Review response to discussion article “Towards standardized processing of eddy covariance flux measurements of carbonyl sulfide” by Kohonen et al.

Reviewer comments in black
Author response in purple
Altered text in the manuscript in italic

Reviewer #1: Georg Wohlfahrt

General assessment: Kohonen et al. report on the effects of varying various post-processing steps required for eddy covariance COS flux measurements with the aim, as stated in the title, to standardize these. COS EC flux measurements are increasingly making their way into the literature as COS offers a novel means of constraining GPP and stomatal conductance. Yet, the necessary processing steps are way not as harmonized as is the case for CO2, potentially causing systematic bias between studies using different processing schemes, thus impeding synthesis activities. Overall I think this is a timely and relevant addition to the literature, which fits with the scope of the journal. I though also believe that the manuscript suffers from several issues, which will require significant changes, as detailed below.

Major comments:
(1) First, I have several formal issues with the manuscript: English style is often poor, which creates situations in which the intended meaning is not entirely clear (e.g. l. 32 the explanation of footprint limitations during stable stratification). Some of the formulations are too sloppy and thus misleading (e.g. l. 33 where “operation at high frequency” is mixed with “fast time response”). Some text is trivial or circular (e.g. l. 434-435), some of the concepts are wrong (e.g. l. 49) and some information is missing (e.g. legend of Fig. 11). Often some later, in-house studies are cited instead of the original papers. Finally, a mix of tenses is used when typically the past tense should be used to describe own results.
We will revise the manuscript and try to improve the english language, avoid circular text and repetition and only use the past tense. We cite the original papers as suggested. The caption of Fig. 10 (former Fig. 11) will be improved.

(2) Novelty and justification of the study: In 2017 a methodological paper on COS EC flux measurement post-processing was published in the same journal (Gerdel et al.). The authors justify their paper mainly by stating that their analysis goes beyond this previous paper. While this is partially true (in particular the analyses on lag times is novel), I think the authors should follow what the title of the paper suggests and rather sell their work as contributing towards a standardization of COS EC flux post-processing routines. To this end, I suggest to synthesize, e.g. in a table, the various processing steps that were used by previously published studies as a starting point and use this as a backbone for their analysis and the resulting recommendations. This table would then summarize whether and if so how previous studies detrended their time series, how the lag time was found, how low/high-frequency response corrections were applied, whether data were filtered for low u* (how were thresholds found) and which QC/QA was used. Following this suggestion requires at least the introductory section to be more or less completely re-written and would allow the paper to live up to what its title suggests and eventually become a reference for COS EC flux measurements.

Thank you for this suggestion. The Introduction section will be reorganized and partly rewritten. The study's objectives will be more clear. A table summarizing previous studies is a very good idea and will be implemented as Table 1 in the new version of the manuscript. We will revise the whole manuscript and extend the discussion to processing routines used in earlier studies.

(3) Vertical advection: This section is somewhat odd – the authors acknowledge that knowing the magnitude of vertical advection is meaningless unless the magnitude of horizontal advection is known as well, yet vertical advection is reported even though horizontal advection has not been quantified. Unless the authors can come up with a discussion of what their results on the sign and magnitude of vertical advection actually mean in the context of their study, I thus suggest removing the results on vertical advection and all text/material that pertains to it.
We have considered this point carefully and came to the conclusion to leave out the results regarding vertical advection. The reviewer has a good point and as we are aiming for harmonization of processing protocols - where vertical advection is not used - we decided to leave this section out of the revised manuscript.

(4) Corrections for high-frequency flux loss: Comparing two different approaches is novel for COS, yet surprisingly none of the underlying results are shown – I suggest to expand this section.
We discuss and show the difference to the reference processing scheme and show in Table 2 the effect of spectral corrections to final fluxes. Histograms and PDFs of different spectral correction schemes will be added to the Supplement (Fig. S2) and daytime and night-time median fluxes added to Table 2 and discussed in the text. We will also add a figure on flux attenuation to Supplementary material Fig. S3 and scatter plot comparing the final fluxes to Fig. S7.

