas
- 10 monitoring efforts, *Atmos. Meas. Tech.*, 7, 647-687, 2014.
- Aubinet, M., Joly, L., Loustau, D., De Ligne, A., Chopin, H., Cousin, J., Chauvin, N., Decarpenterie, T., and Gross, P.: Dimensioning IRGA gas sampling systems: laboratory and field experiments, *Atmos. Meas. Tech.*, 9, 1361-1367, 2016.
- Barrick, J. D., Ritter, J. A., Watson, C. E., Wynkoop, M. W., Quinn, J. K., and Norfolk, D. R.: Calibration of NASA Turbulent Air Motion Measurement System, NASA Technical Reports, NASA-TP-3610, 1996.
- 15 Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO₂ fluxes estimated from GOSAT retrievals of total column CO₂, *Atmos. Chem. Phys.*, 13, 8695-8717, 2013.
- Blonquist, J. M., Montzka, S. A., Munger, J. W., Yakir, D., Desai, A. R., Dragoni, D., Griffis, T. J., Monson, R. K., Scott, R. L., and Bowling, D. R.: The potential of carbonyl sulfide as a proxy for gross primary production at flux tower sites,
- 20 *Journal of Geophysical Research-Biogeosciences*, 116, 2011.
- Boden, T. A., Krassovski, M., and Yang, B.: The AmeriFlux data activity and data system: an evolving collection of data management techniques, tools, products and services, *Geoscientific Instrumentation Methods and Data Systems*, 2, 165-176, 2013.
- Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E. J., Brunke, E. G., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D. A., Krummel, P. B., Langenfelds, R. L., Ramonet, M., Schmidt, M., Steele, L. P., Szopa, S., Yver, C., Viovy, N., and Ciais, P.: Source attribution of the changes in atmospheric methane for 2006–2008,
- 25 *Atmos. Chem. Phys.*, 11, 3689-3700, 2011.
- Brown, E. N., Friehe, C. A., and Lenschow, D. H.: THE USE OF PRESSURE-FLUCTUATIONS ON THE NOSE OF AN AIRCRAFT FOR MEASURING AIR MOTION, *Journal of Climate and Applied Meteorology*, 22, 171-180, 1983.
- 30 Cambaliza, M. O. L., Bogner, J. E., Green, R. B., Shepson, P. B., Harvey, T. A., Spokas, K. A., Stirm, B. H., and Corcoran, M.: Field measurements and modeling to resolve m(2) to km(2) CH₄ emissions for a complex urban source: An Indiana landfill study, *Elementa-Science of the Anthropocene*, 5, 2017.
- Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirm, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L., and Richardson, S. J.: Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions, *Atmos. Chem. Phys.*, 14, 9029-9050, 2014.
- 35 Chang, R. Y. W., Miller, C. E., Dinardo, S. J., Karion, A., Sweeney, C., Daube, B. C., Henderson, J. M., Mountain, M. E., Eluszkiewicz, J., Miller, J. B., Bruhwiler, L. M. P., and Wofsy, S. C.: Methane emissions from Alaska in 2012 from CARVE airborne observations, *P. Nat. Acad. Sci. USA*, 111, 16694-16699, 2014.
- 40 Chen, J. M., Leblanc, S. G., Cihlar, J., Desjardins, R. L., and MacPherson, J. I.: Extending aircraft- and tower-based CO₂ flux measurements to a boreal region using a Landsat thematic mapper land cover map, *J. Geophys. Res. Atmos.*, 104, 16859-16877, 1999.
- Chen, H., Winderlich, J., Gerbig, C., Hofer, A., Rella, C. W., Crosson, E. R., Van Pelt, A. D., Steinbach, J., Kolle, O., Beck, V., Daube, B. C., Bottlieb, E. W., Chow, V. Y., Gantoni, G. W., and Wofsy, S. C.: High-accuracy continuous airborne measurements of greenhouse gases (CO₂ and CH₄) using the cavity ring-down spectroscopy (CRDS) technique,
- 45 *Atmos. Meas. Tech.*, 3, 375-386, 2010.
- Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Frohlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morgui, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L.,

- Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *J. Geophys. Res. Atmos.*, 115, 2010.
- 5 Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., Jones, C., Quéré, C. L., Myneni, R. B., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.*, 2013.
- 10 Commane, R., Lindaas, J., Benmergui, J., Luus, K. A., Chang, R. Y., Daube, B. C., Euskirchen, E. S., Henderson, J. M., Karion, A., Miller, J. B., Miller, S. M., Parazoo, N. C., Randerson, J. T., Sweeney, C., Tans, P., Thoning, K., Veraverbeke, S., Miller, C. E., and Wofsy, S. C.: Carbon dioxide sources from Alaska driven by increasing early winter respiration from Arctic tundra, *Proc Natl Acad Sci U S A*, 114, 5361-5366, 2017.
- Conley, S. A., Faloon, I. C., Lenschow, D. H., Campos, T., Heizer, C., Weinheimer, A., Cantrell, C. A., Mauldin, R. L., Hornbrook, R. S., Pollack, I., and Bandy, A.: A complete dynamical ozone budget measured in the tropical marine boundary layer during PASE, *J. Atmos. Chem.*, 68, 55-70, 2011.
- 15 Dabberdt, W. F., Lenschow, D. H., Horst, T. W., Zimmerman, P. R., Oncley, S. P., and Delany, A. C.: *ATMOSPHERE-SURFACE EXCHANGE MEASUREMENTS*, *Science*, 260, 1472-1481, 1993.
- Desjardins, R. L., Brach, E. J., Alvo, P., and Schuepp, P. H.: Aircraft Monitoring of Surface Carbon Dioxide Exchange, *Science*, 216, 733-735, 1982.
- Desjardins, R. L., Worth, D. E., Pattey, E., VanderZaag, A., Srinivasan, R., Mauder, M., Worthy, D., Sweeney, C., and Metzger, S.: The challenge of reconciling bottom-up agricultural methane emissions inventories with top-down measurements, *Agr. Forest Met.*, 248, 48-59, 2018.
- 20 Diskin, G. S., Podolske, J. R., Sachse, G. W., and Slate, T. A.: Open-path airborne tunable diode laser hygrometer. In: *Diode Lasers and Applications in Atmospheric Sensing*, Fried, A. (Ed.), *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2002.
- 25 Dlugokencky, E. J., Nisbet, E. G., Fisher, R., and Lowry, D.: Global atmospheric methane: budget, changes and dangers, *Philos Trans A Math Phys Eng Sci*, 369, 2058-2072, 2011.
- Finkelstein, P. L. and Sims, P. F.: Sampling error in eddy correlation flux measurements, *J. Geophys. Res. Atmos.*, 106, 3503-3509, 2001.
- Foken, T. and Wichura, B.: Tools for quality assessment of surface-based flux measurements, *Agr. Forest Met.*, 78, 83-105, 1996.
- 30 Frick, P., Baliunas, S. L., Galyagin, D., Sokoloff, D., and Soon, W.: Wavelet Analysis of Stellar Chromospheric Activity Variations, *The Astrophysical Journal*, 483, 426-434, 1997.
- Frick, P., Grossmann, A., and Tchamitchian, P.: Wavelet analysis of signals with gaps, *Journal of Mathematical Physics*, 39, 4091-4107, 1998.
- 35 Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K. G., Schnur, R., Strassmann, K., Weaver, A. J., Yoshikawa, C., and Zeng, N.: Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison, *J. Climate*, 19, 3337-3353, 2006.
- 40 Gioli, B., Miglietta, F., De Martino, B., Hutjes, R. W. A., Dolman, H. A. J., Lindroth, A., Schumacher, M., Sanz, M. J., Manca, G., Peressotti, A., and Dumas, E. J.: Comparison between tower and aircraft-based eddy covariance fluxes in five European regions, *Agr. Forest Met.*, 127, 1-16, 2004.
- Gu, D., Guenther, A. B., Shilling, J. E., Yu, H., Huang, M., Zhao, C., Yang, Q., Martin, S. T., Artaxo, P., Kim, S., Seco, R., Stavrakou, T., Longo, K. M., Tota, J., de Souza, R. A. F., Vega, O., Liu, Y., Shrivastava, M., Alves, E. G., Santos, F. C., 45 Leng, G., and Hu, Z.: Airborne observations reveal elevational gradient in tropical forest isoprene emissions, *Nat Commun*, 8, 15541, 2017.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models, *Nature*, 415, 626-630, 2002.
- 50

- Hayes, D. J., Turner, D. P., Stinson, G., McGuire, A. D., Wei, Y. X., West, T. O., Heath, L. S., Dejong, B., McConkey, B. G., Birdsey, R. A., Kurz, W. A., Jacobson, A. R., Huntzinger, D. N., Pan, Y. D., Mac Post, W., and Cook, R. B.: Reconciling estimates of the contemporary North American carbon balance among terrestrial biosphere models, atmospheric inversions, and a new approach for estimating net ecosystem exchange from inventory-based data, *Global Change Biology*, 18, 1282-1299, 2012.
- Heimburger, A. M. F., Harvey, R. M., Shepson, P. B., Stirm, B. H., Gore, C., Turnbull, J., Cambaliza, M. O. L., Salmon, O. E., Kerlo, A. E. M., Lavoie, T. N., Davis, K. J., Lauvaux, T., Karion, A., Sweeney, C., Brewer, W. A., Hardesty, R. M., and Gurney, K. R.: Assessing the optimized precision of the aircraft mass balance method for measurement of urban greenhouse gas emission rates through averaging, *Elementa-Science of the Anthropocene*, 5, 2017.
- Hiller, R. V., Neininger, B., Brunner, D., Gerbig, C., Bretscher, D., Künzle, T., Buchmann, N., and Eugster, W.: Aircraft-based CH₄ flux estimates for validation of emissions from an agriculturally dominated area in Switzerland, *J. Geophys. Res. Atmos.*, 119, 4874-4887, 2014.
- Horst, T. W.: A simple formula for attenuation of eddy fluxes measured with first-order-response scalar sensors, *Boundary-Layer Met.*, 82, 219-233, 1997.
- Houweling, S., Baker, D., Basu, S., Boesch, H., Butz, A., Chevallier, F., Deng, F., Dlugokencky, E. J., Feng, L., Ganshin, A., Hasekamp, O., Jones, D., Maksyutov, S., Marshall, J., Oda, T., O'Dell, C. W., Oshchepkov, S., Palmer, P. I., Peylin, P., Poussi, Z., Reum, F., Takagi, H., Yoshida, Y., and Zhuravlev, R.: An intercomparison of inverse models for estimating sources and sinks of CO₂ using GOSAT measurements, *J. Geophys. Res. Atmos.*, 120, 5253-5266, 2015.
- Huntzinger, D. N., Post, W. M., Wei, Y., Michalak, A. M., West, T. O., Jacobson, A. R., Baker, I. T., Chen, J. M., Davis, K. J., Hayes, D. J., Hoffman, F. M., Jain, A. K., Liu, S., McGuire, A. D., Neilson, R. P., Potter, C., Poulter, B., Price, D., Raczka, B. M., Tian, H. Q., Thornton, P., Tomelleri, E., Viovy, N., Xiao, J., Yuan, W., Zeng, N., Zhao, M., and Cook, R.: North American Carbon Program (NACP) regional interim synthesis: Terrestrial biospheric model intercomparison, *Ecological Modelling*, 232, 144-157, 2012.
- Jensen, E. J., Pfister, L., Jordan, D. E., Bui, T. V., Ueyama, R., Singh, H. B., Thornberry, T. D., Rollins, A. W., Gao, R.-S., Fahey, D. W., Rosenlof, K. H., Elkins, J. W., Diskin, G. S., DiGangi, J. P., Lawson, R. P., Woods, S., Atlas, E. L., Rodriguez, M. A. N., Wofsy, S. C., Pittman, J., Bardeen, C. G., Toon, O. B., Kindel, B. C., Newman, P. A., McGill, M. J., Hlavka, D. L., Lait, L. R., Schoeberl, M. R., Bergman, J. W., Selkirk, H. B., Alexander, M. J., Kim, J.-E., Lim, B. H., Stutz, J., and Pfeilsticker, K.: The NASA Airborne Tropical Tropopause Experiment: High-Altitude Aircraft Measurements in the Tropical Western Pacific, *Bull. Am. Met. Soc.*, 98, 129-143, 2017.
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J. Q., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations (vol 116, G00J07, 2011), *Journal of Geophysical Research-Biogeosciences*, 117, 2012.
- Kaimal, J. C., Izumi, y., Wyngaard, J. C., and Cote, R.: Spectral characteristics of surface-layer turbulence, *Quarterly Journal of the Royal Meteorological Society*, 98, 563-&, 1972.
- Kaimal, J. C., Wyngaard, J. C., HAUGEN, D. A., Cote, O. R., Izumi, Y., Caughey, S. J., and Readings, C. J.: TURBULENCE STRUCTURE IN CONVECTIVE BOUNDARY-LAYER, *J. Atmos. Sci.*, 33, 2152-2169, 1976.
- Karion, A., Sweeney, C., Petron, G., Frost, G., Hardesty, R. M., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R., and Conley, S.: Methane emissions estimate from airborne measurements over a western United States natural gas field, *Geophys. Res. Lett.*, 40, 4393-4397, 2013a.
- Karion, A., Sweeney, C., Wolter, S., Newberger, T., Chen, H., Andrews, A., Kofler, J., Neff, D., and Tans, P.: Long-term greenhouse gas measurements from aircraft, *Atmos. Meas. Tech.*, 6, 511-526, 2013b.
- Karl, T., Apel, E., Hodzic, A., Riemer, D., Blake, D., and Wiedinmyer, C.: Emissions of volatile organic compounds inferred from airborne flux measurements over a megacity, *Atmos. Chem. Phys.*, 9, 271-285, 2009.
- Karl, T., Misztal, P., Jonsson, H., Shertz, S., Goldstein, A., and Guenther, A.: Airborne Flux Measurements of BVOCs above Californian Oak Forests: Experimental Investigation of Surface and Entrainment Fluxes, OH Densities, and Damkohler Numbers, *J. Atmos. Sci.*, 70, 3277-3287, 2013.

