This manuscript presents an analysis of random error in the measurement of eddy correlation trace gas fluxes at sea and the effects of measurement error on the interpretation of direct air-sea gas transfer velocity observations. The analysis employs field measurements from four research cruises – two N-S transects of the Atlantic and two high-latitude Arctic projects. The study includes two state-of-art measurement systems for carbon dioxide flux – a broadband infrared gas analyzer (IRGA, LI-7200) and a laser cavity ring-down spectrometer (CRDS, Picarro G2311-f).

This manuscript provides a very useful overview of various approaches developed over the years to assess random error in flux measurements and analyzes these methods under conditions where the covariance signal is often near the measurement detection limit in the presence of various interferences such as platform motion, flow distortion and large water vapor fluxes.

The paper is very well written and well organized. I don't have significant comments with respect to usage or punctuation and will confine the following comments to a few issues of substance. Overall, this is a very good paper and a welcome contribution to the field and should be accepted after checking a few of the issues mentioned below.

I am not sure about the merits of the CO_2/H_2O decorrelation of the LI-7200 data, described on lines 183-185 (based on Landwehr et al. 2018). This procedure has potential to remove real turbulent flux signal for CO_2 since the water vapor and CO_2 fluxes are both driven by the same turbulent eddies, and therefore correlated with each other. L2018 state that due to the long inlet lag time and air drier in their configuration the gas signals are decoupled from the vertical wind measurements (which is true) and therefore this decorrelation doesn't remove real flux signal (which I'm not sure about). The decorrelation applied here is not with respect to vertical wind – it is between the two gas concentrations measured simultaneously by the same analyzer. If these signals have approximately the same lag time, then it seems to me the decorrelation could indeed remove actual CO_2 flux signal by removing variance due to low-frequency turbulent eddies present in both signals which pass through the air drier (the drier is basically a low-pass filter on the water vapor signal). Did this decorrelation yield a significant adjustment to the measured fluxes? If not, maybe it's unnecessary.

Did the authors check for a positive bias to the CO_2 fluxes due to the demonstrated crosstalk between water vapor and CO_2 signals in the IRGA? The use of a drier to precondition the sample air is necessary to remove this artifact, and I'm sure the authors approach is fairly effective in this respect. But it might be useful to check the correlation/covariance/cospectra of the water vapor and vertical wind signals on AMT29 to see if low-frequency latent heat flux signal is nevertheless bleeding through the drier and affecting the CO_2 measurement (as mentioned above). (Note, the lag time adjustment may be a bit different for water vapor and CO_2.) I mention this because the AMT cruises are the primary comparison between the two methods and the corrected IRGA CO_2 fluxes on AMT29 are a bit larger than those from the CRDS on AMT28, which is what you might expect if water vapor cross talk is bleeding into the IRGA CO_2 flux measurement (at equatorial latitudes where we expect large latent heat fluxes!).

Of course, there could be other reasons for the observed difference between cruises separated by a year, as mentioned by the authors on p.23. It's a shame both analyzers were not deployed simultaneously on one of the cruises.

There seems to be an error in equation 5. Flux uncertainty goes as the square root of sampling time and the entire fractional term on the RHS of this equation should be to the ½ power. The authors have chosen to use the square root of the product of the two integral time scales in the numerator, which is different from the more common minimum value of the two integral scales, but this is OK. The missing square root may be just a typo, but if this equation was in fact used to estimate error, then that should be recomputed.

I can provide an update for the discussion of the integral time constant in Appendix B. Equation B2 in this manuscript and the associated stability function (both from Blomquist et al. 2010) are a bit dated. They were based on measurements from R/P Flip during the SCOPE field campaign and do not include much information for stable conditions. A more recent analysis (as yet unpublished) of the entire NOAA PSL flux database (41 research cruises spanning 21 years) has updated the empirical relationship for τ as a function of the nondimensional frequency maximum of the cospectrum, η_m