(5) Changes of co-spectral peak frequency with stability: Among the results of this study is a figure comparing the changes in the co-spectral peak frequency with stability for the Horst model and this study. While interesting, this analysis and the results are not motivated in the introduction and are barely discussed. Again, unless the authors are able to come up with a discussion of what the observed differences mean for their study, I suggest removing this material (or possibly moving it into a supplement).
We will move the figure to supplementary material (Fig. S4), as suggested. Equations 15 and 16 of the former version will be moved to Methods-section and presented as equations 6 and 7 in the revised version of the manuscript.

(6) Gap-filling: This is an indeed novel aspect, however way underexploited by the authors. Only a single arbitrarily chosen gap-filling algorithm is tested, the authors miss to put it to a true test and results of gap-filling (e.g. time course of estimated parameters and selected results illustrating gap-filling behaviour) are lacking.
A time series of gap-filling parameters will be shown in the supplementary material Fig. S5. We will add a diurnal plot of the measured flux compared to different gap-filling functions in Fig. 11 in the revised manuscript as well as residuals of different methods in Fig. S6 in the supplementary material. We are not going into further detail in comparing different gap-filling methods, as that would result in a whole new paper. Long-term budgets are not usually the interest of the COS community, as GPP calculations from COS often aim at understanding the CO2 exchange dynamics rather than calculating long-term budgets. However, it is important to fill short gaps (individual 30 mins or a bit longer) to e.g. get diurnal variations. The presented gap-filling method is just one example that can be used for gap-filling COS data. We will add discussion in Section 3.6

(former Section 3.7): "*Three combinations of environmental variables (PAR, PAR and relative humidity, PAR and VPD) were tested using the gap-filling function Eq. 16. These environmental parameters were chosen because COS exchange has been found to depend on stomatal conductance, which in turn depends especially on radiation and humidity (Kooijmans et al., 2019). Development of the gap-filling parameters a, b, c and d over the measurement period is presented in the Supplementary material Fig. S5. While saturating function of PAR only captured the diurnal variation already relatively well, adding a linear dependency on VPD or RH made the diurnal pattern even closer to the measured one (Fig. 11). Therefore, the combination of saturating light response curve and linear VPD dependency was chosen. Furthermore, we chose a linear VPD dependency instead of a linear RH dependency due to smaller residuals in the former (Fig. S6).*"

(7) Li-6262 and QCL CO2 cross-covariance maximisation: An important detail of this study is the setup, which includes a closed-path IRGA that is used to measure CO2 and H2O concentrations (not clear whether from the same tube as the QCL). These data are used to account for the drift in the computer clocks acquiring the sonic (& IRGA) and QCL data, respectively. While this nicely shows the benefit of having a complementary suite of measurements at "super-sites" such as Hyytiälä, in my view the reliance on an additional instrument is a drawback of this study as it limits the applicability of the proposed approach at other sites where no additional IRGA or an open-path model or closed-path model with a short tube is deployed. Even more so, this approach is unnecessary, as there are simpler, software-based, solutions available to keep computer clocks in a network synchronized. In addition, by aligning data this way, the authors ruin a truly independent means of cross-comparing the QCL CO2 and H2O fluxes. I thus think it would be useful to explore the possibility of aligning the data sets in time without the help of the IRGA data. This is possible by expanding the time window in which the lag determination algorithm searches for, as with computer clocks being reset only once a day, time shifts of several seconds may result (in both directions). The reliability of this approach may then be checked by comparing against lag times and fluxes calculated with the IRGA.

We thank the Referee for this point. However, we recognize that standard CO2/H2O flux towers measuring other gases than CO2 typically use QCLs, which are nowadays providing also H2O and CO2, beside the target gas (CH4, N2O, COS, etc). In our study we have taken advantage of this setup. We acknowledge that better synchronization approach should be used already in the data logging system and a short text will be added to Section 2.2. Moreover, we have considered the Referee's suggestion of using a larger time window for COS, but it would be problematic because the covariance peak is not always clear and with a large time window there might be multiple peaks (for example related to low frequency variations). Instead, we have tested the file combination maximizing the covariance of CO2 (QCL) and w, and this resulted in a very similar outcome as the previous method. The results are shown as histograms in the supplement (Fig. S1). We rewrote this part of the manuscript as: "*The following procedure was done to combine two data files of 30 min length (of which one includes sonic anemometer and LI-6262 data and the other includes Aerodyne QCLS data): 1) the cross-covariance of the two CO_2 signals (QCLS and LI-6262) was calculated 2) the QCLS data were shifted so that the cross-covariance of the CO2 signals was maximized. Note that this will result in having the same lag time for QCL and LI-6262. The time shift was a maximum of 10 seconds, with most varying between 0 s to 2 s during one day. It is also possible to shift the time series by maximizing the covariance of CO_2 and w, which will then already account for the lag time (Fig. S1) or combine files according to their time stamps and allow a longer window in which the lag time is searched. However, in this case it is important that the lag time (and time shift) is determined from CO_2 measurements only, as using COS data* *might result in several covariance peaks in longer time frames due to low signal-to-noise ratios and small fluxes.*"