- Kaser, L., Karl, T., Yuan, B., Mauldin III, R. L., Cantrell, C. A., Guenther, A. B., Patton, E. G., Weinheimer, A. J., Knote, C., Orlando, J., Emmons, L., Apel, E., Hornbrook, R., Shertz, S., Ullmann, K., Hall, S., Graus, M., de Gouw, J., Zhou, X., and Ye, C.: Chemistry-turbulence interactions and mesoscale variability influence the cleansing efficiency of the atmosphere, *Geophys. Res. Lett.*, 42, 10894-10903, 2015.
- 5 Kawa, S. R. and Pearson, R.: Ozone Budgets from the Dynamics and Chemistry of Marine Stratocumulus Experiment, *J. Geophys. Res. Atmos.*, 94, 9809-9817, 1989.
- Kljun, N., Calanca, P., Rotach, M. W., and Schmid, H. P.: A simple parameterization for flux footprint predictions, *Boundary-Layer Met.*, 112, 503-523, 2004.
- Kljun, N., Calanca, P., Rotach, M. W., and Schmid, H. P.: A simple two-dimensional parameterisation for Flux Footprint Prediction (FFP), *Geosci. Mod. Dev.*, 8, 3695-3713, 2015.
- 10 Kustas, W. P., Anderson, M. C., French, A. N., and Vickers, D.: Using a remote sensing field experiment to investigate flux-footprint relations and flux sampling distributions for tower and aircraft-based observations, *Advances in Water Resources*, 29, 355-368, 2006.
- 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, 2015.
- 15 Le Quere, C., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., Manning, A. C., Boden, T. A., Tans, P. P., Houghton, R. A., Keeling, R. F., Alin, S., Andrews, O. D., Anthoni, P., Barbero, L., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Currie, K., Delire, C., Doney, S. C., Friedlingstein, P., Gkritzalis, T., Harris, I., Hauck, J., Haverd, V., Hoppema, M., Goldewijk, K. K., Jain, A. K., Kato, E., Kortzinger, A., Landschutzer, P., Lefevre, N., Lenton, A., Lienert, S., Lombardozi, D., Melton, J. R., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J., Nakaoka, S., O'Brien, K., Olsen, A., Omar, A. M., Ono, T., Pierrot, D., Poulter, B., Rodenbeck, C., Salisbury, J., Schuster, U., Schwinger, J., Seferian, R., Skjelvan, I., Stocker, B. D., Sutton, A. J., Takahashi, T., Tian, H. Q., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2016, *Earth System Science Data*, 8, 605-649, 2016.
- 20 Lenschow, D. H., Delany, A. C., Stankov, B. B., and Stedman, D. H.: Airborne Measurements of the Vertical Flux of Ozone in the Boundary-Layer, *Boundary-Layer Met.*, 19, 249-265, 1980.
- Lenschow, D. H., Mann, J., and Kristensen, L.: How long is long enough when measuring fluxes and other turbulence statistics, *J. Atmos. Ocean. Technol.*, 11, 661-673, 1994.
- Lenschow, D. H., Pearson, R., and Stankov, B. B.: Estimating the ozone budget in the boundary-layer by use of aircraft measurements of ozone eddy flux and mean concentration, *J. Geophys. Res. Oceans Atmos.*, 86, 7291-7297, 1981.
- 30 Lenschow, D. H., Wulfmeyer, V., and Senff, C.: Measuring Second- through Fourth-Order Moments in Noisy Data, *J. Atmos. Ocean. Technol.*, 17, 1330-1347, 2000.
- Liu, Y. J., San Liang, X., and Weisberg, R. H.: Rectification of the Bias in the Wavelet Power Spectrum, *J. Atmos. Ocean. Technol.*, 24, 2093 - 2102, 2007.
- 35 Mauder, M., Cuntz, M., Drue, C., Graf, A., Rebmann, C., Schmid, H., Schmidt, M., and Steinbrecher, R.: A strategy for quality and uncertainty assessment of long-term eddy-covariance measurements, *Agr. Forest Met.*, 169, 122-135, 2013.
- Mauder, M., Desjardins, R., and MacPherson, I.: Scale analysis of airborne flux measurements over heterogeneous terrain in a boreal ecosystem, *J. Geophys. Res. Atmos.*, 112, 2007.
- 40 Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Eddy-covariance flux measurements with a weight-shift microlight aircraft, *Atmos. Meas. Tech.*, 5, 1699-1717, 2012.
- Metzger, S., Junkermann, W., Mauder, M., Butterbach-Bahl, K., Widemann, B. T. Y., Neidl, F., Schafer, K., Wieneke, S., Zheng, X. H., Schmid, H. P., and Foken, T.: Spatially explicit regionalization of airborne flux measurements using environmental response functions, *Biogeosci.*, 10, 2193-2217, 2013.
- 45 Miglietta, F., Gioli, B., Hutjes, R. W. A., and Reichstein, M.: Net regional ecosystem CO₂ exchange from airborne and ground-based eddy covariance, land-use maps and weather observations, *Global Change Biology*, 13, 548-560, 2007.
- Misztal, P. K., Avise, J. C., Karl, T., Scott, K., Jonsson, H. H., Guenther, A. B., and Goldstein, A. H.: Evaluation of regional isoprene emission factors and modeled fluxes in California, *Atmos. Chem. Phys.*, 16, 9611-9628, 2016.
- Misztal, P. K., Karl, T., Weber, R., Jonsson, H. H., Guenther, A. B., and Goldstein, A. H.: Airborne flux measurements of biogenic volatile organic compounds over California, *Atmos. Chem. Phys.*, 14, 10631-10647, 2014.

- Mondal, D. and Percival, D. B.: Wavelet variance analysis for gappy time series, *Annals of the Institute of Statistical Mathematics*, 62, 943-966, 2008.
- Rannik, Ü., Peltola, O., and Mammarella, I.: Random uncertainties of flux measurements by the eddy covariance technique, *Atmos. Meas. Tech.*, 9, 5163-5181, 2016.
- 5 Rella, C.: Accurate Greenhouse Gas Measurements in Humid Gas Streams Using the Picarro G1301 Carbon Dioxide / Methane / Water Vapor Gas Analyzer, PICARRO, INC., 2010.
- Ritter, J. A., Barrick, J. D. W., Sachse, G. W., Gregory, G. L., Woerner, M. A., Watson, C. E., Hill, G. F., and Collins, J. E.: Airborne flux measurements of trace species in an Arctic boundary layer, *J. Geophys. Res. Atmos.*, 97, 16601-16625, 1992.
- 10 Ritter, J. A., Barrick, J. D. W., Watson, C. E., Sachse, G. W., Gregory, G. L., Anderson, B. E., Woerner, M. A., and Collins, J. E.: Airborne boundary layer flux measurements of trace species over Canadian boreal forest and northern wetland regions, *J. Geophys. Res. Atmos.*, 99, 1671-1685, 1994.
- Ritter, J. A., Lenschow, D. H., Barrick, J. D. W., Gregory, G. L., Sachse, G. W., Hill, G. F., and Woerner, M. A.: AIRBORNE FLUX MEASUREMENTS AND BUDGET ESTIMATES OF TRACE SPECIES OVER THE AMAZON
- 15 BASIN DURING THE GTE ABLE-2B EXPEDITION, *J. Geophys. Res. Atmos.*, 95, 16875-16886, 1990.
- Rollins, A. W., Thornberry, T. D., Gao, R. S., Smith, J. B., Sayres, D. S., Sargent, M. R., Schiller, C., Krämer, M., Spelten, N., Hurst, D. F., Jordan, A. F., Hall, E. G., Vömel, H., Diskin, G. S., Podolske, J. R., Christensen, L. E., Rosenlof, K. H., Jensen, E. J., and Fahey, D. W.: Evaluation of UT/LS hygrometer accuracy by intercomparison during the NASA MACPEX mission, *J. Geophys. Res. Atmos.*, 119, 1915-1935, 2014.
- 20 Rothman, L. S., Gordon, I. E., Babikov, Y., Barbe, A., Benner, D. C., Bernath, P. F., Birk, M., Bizzocchi, L., Boudon, V., Brown, L. R., Campargue, A., Chance, K., Cohen, E. A., Coudert, L. H., Devi, V. M., Drouin, B. J., Fayt, A., Flaud, J. M., Gamache, R. R., Harrison, J. J., Hartmann, J. M., Hill, C., Hodges, J. T., Jacquemart, D., Jolly, A., Lamouroux, J., Le Roy, R. J., Li, G., Long, D. A., Lyulin, O. M., Mackie, C. J., Massie, S. T., Mikhailenko, S., Muller, H. S. P., Naumenko, O. V., Nikitin, A. V., Orphal, J., Perevalov, V., Perrin, A., Polovtseva, E. R., Richard, C., Smith, M. A. H., Starikova, E.,
- 25 Sung, K., Tashkun, S., Tennyson, J., Toon, G. C., Tyuterev, V. G., and Wagner, G.: The HITRAN2012 molecular spectroscopic database, *Journal of Quantitative Spectroscopy & Radiative Transfer*, 130, 4-50, 2013.
- Sayres, D. S., Dobosy, R., Healy, C., Dumas, E., Kochendorfer, J., Munster, J., Wilkerson, J., Baker, B., and Anderson, J. G.: Arctic regional methane fluxes by ecotope as derived using eddy covariance from a low-flying aircraft, *Atmos. Chem. Phys.*, 17, 8619-8633, 2017.
- 30 Schaefer, K., Collatz, G. J., Tans, P., Denning, A. S., Baker, I., Berry, J., Prihodko, L., Suits, N., and Philpott, A.: Combined Simple Biosphere/Carnegie-Ames-Stanford Approach terrestrial carbon cycle model, *Journal of Geophysical Research-Biogeosciences*, 113, 2008.
- Schaller, C., Gockede, M., and Foken, T.: Flux calculation of short turbulent events - comparison of three methods, *Atmos. Meas. Tech.*, 10, 869-880, 2017.
- 35 Schimel, D., Stephens, B. B., and Fisher, J. B.: Effect of increasing CO₂ on the terrestrial carbon cycle, *P. Nat. Acad. Sci. USA*, 112, 436-441, 2015.
- Sellers, P. J., Hall, F. G., Kelly, R. D., Black, A., Baldocchi, D., Berry, J., Ryan, M., Ranson, K. J., Crill, P. M., Lettenmaier, D. P., Margolis, H., Cihlar, J., Newcomer, J., Fitzjarrald, D., Jarvis, P. G., Gower, S. T., Halliwell, D., Williams, D., Goodison, B., Wickland, D. E., and Guertin, F. E.: BOREAS in 1997: Experiment overview, scientific results, and future
- 40 directions, *J. Geophys. Res. Atmos.*, 102, 28731-28769, 1997.
- Silver, J. A.: FREQUENCY-MODULATION SPECTROSCOPY FOR TRACE SPECIES DETECTION - THEORY AND COMPARISON AMONG EXPERIMENTAL METHODS, *Appl. Opt.*, 31, 707-717, 1992.
- Stickney, T. M., Shedlov, M. W., and Thompson, D. I.: Rosemount Total Temperature Sensors, Tech. Rep. 5755, Rev. B., Rosemount, Inc., 1990.
- 45 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, 2003.
- Torrence, C. and Compo, G.: A practical guide to wavelet analysis, *Bull. Am. Met. Soc.*, 79, 61-78, 1998.

- Trousdell, J. F., Conley, S. A., Post, A., and Faloon, I. C.: Observing entrainment mixing, photochemical ozone production, and regional methane emissions by aircraft using a simple mixed-layer framework, *Atmos. Chem. Phys.*, 16, 15433-15450, 2016.
- 5 Vaughan, A. R., Lee, J. D., Misztal, P. K., Metzger, S., Shaw, M. D., Lewis, A. C., Purvis, R. M., Carslaw, D. C., Goldstein, A. H., Hewitt, C. N., Davison, B., Beevers, S. D., and Karl, T. G.: Spatially resolved flux measurements of NO_x from London suggest significantly higher emissions than predicted by inventories, *Faraday Disc.*, 189, 455-472, 2016.
- Webb, E. K., Pearman, G. I., and Leuning, R.: Correction of flux measurements for density effects due to heat and water vapour transfer, *Quarterly Journal of the Royal Meteorological Society*, 106, 85-100, 1980.
- 10 Weil, J. C., Horst, T. W., Schwartz, S. E., and Slinn, W. G. N.: Footprint estimates for atmospheric flux measurements in the convective boundary-layer, *Precipitation Scavenging and Atmosphere-Surface Exchange*, Vols 1-3, 1992. 717-728, 1992.
- Wolfe, G. M., Hanisco, T. F., Arkinson, H. L., Bui, T. P., Crouse, 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, 2015.
- 15 Worden, J. R., Bloom, A. A., Pandey, S., Jiang, Z., Worden, H. M., Walker, T. W., Houweling, S., and Röckmann, T.: Reduced biomass burning emissions reconcile conflicting estimates of the post-2006 atmospheric methane budget, *Nature Communications*, 8, 2227, 2017.
- 20 Yu, H., Guenther, A., Gu, D., Warneke, C., Geron, C., Goldstein, A., Graus, M., Karl, T., Kaser, L., Misztal, P., and Yuan, B.: Airborne measurements of isoprene and monoterpene emissions from southeastern U.S. forests, *Sci Total Environ*, 595, 149-158, 2017.
- Yuan, B., Kaser, L., Karl, T., Graus, M., Peischl, J., Campos, T. L., Shertz, S., Apel, E. C., Hornbrook, R. S., Hills, A., Gilman, J. B., Lerner, B. M., Warneke, C., Flocke, F. M., Ryerson, T. B., Guenther, A. B., and de Gouw, J. A.: Airborne flux measurements of methane and volatile organic compounds over the Haynesville and Marcellus shale gas production regions, *J. Geophys. Res. Atmos.*, 120, 6271-6289, 2015.
- 25 Zulueta, R. C., Oechel, W. C., Verfaillie, J. G., Hastings, S. J., Gioli, B., Lawrence, W. T., and U, K. T. P.: Aircraft Regional-Scale Flux Measurements over Complex Landscapes of Mangroves, Desert, and Marine Ecosystems of Magdalena Bay, Mexico, *J. Atmos. Ocean. Technol.*, 30, 1266-1294, 2013.

Table 1. Instrument specifications for the CARAFE payload.