$$\tau = \frac{z}{2\pi U_r \eta_m}$$

Where the best fit for η_m as a function of z/L is given by

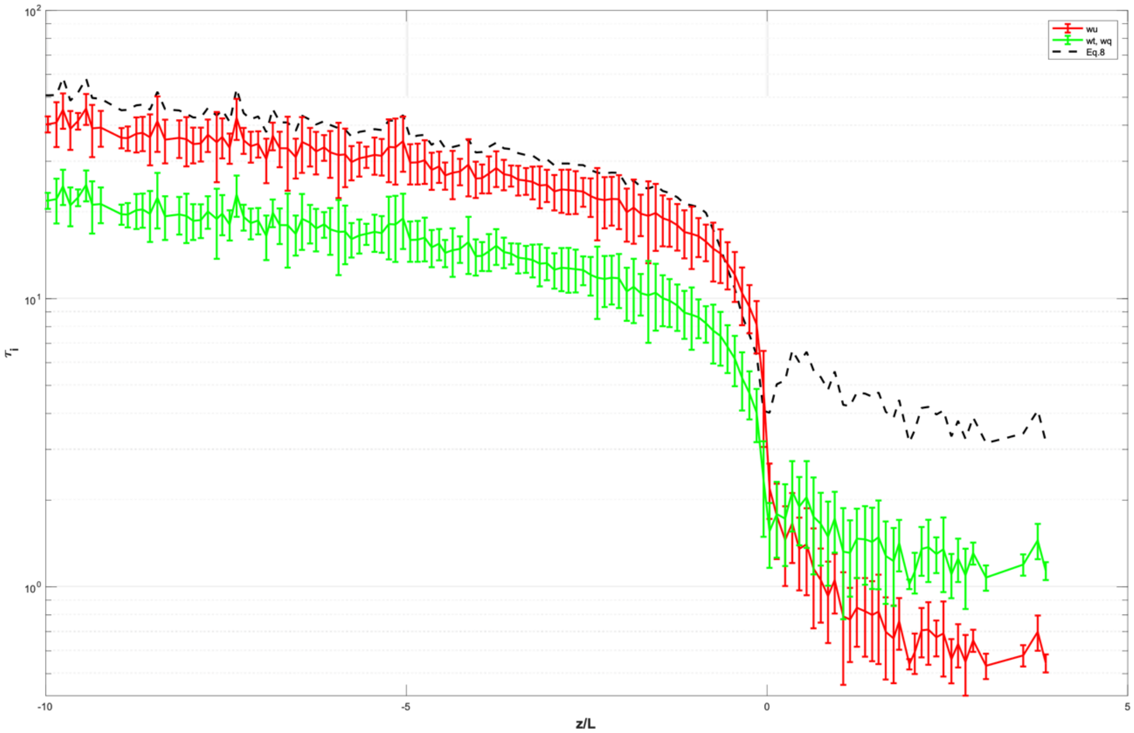
$$\eta_m(z/L) = A1 + \frac{1}{A2 + A3|z/L|} \quad for\ z/L < 0$$

$$\eta_m(z/L) = B1 + B2(z/L)^{2/3} \quad for\ z/L > 0$$

And the best-fit constants A and B differ for momentum and scalar fluxes:

	$A1$	$A2$	$A3$	$B1$	$B2$
$\eta_m\ (w'u')$	0.033	25	400	0.069	0.42
$\eta_m\ (w't')$	0.06	13	120	0.134	0.16
$\eta_m\ (w'q')$	0.06	33	120	0.089	0.20

I've attached a plot below, where the black dashed line represents Equation B2 and the green line is the updated scalar flux integral time constant from the equations above. U_r differs a bit between z/L bins in the flux database, which causes a little scatter in the trend of each line, but it's clear the updated function in green yields a time constant considerably smaller than Equation B2 in black, especially in stable conditions, and this is more or less in agreement with what is shown in Figure B1 of this manuscript.



I’m not suggesting you include all this in the manuscript, but you can mention that based on recent analysis the equation B2 formulation is now thought to be an overestimate.

Note, your figure caption for Fig B1 has a couple typos: the peak frequency equation is B3 and the similarity relationship is B2.

I’m convinced the authors have demonstrated their principal conclusion – that for state-of-art gas analyzers sampling error is a more important contributor to flux uncertainty than analyzer noise, and this is the reason why we usually need to average over hourly timescales to achieve reasonable measurement precision. This is also why it is very difficult to make credible CO_2 flux measurements in the presence of significant turbulent disruptions and pollution plumes or other sources of CO_2 variability related to airmass advection. Threshold criteria for stationarity and homogeneity are sometimes also helpful in reducing measurement uncertainty.