(8) Conclusions: The authors should conclude with referencing against what processing steps have been used by previous studies and put their results into perspective with these, by highlighting critical steps and the need for further harmonization.
As suggested by the Referee, we will add more discussion related to previous studies in the "Results and Discussions" chapter. Moreover, in the "Conclusions" we will highlight more clearly the critical steps and needs for further harmonization.

Detailed comments:
l. 1-3: reformulate to better convey intended meaning "... growing in popularity with the aim of estimating gross primary productivity at ecosystem scale, however lack standardized protocols ..."
Clarified. The sentences now read "*Carbonyl sulfide (COS) flux measurements with the eddy covariance (EC) technique are becoming popular for estimating gross primary productivity. To compare COS flux measurements across sites, we need standardized protocols for data processing.*"

l. 17: replace "due to the use" with something more suitable, e.g. "motivated by ..."
Corrected as suggested.

l. 20: ", but in contrast to CO2, COS is destroyed..."
Corrected as suggested.

l. 22: for readers not familiar with the LRU, talking about the radiation-dependency of the LRU without introducing the concept might be highly confusing
Text about LRU removed.

l. 29: "... the assumptions underlying the EC method ..."
Corrected as suggested.

l. 29-34 and l. 35-41: these two paragraphs are in my view too general to meaningfully add to the introduction
We reduced the paragraphs into one sentence: "*To meet the assumptions underlying the EC method, the site, setup design, and instrumentation need to be considered (Aubinet et al., 2000, 2012; Nemitz et al., 2018; Sabbatini et al., 2018).*"

l. 32: reformulate – what you likely mean is that the measurement height should be such that the footprint remains within the ecosystem of interest even during stable stratification
This sentence will be removed as we will harmonize the Introduction section and make it more compact.

l. 33: EC instruments need to have a fast time response, which is different from "operation at high frequency" (a slow-response sensor does not become suitable for EC only because its data are logged at 20 Hz), as it can be shown that fast-response measurements made every few seconds do not cause a systematic bias in the EC flux
Corrected.

l. 38: not sure this sentence applies universally to all closed-path analyzers and anyway I would think this is too much detail for the introduction – suggest to remove
Sentence removed.

l. 41: the first ones to report on this issue were Ibrom et al. (2007)
This sentence will be removed as we harmonize the Introduction section and make it more compact.

l. 43-44: a rotation into the prevailing wind direction is only one step in the coordinate rotation; typically the aim is to align the coordinate system with the prevailing streamlines (2D or 3D) or with respect to some coordinate system that was established over a longer period (e.g. planar fit)
Corrected.

l. 49: the EC flux is fine – the problem is that it may represent a poor estimate of the surface-atmosphere exchange under these conditions
Corrected.

l. 53: in fact it was the following year (1999) that John Finnigan published a commentary on the Lee (1998) paper in which he demonstrated that correcting only for vertical advection is nonsense
All text and discussion regarding vertical advection will be removed from the revised manuscript.

l. 56: if environmental data are lacking too, mean diurnal variation may be used as a last resort
Added mean diurnal variation.

l. 64: if the cross-covariance is flat, then a wrong lag time will not have a large effect
We agree that in case the covariance is flat, lag time does not make a large effect. But in the case when cross-covariance is not flat but noisy, and lag time determination is difficult, it affects the flux magnitude.

l. 66: I do not get the "However, . . ." which links to the previous sentence, which does not appear to make sense here
Removed "However,".

l. 69: Gerdel et al. did study lag determination (their section 3.1) and u*-filtering (their Fig. 6)
It was meant here that different methods for lag time determination have not been studied earlier. We have changed the sentence to "*Gerdel et al. (2017) describes the issues of different detrending methods, high-frequency spectral correction, lag time determination and u* filtering. However, there has not been any study comparing different methods for lag time determination or high frequency spectral correction in terms of their effects on COS fluxes.*"

l. 71: actually you do not discuss the "EC flux measurement setup" at all
Removed.