Measurement	Instrument	Data Rate (Hz)	1s Precision (@ 1Hz)	Accuracy
CO ₂	LGR	10	330 ppb	0.015% <u>600 ppb^a</u>
	Picarro	0.2	50 ppb	<u>200 ppb^b</u>
CH ₄	LGR	10	2.0 ppb	0.03% <u>4 ppb^a</u>
	Picarro	0.2	0.4 ppb	<u>1 ppb^b</u>
H ₂ O	LGR	10	200 ppm	7% ^{ac}
	Picarro	0.07	100 ppm	7% ^{ac}
	DLH	20	10 ppm	5%
3-D winds	TAMMS	20	0.05 m s ⁻¹	5%
Pressure			0.003 mb	5%
Vertical wind	Rosemount 858	20	0.05 m s ⁻¹	5%
Temperature	Rosemount TAT	20	0.05 K	5%
Aircraft position	Applanix 510	20	-	100 m
Aircraft attitude			-	0.005°
Telemetry	NASDAT	20	-	-
Visible imagery	Nikon 7000			
IR imagery	FLIR 325sc	1	-	-
4-band veg. health	MS RedEdge			
PPFD	LI-190R	1	-	10%

^aBased on Picarro accuracy and variance of LGR-Picarro difference in 1 Hz observations.

^bBased on laboratory calibrations and in-flight performance of similar instruments (Chen. et al., 2010; Karion et al., 2013b).

^cBased on in-flight comparison with DLH.

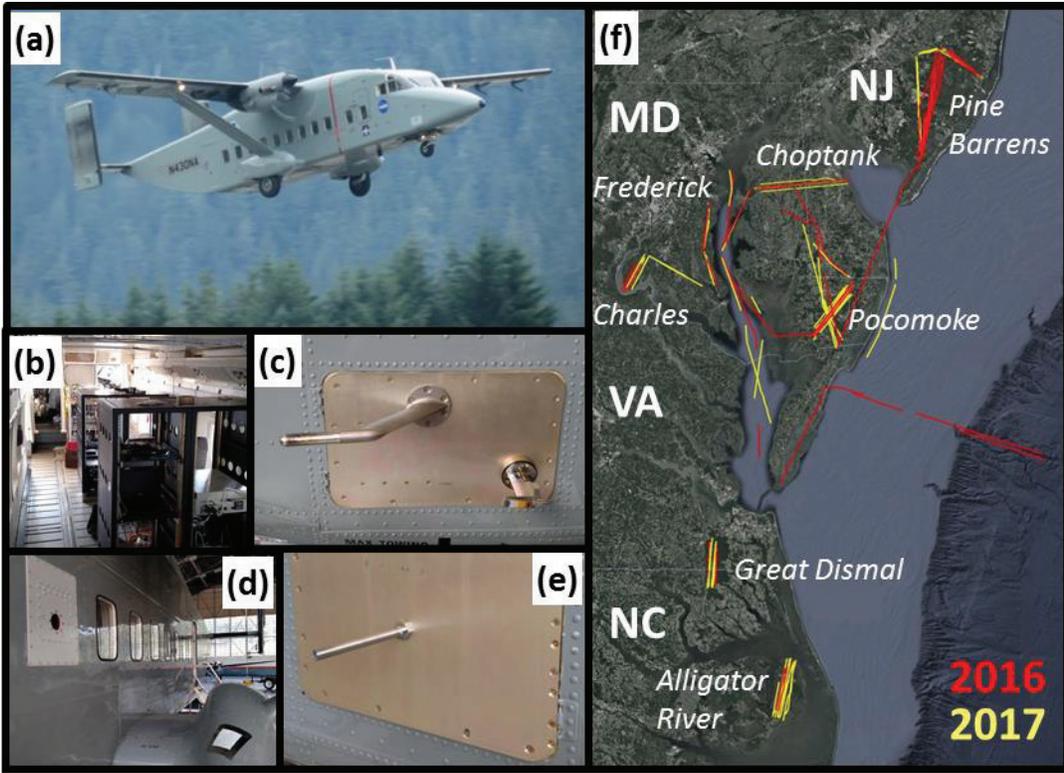
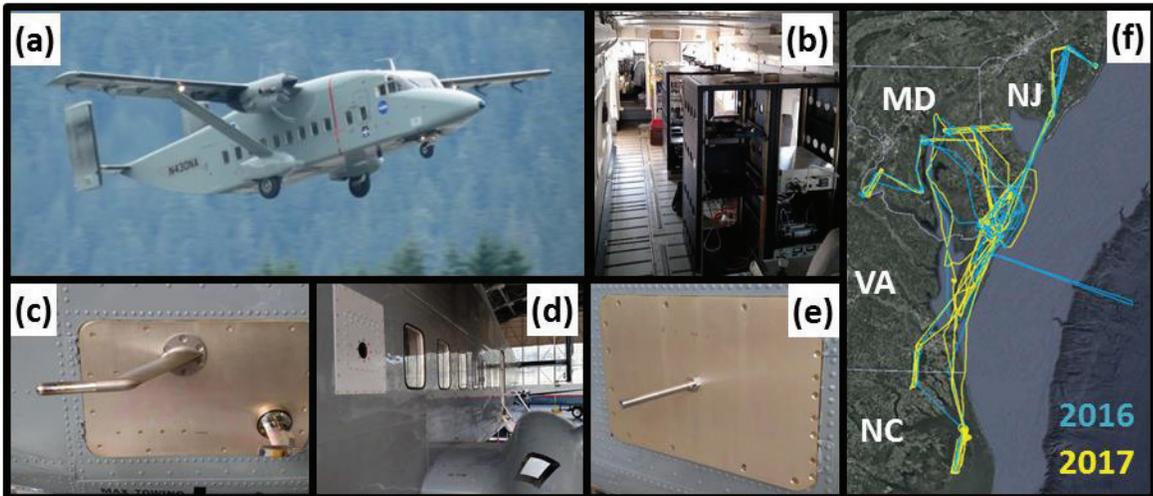


Figure 1. The CARAFE payload. (a) The NASA C-23B Sherpa. (b) A view inside the cabin with all instruments installed. (c) The angle-of-attack (upper) and total air temperature (lower) probes. (d) The DLH window plate (upper left) and fairing-mounted target (lower right). (e) GHG inlet. (f) **Flux leg** flight tracks for both campaigns. Target locations indicated in italicized text correspond to those listed in SI Table S1.

5

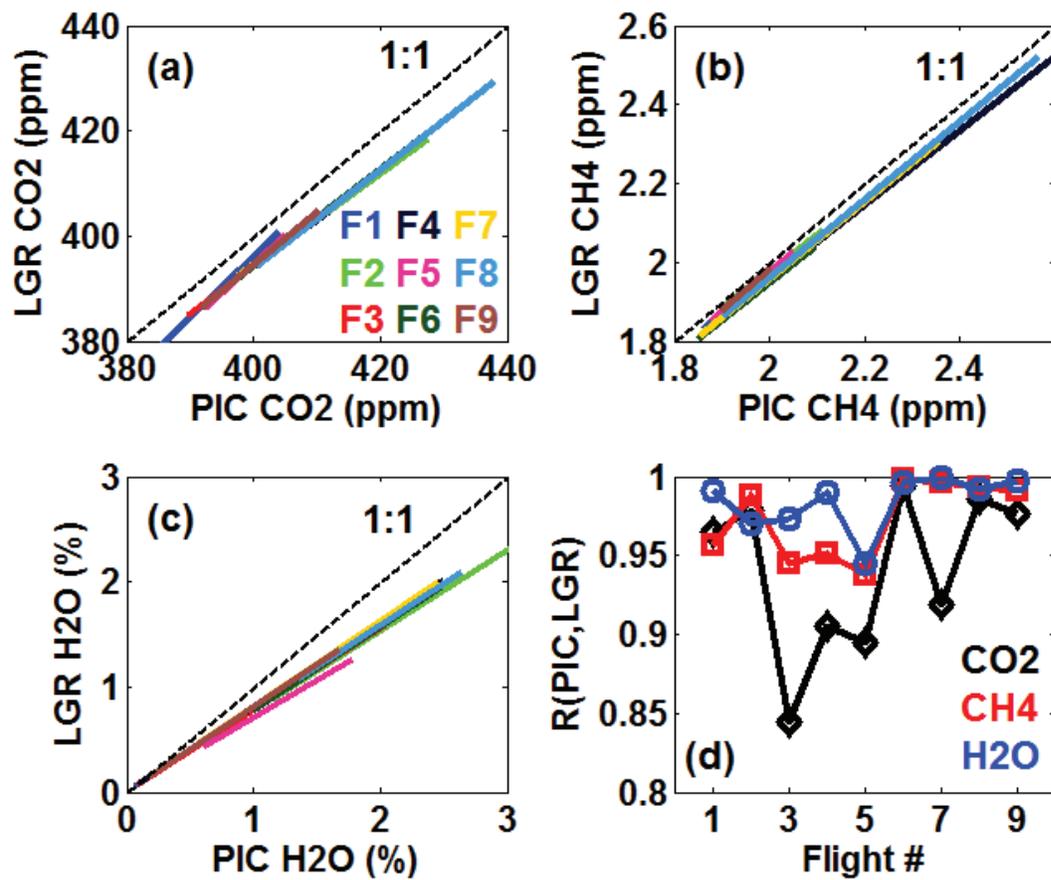
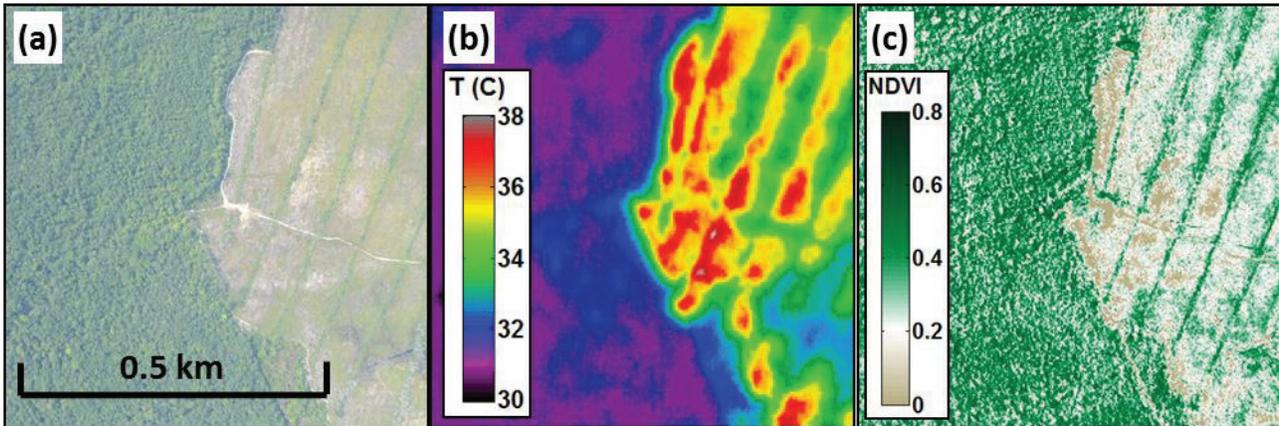
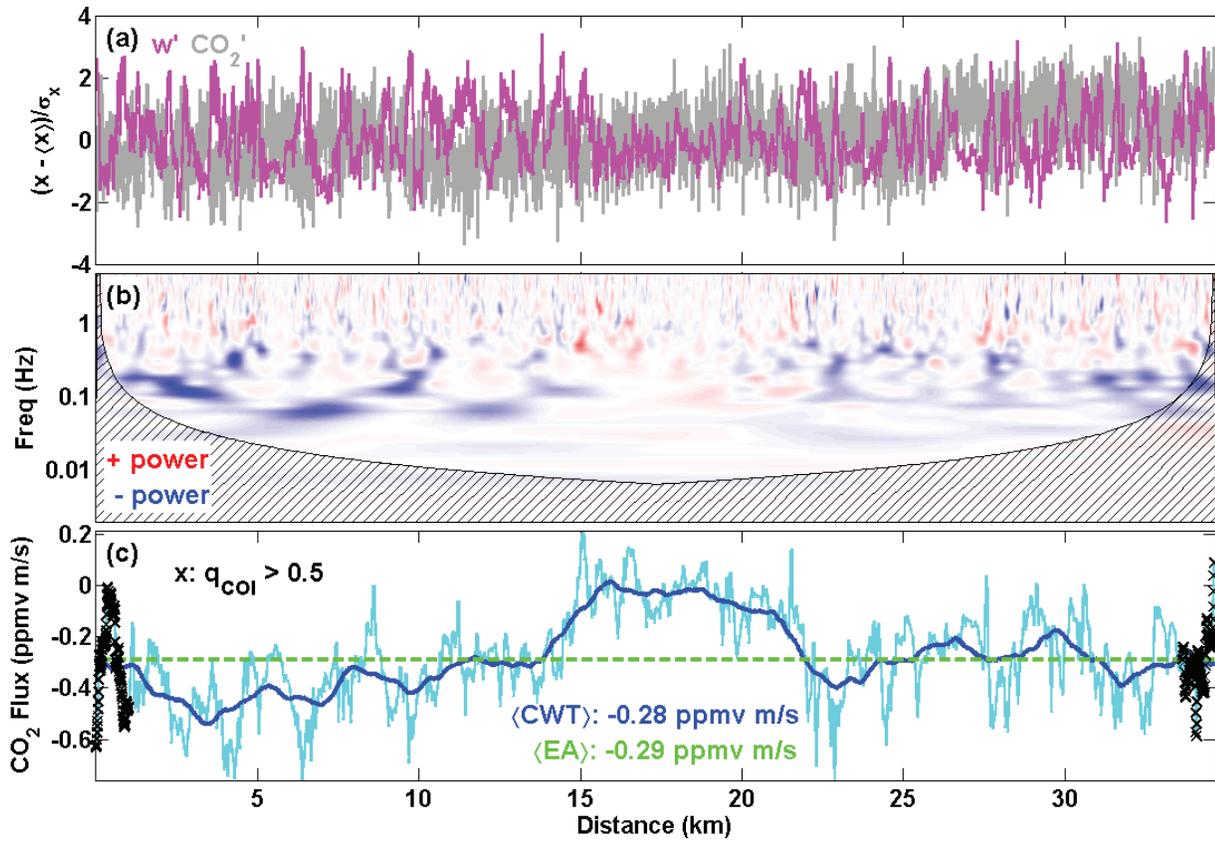


Figure 2. Least squares fit lines (a-c) and correlation coefficients (d) for Picarro and LGR dry mixing ratios of CO₂, CH₄ and H₂O obtained during flights in 2016. Fits are colored by flight and shown only over the range of mixing ratios observed on each flight.

5 Dashed lines in a), b) and c) denote a 1:1 correlation.



5 Figure 3. Example imagery from the SIS recorded during the 09 September 2016 flight over Pocomoke forest. (a) Visible image from the Nikon 7000. (b) surface temperature from the FLIR A325sc. (c) Normalized difference vegetation index (NDVI) derived from the “red” (668 ± 10 nm) and “near IR” (840 ± 40 nm) bands of the MicaSense RedEdge sensor. Note, the color scale for the latter saturates at both ends.