l. 71-73: the introduction should finish with a statement of objectives
We will revise the whole introduction section and end it with a clear statement of objectives: "*In this study, we compare different methods for detrending, lag time determination and high-frequency*

spectral correction. In addition, we compare two methods for storage change flux calculation, discuss the nighttime low turbulence problem in the context of COS EC measurements, introduce a method for gap-filling COS fluxes for the first time and discuss the most important sources of random and systematic errors. Through the evaluation of these processing steps, we aim to settle on a set of recommended protocols for COS flux calculation."

l. 79: coordinate rotation
Corrected.

l. 78-88: does this have any relevance for this study?
We think it is important for the reader to know about the measurement site and its characteristics. Also, we want to mention that the same data has been published before in a paper focusing on different aspects than methodology.

l. 92: and sonic temperature
Corrected.

l. 95: flow rate through Li-6262, same pump as QCL, tube diameter, length, is the same tube as for the QCL?
The two instruments had their own inlet tubings. This is now clarified in the text and LI-6262 tubing information added: "*All measurements were recorded at 10 Hz frequency and were made with a flow rate of approximately 10 liters per minute (LPM) for the QCLS and 14 LPM for LI-6262, respectively. The PTFE sampling tubes were 32 m and 12 m long for QCLS and LI-6262, respectively, and both had an inner diameter of 4 mm. Two PTFE filters were used upstream of the QCLS inlet to prevent any contaminants entering the analyzer sample cell: one coarse filter (0.45 µm, Whatman), followed by a finer filter (0.2 µm, Pall corporation), at approximately 50 cm distance to the analyzer inlet. The Aerodyne QCLS used an electronic pressure control system to control the pressure fluctuations in the sampling cell. The QCLS was run at 20 Torr sampling cell pressure. An Edwards XDS35i scroll pump (Edwards, England, UK) was used to pump air through the sampling cell, while LI-6262 had flow control by a LI-670 flow control unit.*"

l. 106: is this the mean? what is the standard deviation?
The standard deviation was determined from the standard deviation of cylinder air measurements. Now clarified in the text: "*The standard deviation was 19 ppt for COS mixing ratios and 1.3 ppm for CO_2 at 10 Hz measurement frequency, as calculated from the cylinder measurements.*"

l. 111-112: given the precision of typical computer clocks, this will result in clock differences up to several seconds; important to add that most likely the Li-6262 data were acquired by and thus synchronized with the sonic anemometer through it's A/D input?!
Thanks for the comment. The logging system and data flow will be described in more detail in the revised manuscript.

l. 129-131: I am not sure I understand the first step – the QCL clock may be either delayed or advanced with respect to the clock of the sonic anemometer & IRGA and I thus do not understand why you shift the QCL time series by the lag time between w and IRGA CO2? In my understanding you could start off with the second step which actually aligns both time series.

The reviewer is correct, we have removed the first step. Sorry for the misunderstanding, now clarified in the text: "*The following procedure was done to combine two data files of 30 min length (of which one includes sonic anemometer and LI-6262 data and the other includes Aerodyne QCLS data): 1) the cross-covariance of the two CO_2 signals (QCLS and LI-6262) was calculated 2) the QCLS data were shifted so that the cross-covariance of the CO_2 signals was maximized. Note that this will result in having the same lag time for QCL and LI-6262. The time shift was a maximum of 10 seconds, most often varying between 0 s to 2 s during one day. It is also possible to shift the time series by maximizing the covariance of CO_2 and w, which will then already account for the lag time (Fig. S1) or combine files according to their time stamps and allow a longer window in which the lag time is searched. However, in this case it is important that the lag time (and time shift) is determined from CO_2 measurements only, as COS data might result in several covariance peaks in longer time frames due to low signal-to-noise ratio and small fluxes.*"

l. 131: might be worth showing this as a histogram in the supplement?
Time shift (computer drift + QCL lag) is now shown in the supplement Fig. S1.

l. 133-134: shouldn't this sentence come first in this section as this likely was the initial step? Or did the procedure described in l. 126-132 use data before despiking?
File combination was indeed done before despiking.

l. 142-143: not sure that computation time is a relevant issue nowadays with regard to coordinate rotation
Removed the note about computation times.

Fig. 1: the first rotation angle is typically the one that aligns the coordinate system along the main wind direction – why would that rotation be limited to less than 10 ◦ , which would mean rejecting fluxes from 340 ◦ ?
You're right, it was the second rotation angle, as mentioned in the text. Corrected in Fig. 1 now as well.

l. 200: in my memory, the first to propose this approach were Aubinet et al. (2000, 2001)
Already credited earlier, but added the reference here as well.

l. 201: a site-specific cospectral model was already used by Wohlfahrt et al. (2005)
In this line we only refer to Aubinet et al. (2000) as they introduced the technique and to De Ligne et al. (2010) as they compare different frequency response correction methods and make recommendations, but we will add Wohlfahrt et al. (2005) reference to earlier in the text.

l. 213: which rotation angle?
Second rotation angle, now clarified in the text.

l. 232: but fluxes are available half-hourly – how do you come up with a half-hourly storage change estimate?
The storage change estimate from the concentration profile measurements are hourly and from EC measurements half-hourly. Concentration change used in the calculation is the change that occurred during that 30 min (cafter – cbefore), hence we get half-hourly storage estimates from the

EC setup. The profile measurements were subsampled to 30 min to correct for the storage change in measured 30 min fluxes.

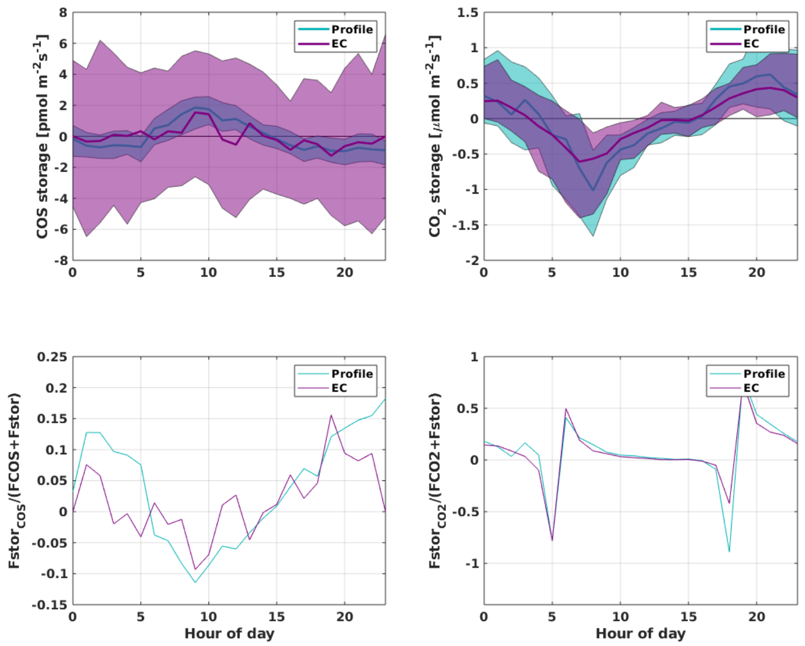
l. 237-238: 5 hours sounds like a really long time to reduce random noise – in fact I would expect a 5 hour moving average to even average out true storage; how exactly did you calculate the one-point storage term?

This is the same time window that was used in smoothing profile measurements (Kooijmans et al., 2017) and we chose the same time window for smoothing EC data for better comparison. If a shorter time window is used (e.g. 1 hour), the diurnal shape of the storage change flux does not change but noise of COS storage change flux increases almost three-fold (see figure below). One-point storage term was calculated by assuming that the concentration change in time is uniform from the measurement height to ground:

FCOSstor = h*ΔC/Δt

Where h is the measurement height (in m) and ΔC/Δt is the concentration change in time (in mol m-3 s-1) at the measurement height.

Below is an example of storage change flux when 1 hour moving average is used for smoothing concentration measurements. Median diurnal variation of COS storage change flux is not different from the 5 hour moving average, but variation has increased notably.



l. 264: clearly here only Lee (1998) is to credit with this approach

All text related to vertical advection will be removed from the revised manuscript.

l. 270: “missing CO2 fluxes”?