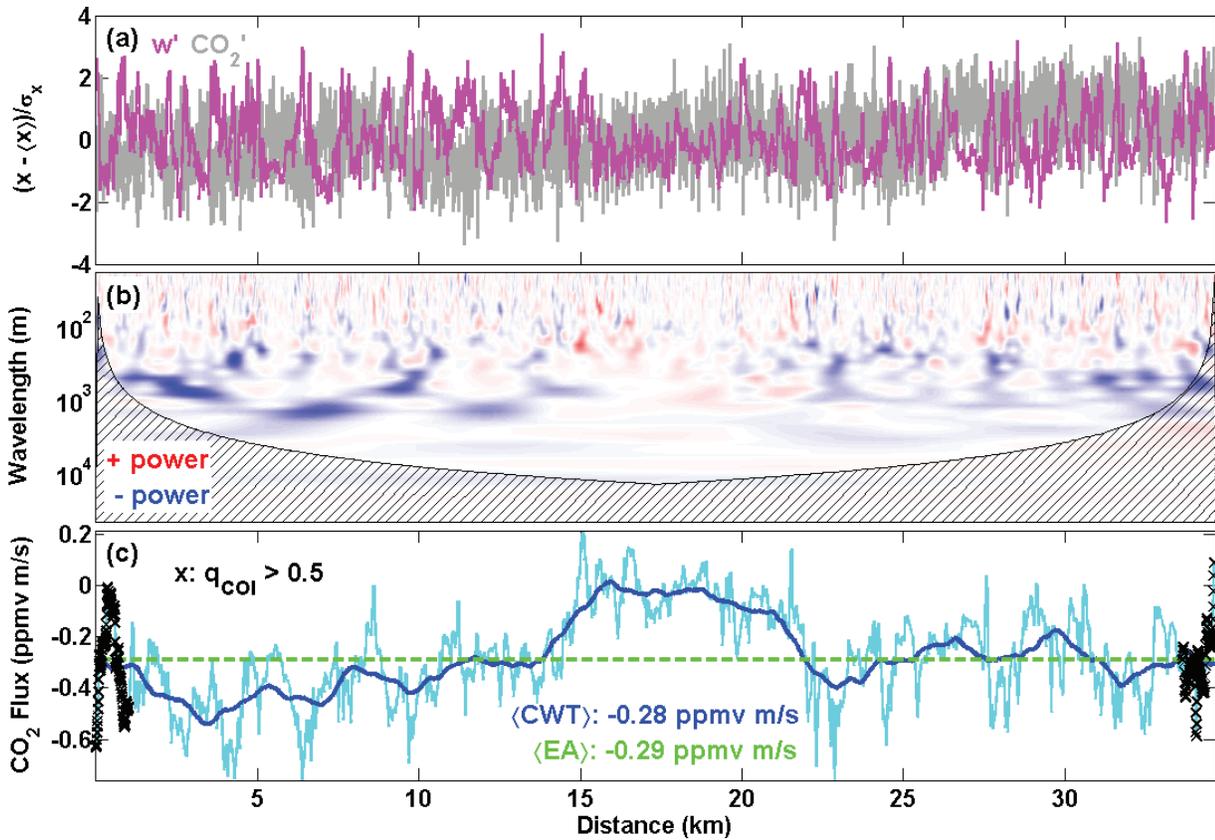
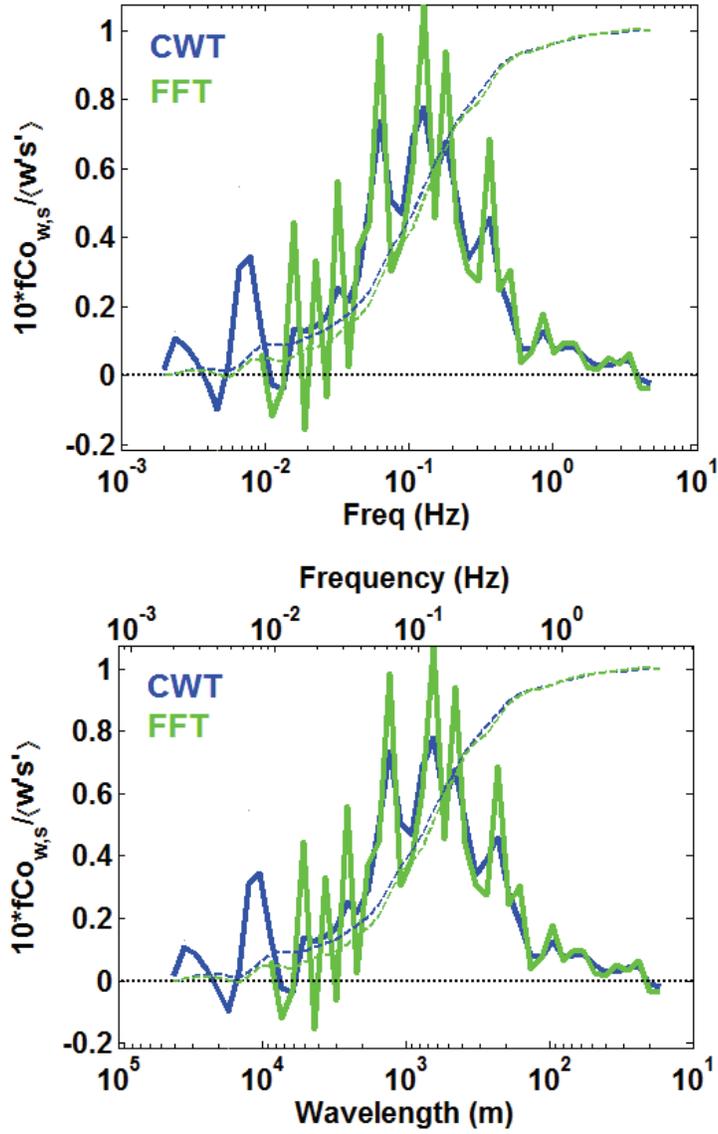
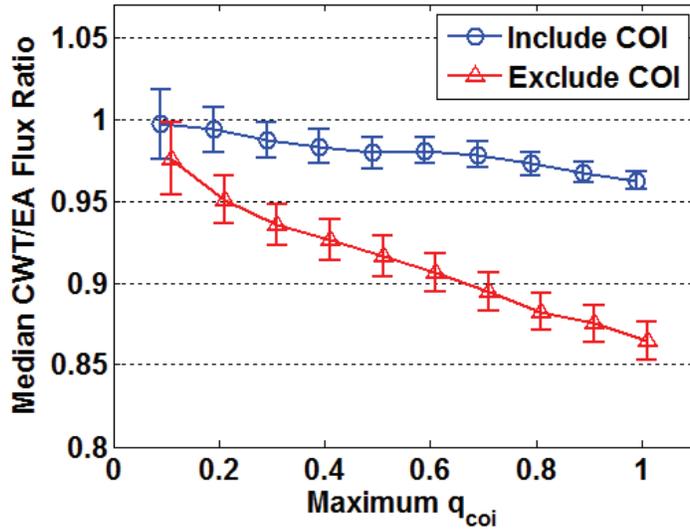


Figure 4. Example wavelet CO_2 flux calculation from a flight leg at 130 m altitude over Great Dismal Swamp, VA on 16 May 2017. (a) Normalized 10 Hz time series of vertical wind speed (w , magenta) and CO_2 (gray) fluctuations. (b) Local wavelet cospectrum. Red areas denote positive power, blue areas negative. Power is bias-corrected (multiplied by scale) as suggested by Liu et al. (2007). Hatched area indicates the cone of influence (COI). (c) Scale-integrated wavelet flux (cyan: 10 Hz, blue: 2 km running mean) and ensemble-average flux (green dash). Crosses denote points where 50% or more of the cospectral power lies within the COI.

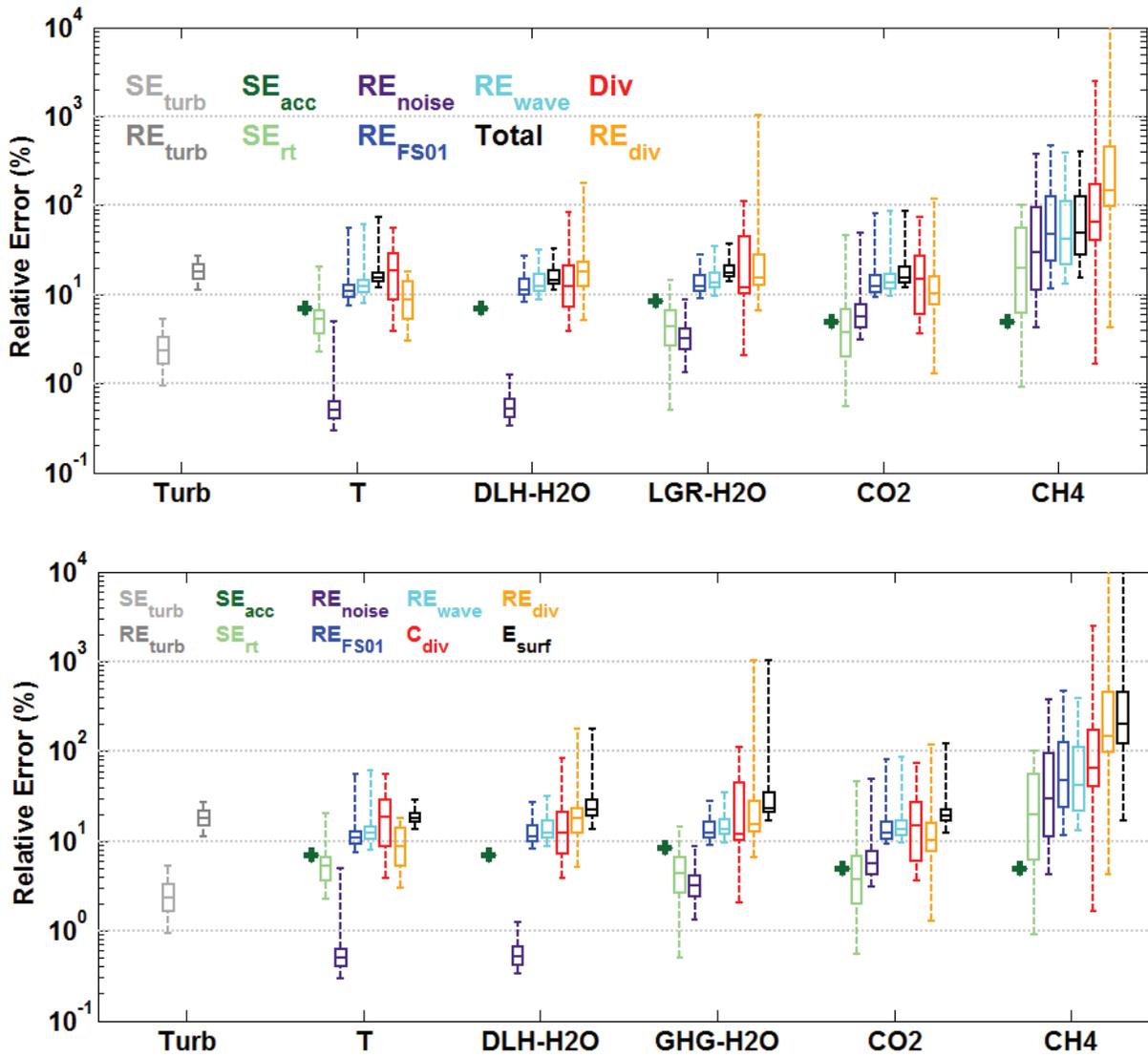
5



5 Figure 5. Global cospectra of vertical wind and potential temperature for the leg described in Fig. 4, calculated by time-averaging the local wavelet cospectrum (blue) and fast Fourier transforms (green). Cospectra are scaled for display. Dashed lines show the cumulative integrals of the cospectra (ogives), normalized by total covariance.

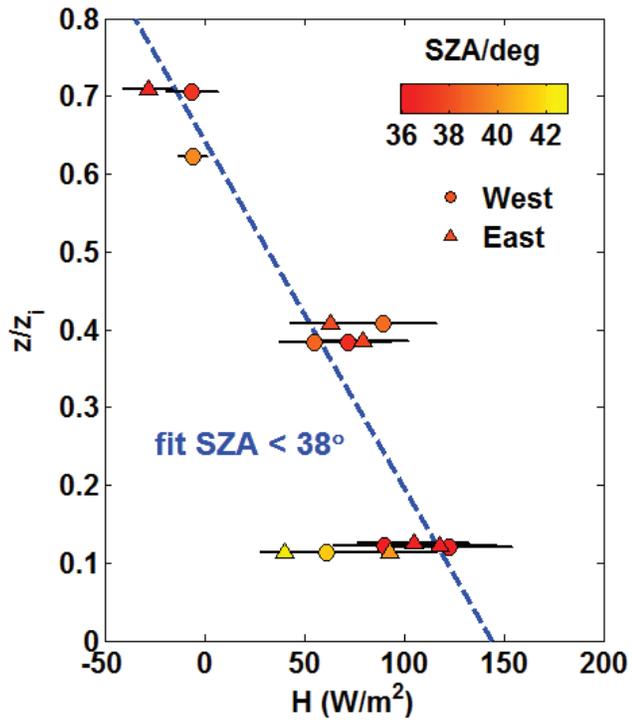


5 Figure 6. Impact of treatment of the COI on wavelet fluxes. Wavelet fluxes are calculated either including (blue circles) or excluding (red triangles) cospectral power within the COI. In addition, wavelet time series are filtered using the q_{coi} quality flag (see text) at various thresholds prior to averaging over each leg. Symbols represent the median ratio of wavelet to ensemble-average fluxes for all scalars (CO_2 , CH_4 , H, LE-LGR and LE-DLH) and legs with stationarity flags of 0.1 or less. Error bars represent standard errors on the ratio.



5 Figure 7. Distribution of errors, normalized by leg-average fluxes, for all 2017 flight legs below 200 m altitude (97 legs total). In the box plots, the center line is the median value, box edges are 25th/75th percentiles and whiskers are 5th/95th percentiles. Turbulence sampling errors (systematic: light gray, random: dark gray) are the same for all scalars. Instrument-specific systematic error sources include measurement accuracy (dark green) and limited response time (light green). Random errors due to uncorrelated noise are shown in indigo. Empirical random errors, which inherently include both the turbulence and instrument noise components, are derived for both the leg ensemble (blue) and the wavelet time series (cyan). **The total error (black) is the sum of SE_{turb} , SE_{FS} , SE_{acc} and RE_{wave} .** Also shown is the distribution of divergence corrections (red) and the additional fractional uncertainty in flux associated with this correction (orange). **The total uncertainty in derived surface fluxes (black) includes the combined contributions from systematic, random, and divergence errors.** Note that SE_{RT} for DLH-H₂O is undefined as this instrument is the time response standard.

10



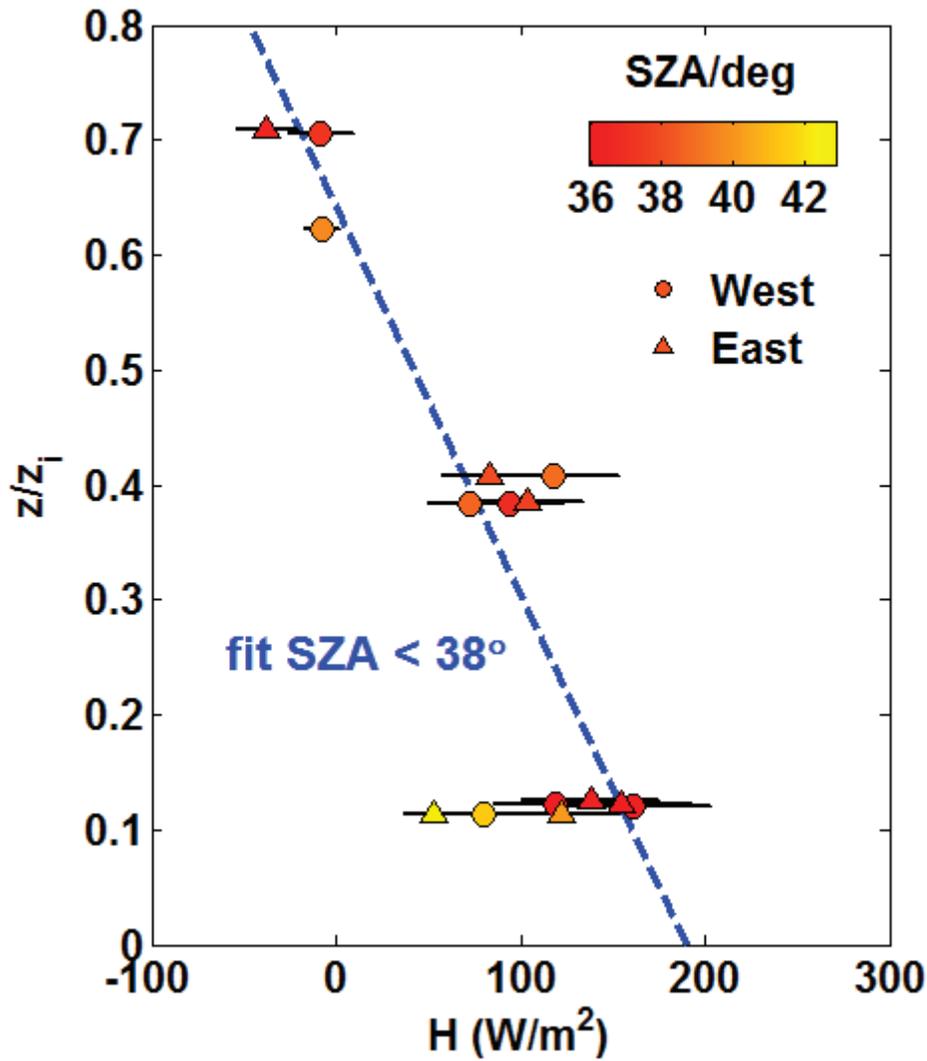
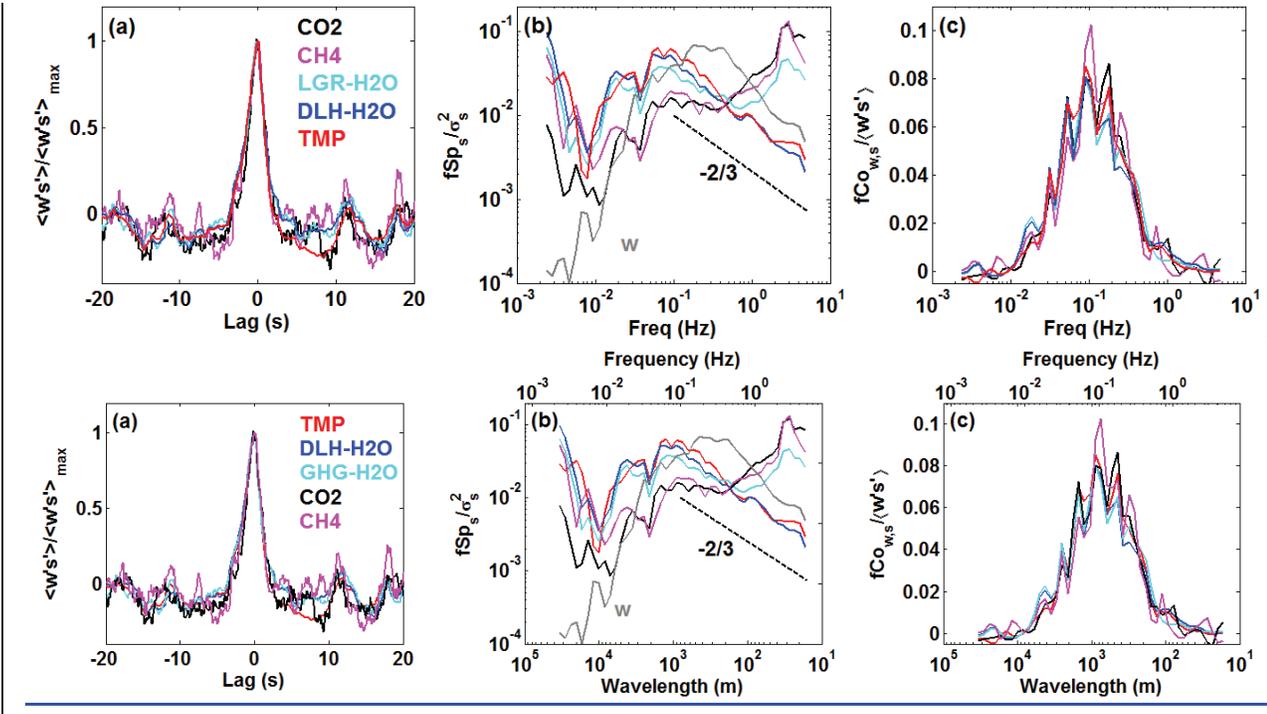


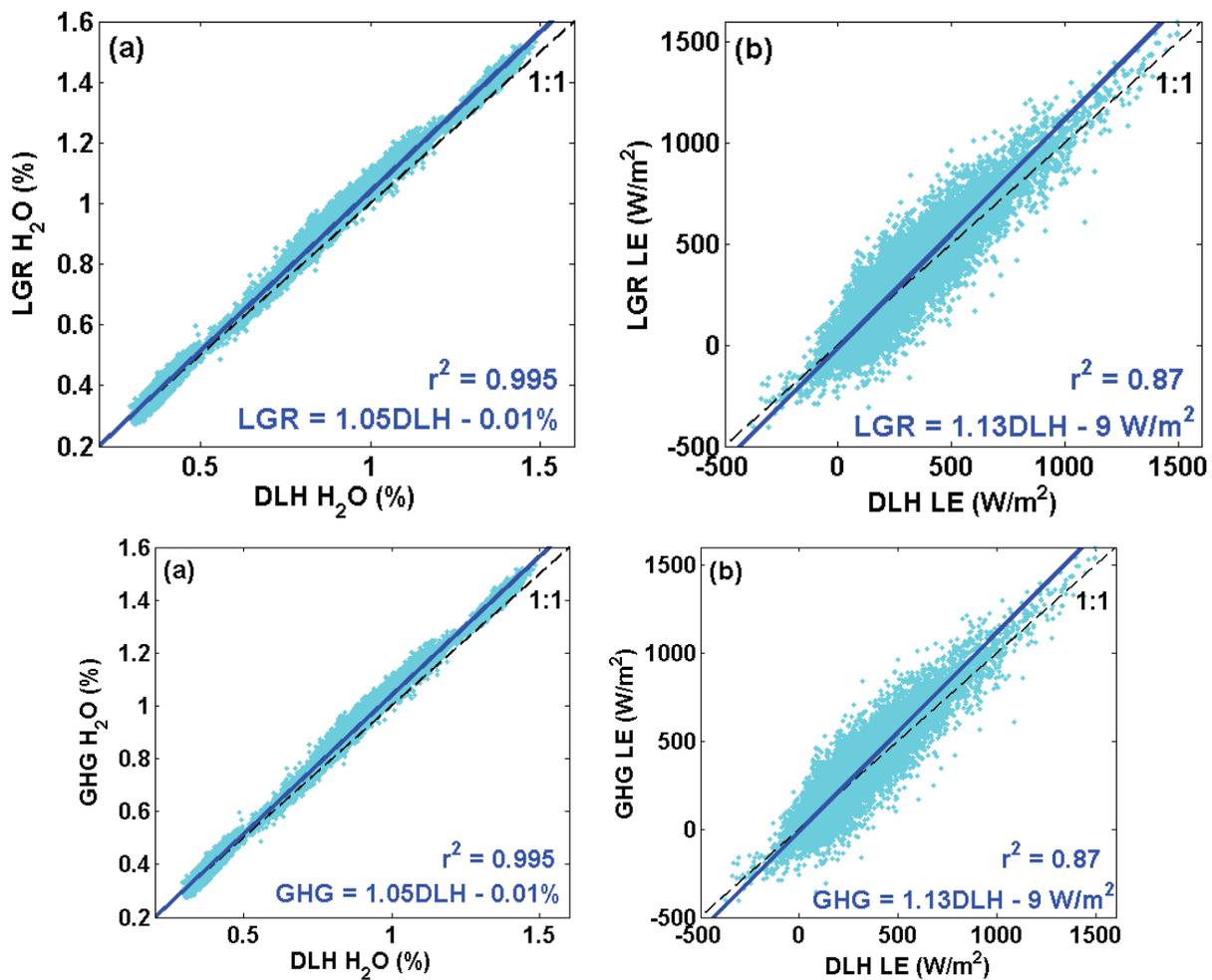
Figure 8. Vertical profile of sensible heat flux observed over Pocomoke forest on 16 September 2016. Points represent mean CWT fluxes accumulated over ~ 43 km (9 min) of flight along two side-by-side tracks spaced ~ 2 km apart (circles: west tracks, triangles: east tracks). CWT fluxes with $q_{coi} > 0.5$ are excluded from the averages. Error bars represent total errors (systematic plus random). Data are colored by the mean solar zenith angle (SZA) during the leg. The blue dashed line is an error-weighted least-squares fit for all fluxes with SZA below 38° . Note that fluxes for this flight ~~may be biased low by $\sim 18\%$ due to the~~ are corrected upwards by a factor of $1/0.76$ for the vertical wind spectral artifact discussed in Sect. 2.2. The boundary layer depth for this flight was $z_i = 1070$ m.

5

10



5 Figure 9. Example vertical wind - scalar cross covariance functions (a), CWT power spectra (b) and CWT cospectra (c) for a 31 km leg at 100 m altitude ($z/z_i \sim 0.1$) over Alligator River, NC on 26 May 2017. Cross covariance functions are normalized by peak covariance and time-shifted to align the peaks. Spectra are frequency-multiplied and (c) variance-normalized. The dashed line in (b) shows the expected $-2/3$ decay in the inertial subrange.



5 Figure 10. Comparison of water vapor mixing ratios (a) and latent heat fluxes (b) for all legs of the 2017 mission. Cyan dots represent 1 Hz average data. Blue lines are reduced major axis linear fits.

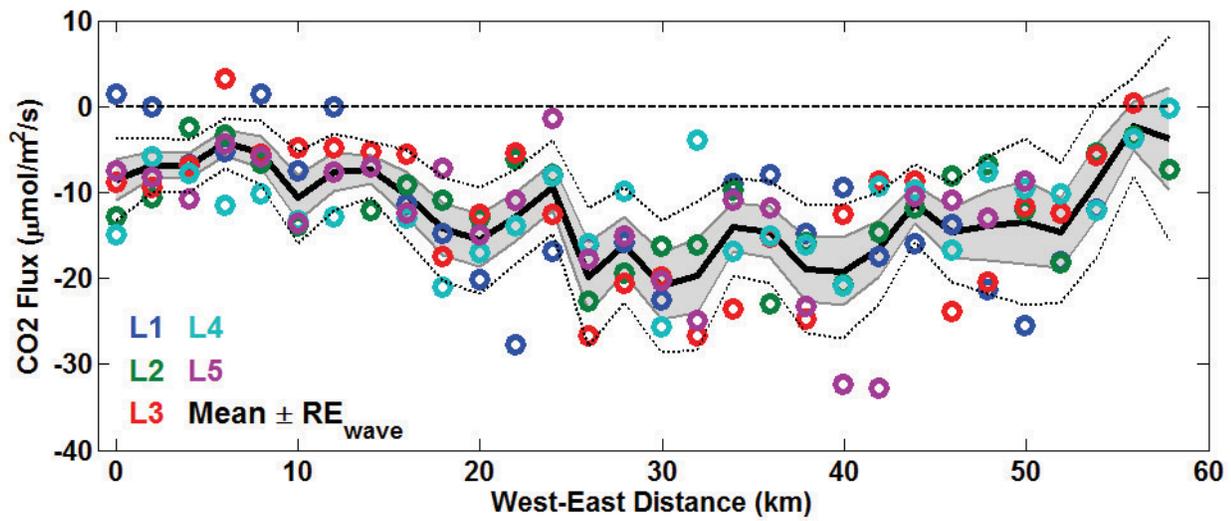


Figure 11. Comparison of horizontal CO₂ flux profiles sampled over the same region of mixed cropland and forest on 04 May 2017. Colored circles represent 2 km average fluxes from each of five consecutive legs. The solid black line, shading, and dotted lines are the mean, 1 σ random error, and 2 σ random error, respectively.

The NASA Carbon Airborne Flux Experiment (CARAFE): Instrumentation and Methodology

5 Glenn M. Wolfe^{1,2}, S. Randy Kawa¹, Thomas F. Hanisco¹, Reem A. Hannun^{1,2}, Paul A. Newman¹,
Andrew Swanson^{1,3}, Steve Bailey¹, John Barrick⁴, K. Lee Thornhill⁴, Glenn Diskin⁴, Josh DiGangi⁴,
John B. Nowak⁴, Carl Sorenson⁵, Geoffrey Bland⁶, James K. Yungel⁷, and Craig A. Swenson⁸

¹Atmospheric Chemistry and Dynamics Lab, NASA Goddard Space Flight Center, Greenbelt, MD, USA

²Joint Center for Earth Systems Technology, University of Maryland Baltimore County, Baltimore, MD, USA

³Goddard Earth Sciences Technology and Research, Universities Space Research Association, Columbia, MD, USA

10 ⁴NASA Langley Research Center, Hampton, VA, USA

⁵NASA Ames Research Center, Moffett Field, CA, USA

⁶Wallops Flight Facility, NASA Goddard Space Flight Center, Wallops Island, VA, USA

⁷AECOM, ATM Project, Wallops Flight Facility, NASA Goddard Space Flight Center, Wallops Island, VA, USA

⁸SSAI, ATM Project, Wallops Flight Facility, NASA Goddard Space Flight Center, Wallops Island, VA, USA

15 Correspondence to: Glenn M. Wolfe (glenn.m.wolfe@nasa.gov)

Supplementary Information

S1 PPFD corrections

The LI-COR LI-190R PPFD sensor is designed to be stationary and level; thus, corrections are required to account for sensor tilt under flight conditions. The attitude of the aircraft with respect to the Earth's surface is characterized by two angles: pitch (P) and roll (R). The sensor is not completely level with respect to the aircraft axes, necessitating additional terms to account for offsets (e_P and e_R). The sensor can also experience perturbations to the "direct sun" component, which depends on the aircraft heading (H , with potential offset e_H) and attitude as well as the solar zenith (θ) and azimuth (ϕ) angles. For example, if the sun lies to the south and the aircraft is heading north with a positive pitch (nose pointed upwards), the sensor will be pointed more towards direct sunlight than if it were fully level. The sensor can also experience shading under high roll or pitch conditions. These effects are evident in the raw flight data (Fig. S2S3).

25 These considerations give rise to a trigonometric empirical correction for varying Earth-sensor and sun-sensor angles. The correction takes the following form:

$$PPFD_{corr} = \frac{PPFD_{raw} \cos \theta}{\cos[\theta + (P + e_P) \cos(H + e_H - \phi) + (R + e_R) \sin(H + e_H - \phi)]} \quad (S1)$$

30 The offset angles ($e_P = 6.1^\circ$, $e_R = 2.6^\circ$, $e_H = 1.4^\circ$) are estimated by optimizing the correlation between corrected PPFD and $\cos \theta$ for two clear-sky flights. Fig. S2-S3 demonstrates the quality of this correction. In addition, PPFD data are discarded when roll angles exceed 5° to minimize sensor shading artifacts.

S2 Data gap filling

The alteration of a time series to eliminate gaps can markedly perturb wavelet-derived fluxes. Linear interpolation is adequate for gaps that are small relative to the integral time scale (of order 1 s), but it is a poorer approximation for larger gaps. It is also possible to “stitch” a time series together (literally removing the gaps). This method can create scale-dependent artifacts; in the Fourier analogy, this is similar to removing some portion of a sine wave and thereby creating a discontinuity. Figure S6-S8 shows example errors introduced through these gap-filling methods for an artificially-gapped segment of potential temperature flux observations. Errors are particularly pronounced near gap edges but can propagate throughout the whole leg. Note that this is only an example, and the magnitude and breadth of errors will depend on the width and location of a gap as well as the nature of the underlying observations.

To reduce potential artifacts near gap edges, we have developed an empirical method that we refer to as “covariance filling.” Covariance filling exploits the fact that we have two time series, x and w , that are expected to co-vary. For illustration, let us assume that only x contains gaps. The basic steps are as follows:

1. For each gap of width N , define a region of width $3N$ centered on the gap.

2. Calculate the covariance within the region using all valid data pairs.

$$\text{cov}(w, x) = \overline{(w - \bar{w})(x - \bar{x})} \quad (\text{S2})$$

Overbars indicate regional means that are restricted to times when both x and w are valid.

3. Assume the gap contains the same covariance as the region. Then the predicted standard deviation of x within the gap, σ_x , is calculated from the identity

$$\text{cov}(w, x) = r_{wx} \sigma_w \sigma_x \quad (\text{S3})$$

Here, σ_w refers to the standard deviation of w within the gap. The correlation coefficient r_{wx} is assumed equal to one, as we will directly scale w below.

4. Scale w within the gap to create a predicted time series of x and use this to fill the gap.

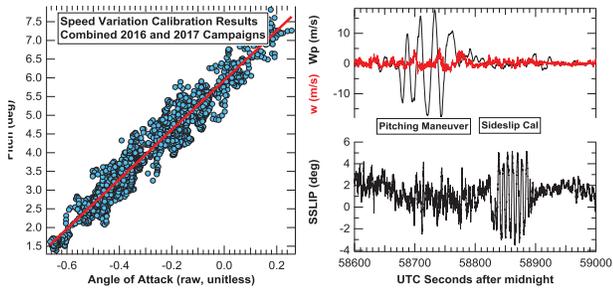
$$x_g(t) = \frac{\sigma_x}{\sigma_w} (w_g(t) - \overline{w_g(t)}) \quad (\text{S4})$$

As evident in Figure S6S8, covariance filling can reduce artifacts in derived fluxes compared to the interpolation and stitching methods. We find this to be true in general, based on a Monte Carlo experiment where 10,000 gaps of random size (1 to 600 points for 10 Hz data) and location were inserted into a potential temperature time series before wavelet flux calculation. Median maximum errors in this simulation are 30% for covariance filling, 45% for interpolation and 65% for stitching. Covariance filling does carry some limitations. Linear interpolation is used instead in the following cases: 1) a gap is small (less than 3 points), or 2) more than 50% of data pairs are missing in the region around a gap (due to adjacent gaps), or 3) overlapping gaps exist in both time series.

Table S1. Summary of flights for the 2016 and 2017 CARAFE missions.

Flight #	Flight Date	Flight Time (EDT)	Destination	Flux Legs	Tower Overpass^a
1	7 Sep 2016	12:25 - 15:59	Pocomoke Forest (MD)	6	
2	9 Sep 2016	11:23 - 15:20	Pocomoke Forest (MD)	13	
3	12 Sep 2016	11:06 - 15:03	Choptank Ag (MD/DE)	9	USDA-Chop, US-StJ
4	14 Sep 2016	11:07 - 15:17	Pine Barrens (NJ)	12	US-Slt, US-Ced
5	16 Sep 2016	11:21 - 15:10	Pocomoke Forest (MD)	15	
6	22 Sep 2016	11:58 - 16:00	Prince Frederick (MD) Charles County (MD)	6 8	
7	23 Sep 2016	11:12 - 15:05	Pine Barrens (NJ)	12	US-Slt, US-Ced
8	24 Sep 2016	10:47 - 14:35	Great Dismal (VA/NC) Alligator River (NC)	7 6	US-NC4
9	26 Sep 2016	11:04 - 15:10	Atlantic Ocean Chesapeake Bay Pocomoke Forest (MD)	4 5 2	
1	3 May 2017	15:00 - 17:27	Pocomoke Forest (MD)	4	
2	4 May 2017	11:12 - 15:03	Choptank Ag (MD/DE) Pocomoke Forest (MD)	7 3	USDA-Chop, US-StJ
3	8 May 2017	10:42 - 12:23	Pocomoke Forest (MD)	7	
4	8 May 2017	14:02 - 17:40	Pocomoke Forest (MD)	16	
5	9 May 2017	11:01 - 15:05	Pine Barrens (NJ)	16	US-Slt, US-Ced
6	10 May 2017	11:56 - 14:32	Prince Frederick (MD) Charles County (MD)	7 8	
7	15 May 2017	12:59 - 15:49	Alligator River (NC)	6	US-NC4
8	16 May 2017	11:08 - 15:02	Great Dismal (VA/NC) Chesapeake Bay	4 6	
9	18 May 2017	11:05 - 15:05	Choptank Ag (MD/DE)	10	USDA-Chop, US-StJ
10	19 May 2017	10:04 - 13:29	Pocomoke Forest (MD)	13	
11	26 May 2017	08:51 - 12:55	Alligator River (NC)	11	US-NC4

^aAbbreviations: USDA-Chop = Choptank, MD; US-StJ = St. Jones Preserve, DE; US-Slt = Silas Little, NJ; US-Ced = Cedar Bridge, NJ; US-NC4 = Alligator River, NC. More information on towers available at the ameriflux.lbl.gov/ and <https://www.ars.usda.gov/northeast-area/beltsville-md/beltsville-agricultural-research-center/hydrology-and-remote-sensing-laboratory/docs/research-sites/>.

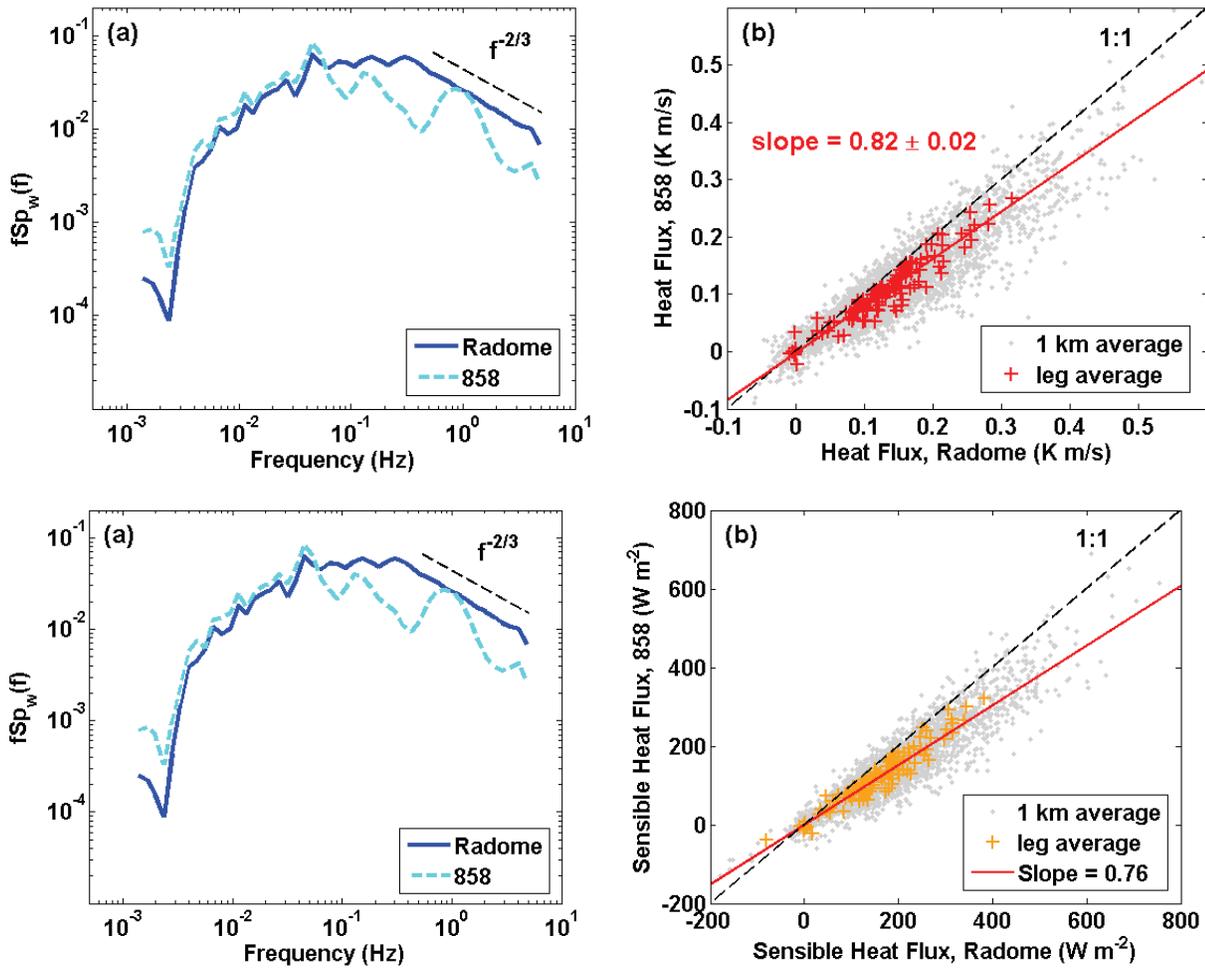


(a)

(b)

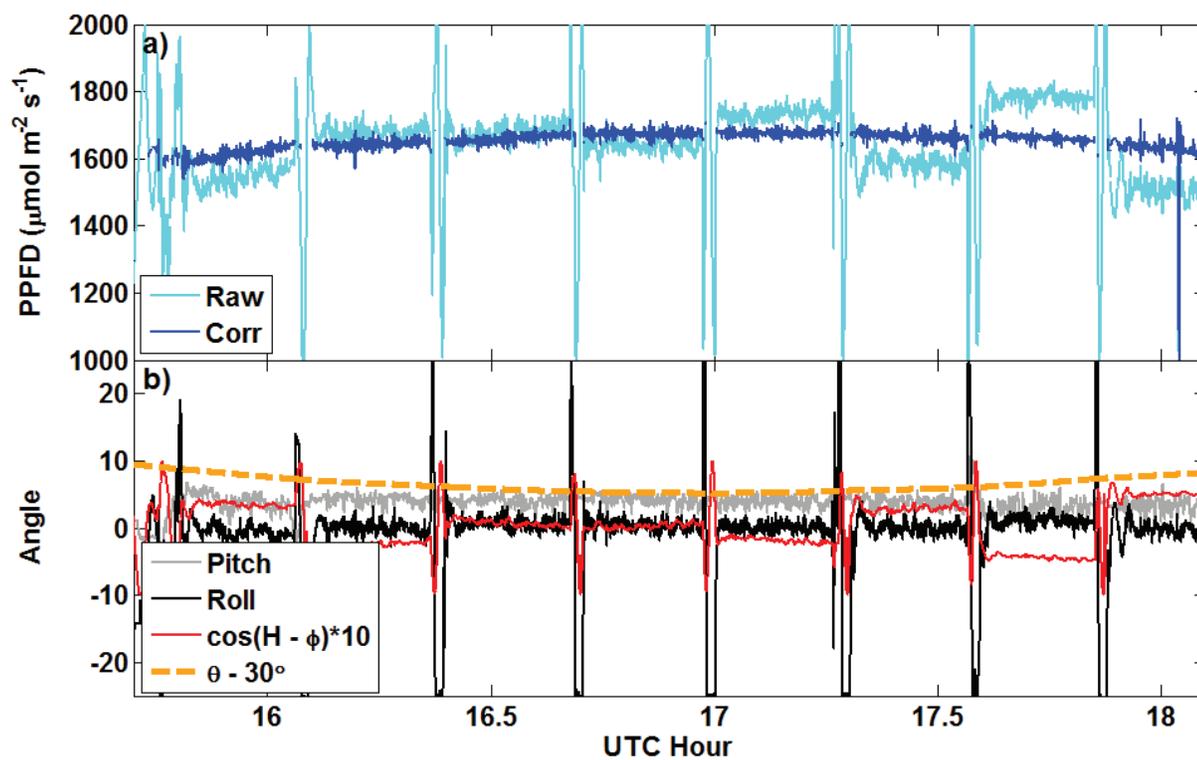
(c)

Fig S1. (a) Results for the angle of attack response during all speed variation maneuvers during both the 2016 and 2017 campaigns. (right) Derived vertical wind speed (red) measured during pitching maneuvers and sideslip calibrations showing that measured vertical wind speed is unaffected by changes in both pitch (b) and sideslip angle (c).



5 Figure S4S2. (a) Example frequency-multiplied vertical wind spectra for a single leg from the 2017 mission. Spectra are calculated from wavelet transforms for vertical wind data from the radome (solid blue) and 858 probe (dashed cyan) systems. The black dashed line denotes the $-2/3$ slope expected in the inertial subrange. (b) Comparison of kinematic sensible heat fluxes for all 2017 flights calculated using vertical winds from the radome or 858 probe. Gray dots represent 1 km average fluxes, red-orange crosses are averaged over whole legs. The red line is a represents the systematic error in 858 fluxes, based on a zero-intercept linear regression through all 2017 fluxes. The slope of this line is used to correct for the systematic bias in 2016 fluxes as described in Sect. 2.2. the leg average data.

10



5 | Figure S2S3. (a) 1 Hz PPFDF observations from the 12 September 2016 flight under mostly clear-sky conditions. Raw data is shown in cyan, corrected/filtered data in blue. (b) Corresponding aircraft attitude (pitch: gray, roll: black) and solar zenith angle (orange dash) and cosine of difference of aircraft heading and solar azimuth (red). The latter is the relevant quantity for sun-sensor angle corrections.

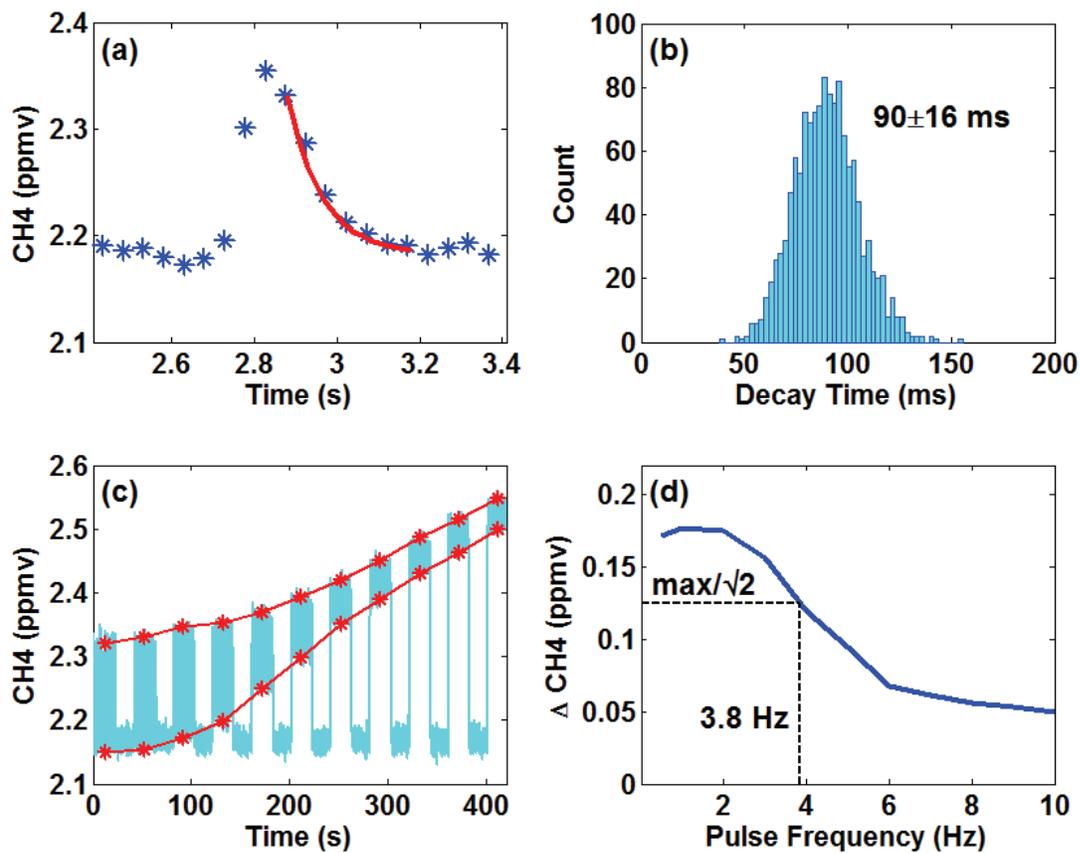
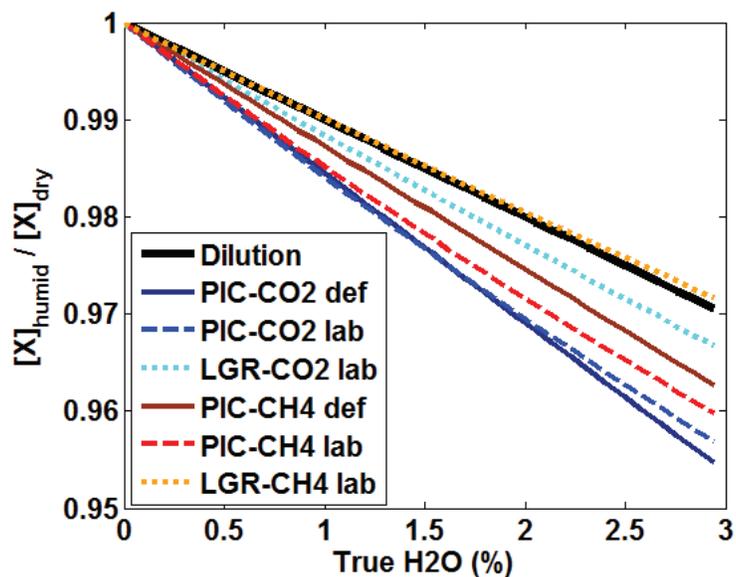


Figure S4. Laboratory tests to characterize the time response of the LGR/nXDS15i combination. In these experiments, a miniature solenoid valve (The Lee Company) was used to add 5 ms pulses of 100 ppmv CH₄ to room air while sampling at 20 Hz. Each pulse exhibits an exponential decay (a), and fitting the first 0.3 s of 1300 such decays gives a characteristic e-folding time of 90 ± 16 ms (b). We also performed a test analogous to that described in Aubinet et al. (2016), wherein the time between pulses is reduced (pulse frequency increased) and pulses “smear” together (c). The “cutoff frequency” is defined as the point where the difference between maximum and minimum concentrations decreases by $\sqrt{2}$ (d).



5 | Figure S3S5. Empirical correction factors for conversion of ambient CO₂ (blues) and CH₄ (reds) observations from humid to dry mixing ratios. Correction factors are based on laboratory measurements of calibration gas with varying humidity levels controlled by a bubbler/dilution system. Correction factors inherently include both dilution and spectroscopic effects. Also shown for comparison is the correction factor for dilution only (solid black line) and the default correction factors for the Picarro G1301-m (solid blue and red lines). “True H₂O” refers to the calibration-corrected Picarro water vapor mixing ratio.

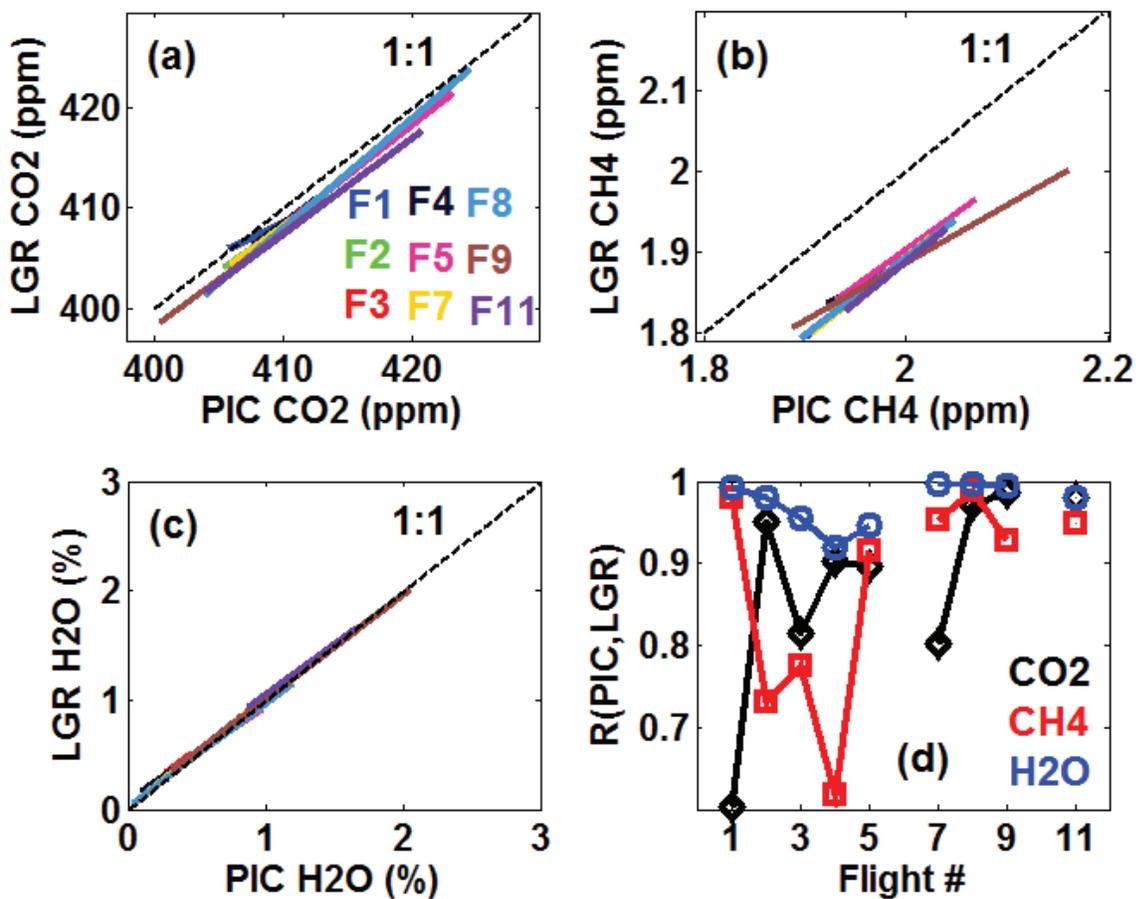


Figure S6. Least squares fit lines (a-c) and correlation coefficients (d) for Picarro and LGR dry mixing ratios of CO₂, CH₄ and H₂O obtained during flights in 2017. Fits are colored by flight and shown only over the range of mixing ratios observed on each flight. Dashed lines in a), b) and c) denote a 1:1 correlation. Fits are not available for flights 6 and 10 due to a malfunction of the Picarro instrument; in these cases, mission-averaged fit coefficients were used to correct the LGR data.

5

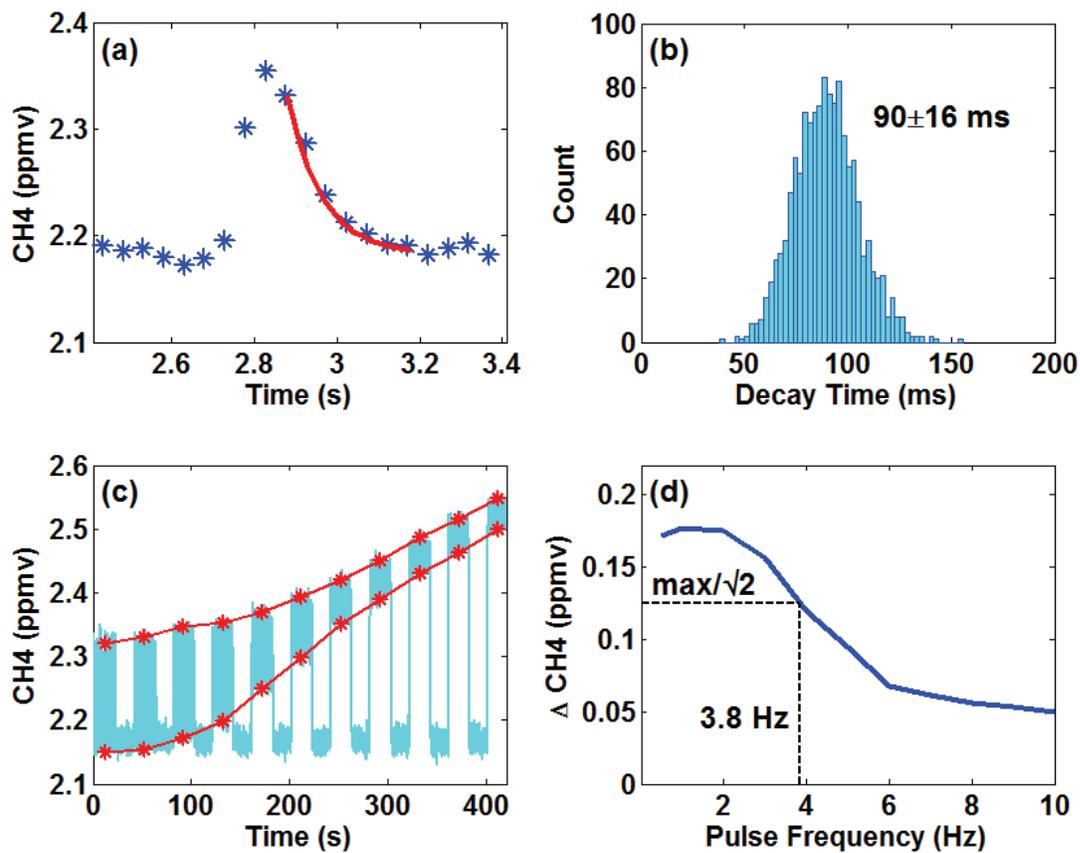


Figure S4. Laboratory tests to characterize the time response of the LGR/nXDS15i combination. In these experiments, a miniature solenoid valve (The Lee Company) was used to add 5-ms pulses of 100-ppmv CH₄ to room air while sampling at 20 Hz. Each pulse exhibits an exponential decay (a), and fitting the first 0.3 s of 1300 such decays gives a characteristic τ -folding time of 90 ± 16 ms (b). We also performed a test analogous to that described in Aubinet et al. (2016), wherein the time between pulses is reduced (pulse frequency increased) and pulses “smear” together (c). The “cutoff frequency” is defined as the point where the difference between maximum and minimum concentrations decreases by $1/\sqrt{2}$ (d).

5

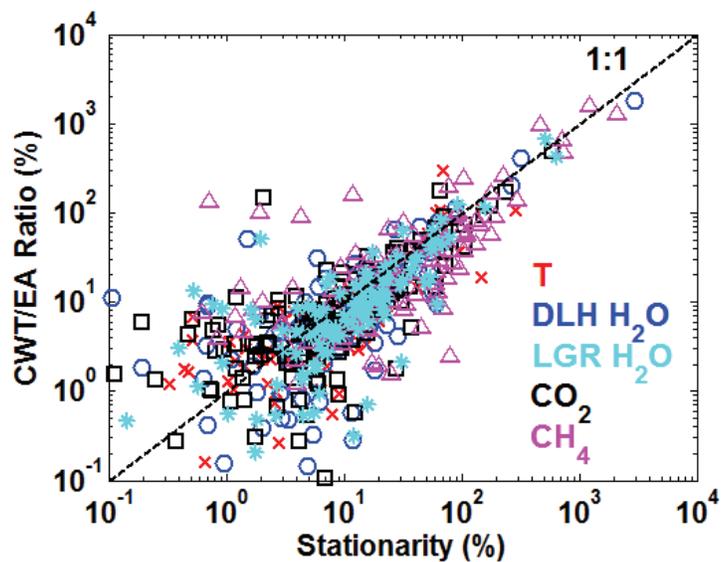
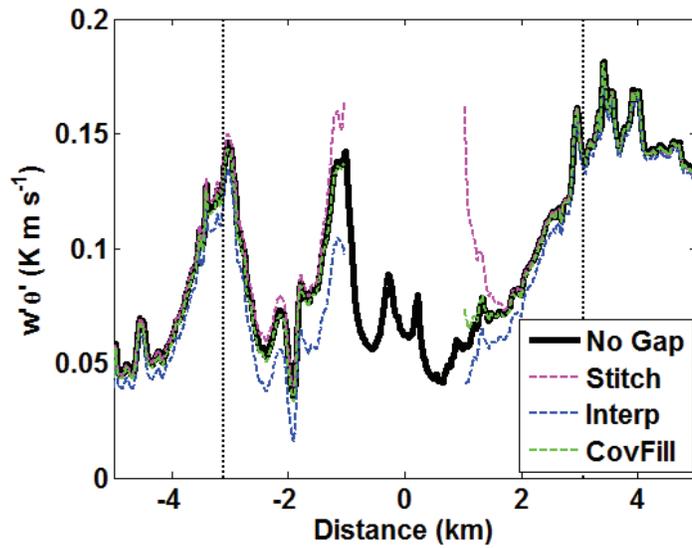


Figure S5S7. Comparison of stationarity and the ratio of CWT to EA fluxes for all scalars and flight legs of the 2017 mission. Stationarity is defined as described in the main text (Eq. (2)) but cast in percentage units. Analogously, the CWT/EA flux ratio is
 5 defined as $100 \times |1 - \langle F_{CWT} \rangle / F_{EA}|$.



5 | Figure S6S8. Example of gap-filling effects on wavelet fluxes. The solid black line shows a portion of a sensible heat flux time series derived from wavelet transforms. A gap of random size (293 points) and location was inserted into the potential temperature time series, and wavelet fluxes were recomputed using one of three gap filling methods: stitching (magenta), linear interpolation (blue), or covariance filling (green). Perturbations near the gap edges are evident. Vertical dotted lines indicate the window within which fluxes would be discarded due to potential influence from gap filling. The window width is thrice the size of the gap and centered thereon.

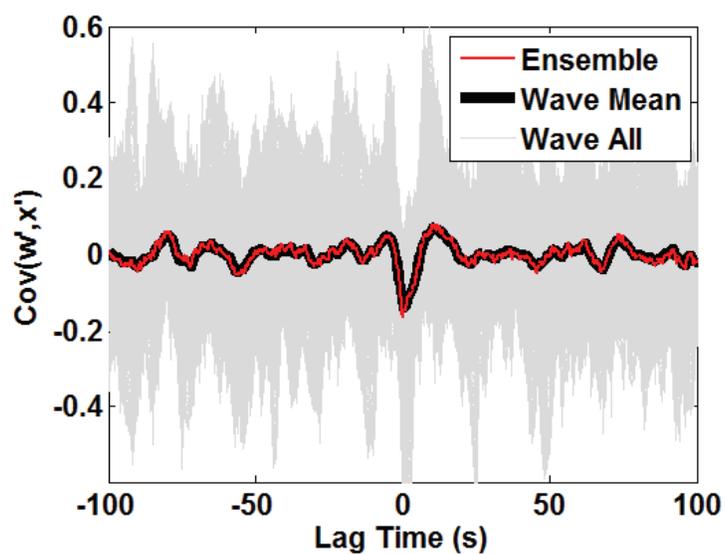
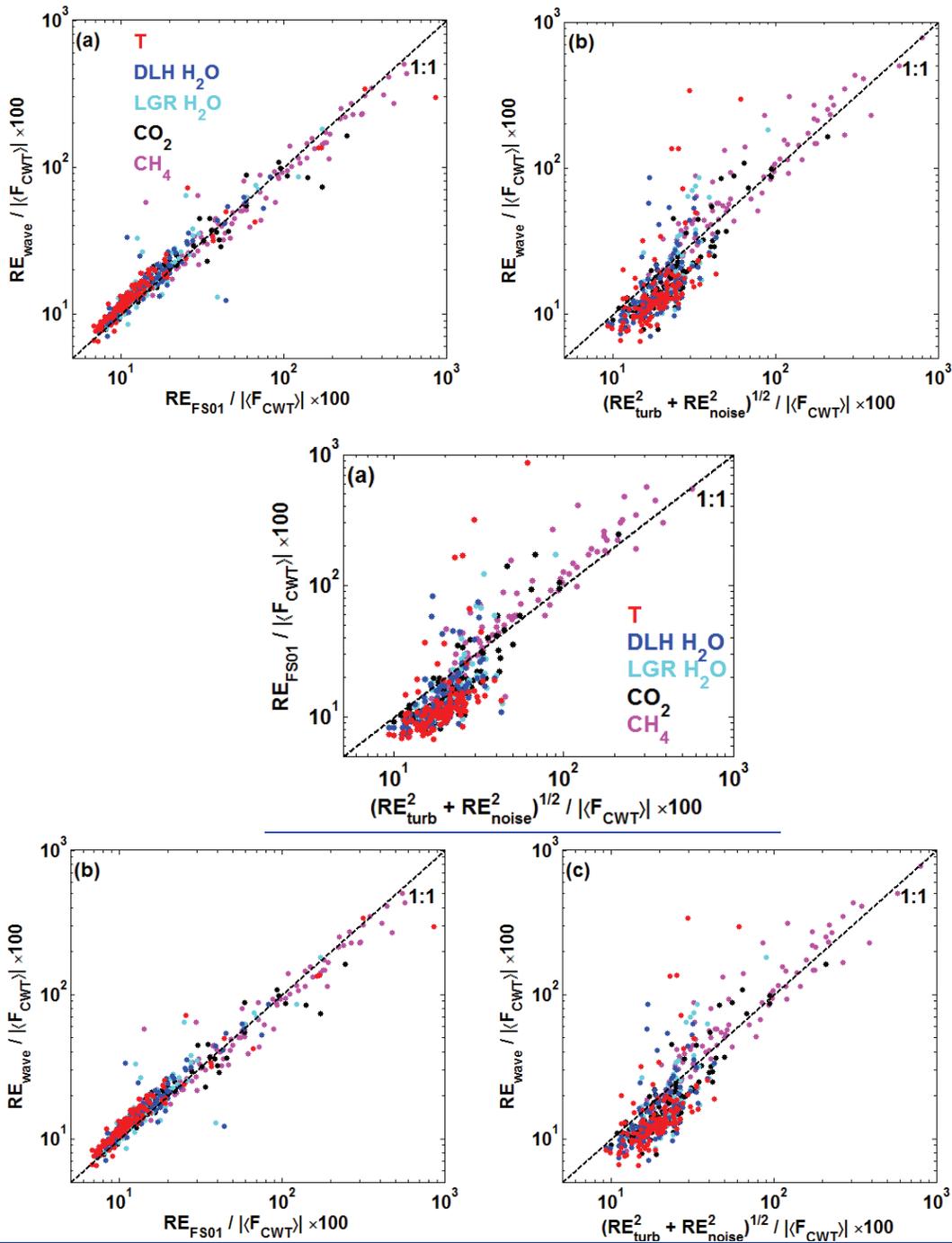


Figure S7. Example of wavelet lag covariance calculation (Sect. 3.4.2) for a single leg of CO_2 fluxes. Gray lines represent 1-Hz wavelet cross-covariance functions of vertical wind and CO_2 for all 960 points in the time series. The mean of all wavelet cross-covariances (black) is nearly identical to the cross-covariance calculated on the ensemble dataset (red).

5



5 Figure S98. Comparison of random error estimates: (a) ensemble empirical vs semi-theoretical, (b) wavelet-derived random errors against ensemble average statistical errors vs ensemble empirical, and (c) wavelet-derived vs semi-theoretical. (a) and the root-sum-square of turbulence and noise errors (b). Errors are shown for all legs and fluxes in the 2017 mission. Errors are given in as

percentage units relative to leg-averaged fluxes. Recall that RE_{turb} is a theoretical upper limit, hence we expect $(RE_{turb}^2 + RE_{noise}^2)^{1/2}$ to exceed RE_{FSOI} and RE_{wave} in cases where turbulence is the dominant driver of total random error.

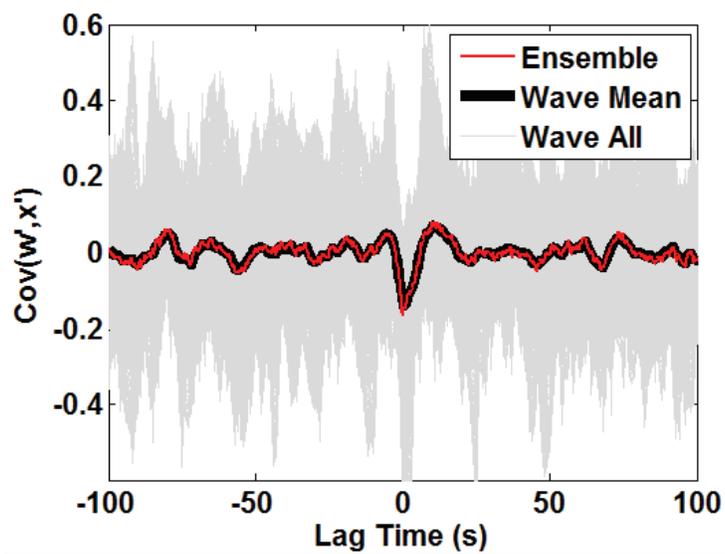


Figure S107. Example of wavelet lag-covariance calculation (Sect. 3.4.2) for a single leg of CO₂ fluxes. Gray lines represent 1 Hz wavelet cross-covariance functions of vertical wind and CO₂ for all 960 points in the time series. The mean of all wavelet cross-covariances (black) is nearly identical to the cross-covariance calculated on the ensemble dataset (red).

5

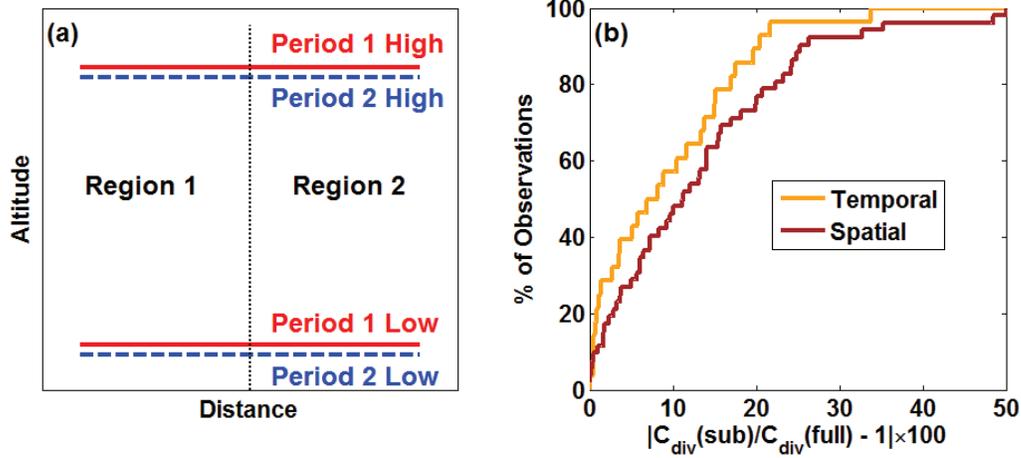


Figure S9S11. Illustration of tests for divergence correction variability in time and space. (a) Cartoon depicting the distribution of flux legs. For a given flight, two periods separated by 1.5 – 2.2 hours are chosen containing a high and low level leg over the same horizontal area. Fluxes of CO_2 , temperature and H_2O (DLH and LGR) are used for these tests. For the temporal variability test, divergence corrections (C_{div}) for period 1 and period 2 (“sub-periods”) are compared to C_{div} calculated using all four legs (“full”). For the spatial variability test, C_{div} for regions 1 and 2 (“sub-regions”) are compared to C_{div} calculated using fluxes from both regions (“full”). The latter test is done separately for each period. (b) cumulative distribution of percent deviation of sub-period and sub-region C_{div} values from their “full” counterparts (essentially a measure of deviation relative to an “average” divergence correction).