Corrected.

l. 282: add units

Units added.

l. 284-286: isn't this a repetition from above?
Yes, we will incorporate all the relevant information into the methods section.

l. 288-289: all measurements are characterised by noise to a certain degree . . .
Yes, we do agree and we have rephrased it as "*As COS measurements are often characterized by low signal-to-noise ratio, the maximization of the absolute value of cross-covariance may determine the lag time from a local maxima, as demonstrated in Fig. 2.*"

l. 289-290: I do see several local minima in Fig. 2, but not that any of the tested algorithms gets stuck in one of these
We noticed that there was an artefact related to lag time calculation, as the lag time results were obtained from a lag time optimization tool. In the revised manuscript, we are using the lag times calculated based on the covariance maximization.

l. 298-299: so what? How does that sentence relate to your results?
Thanks for the comment. The sentence will be rephrased accordingly.

l. 305: reformulate – I guess that what you mean is that with the DetLim method, the COS lag was selected in 40 % of all cases (while the CO2 lag was chosen in 60 %)
Reformulated as "*By using the DetLim lag method, the COS lag time was estimated for 40 % of cases from the* $\overline{w'\chi_{\cos}'}$ *covariance maximization, while the* CO_2 *lag was used as proxy for the COS lag in about 60 % of cases.*"

l. 305: can you compare lag times of COS and CO2 both fluxes are clearly higher than the flux detection limit? Is there a systematic difference between the two (e.g. COS lag time always longer than CO2) and if so could the DetLim Method be improved by adding this offset to the CO2 lag instead of just using the CO2 lag?
We tested this idea, but there was no systematic offset found between the lag times.

l. 309: why did you choose the cumulative flux as a metric? Wouldn't the cumulative flux potentially be affected by compensating effects (over- and underestimation during certain conditions resulting in similar cumulative fluxes)? Unlike for CO2, I also do not see much nee to calculate a daily or longer term budget for COS; I also do not see how you can get units of nmol/m2s for COS as for a cumulative flux you need to integrate over time, i.e. multiply the half-hourly flux by 1800 s and then sum these up this then yields molar units per m2 and time period over which the cumulative was calculated – same for CO2.
It is not a cumulative sum as used in budgets, thus not multiplied with 1800s and has units of nmol m-2s-1, i.e. just simply cumulative sum of all measured (NOT gap-filled!) fluxes. We have considered the reviewer's comment on compensating over- and underestimations in cumulative sum and chose to use median fluxes instead. Median fluxes are also affected by the compensating over- and underestimations but not as much to differences in missing data. We will add overall median fluxes as well as separate night-time and daytime median fluxes of all different processing schemes to Table 2.

Table 1: instead of repeating the values for the reference three times, list it only once?!
Reference fluxes will be added to the table caption and removed from the table.

l. 315: how do you know this difference is large on an annual scale? What is the basis for this recommendation?
Changed the sentence to "*This difference might become important in the annual scale, and we recommend using the detection limit method in lag time determination of small fluxes, as in Nemitz et al. (2018)*" for clarity.

l. 335-336: "low-frequency corrections"
Corrected as suggested.

l. 336-337: combine both sentences
Combined sentences.

l. 337-338: this is kind of trivial
While this recommendation may be stating the obvious, it is not always carried out in field practice and therefore is worth emphasizing. We decided to keep the sentence, as it adds to the discussion.

l. 344: what is the "normal" CO_2 cospectrum? Fig. 7a shows the COS cospectrum, 7b the COS power spectrum – this sentence does not make sense
Corrected the figure reference to 7a and 7c.

Fig. 7: are the (co)spectra in any way filtered for stability or really averaged over the entire month? I suggest to remove the sub-grid lines in all four panels as otherwise these are too busy and become blurry
All the spectra are filtered for stability. Added a sentence to the (co)spectra figure's caption "*All data are filtered for stabilities $-2 < z/L < -0.0625$ and COS data only accepted when the covariance was higher than three times the random error due to instrument noise (Eq. 1).*" Gridlines will be removed.

l. 349: in Fig. 7 you are using normalized frequency, thus no units
Changed the text from "3 s" to "*normalized frequency higher than 3*"

l. 350: while the attenuation is clearly visible for CO_2, it does not show well for COS, whose cospectrum mostly overlaps with the sensible heat cospectrum
Cospectrum is now plotted for only those times when COS flux surpasses the random noise. While there is still quite a lot of noise in the high frequency end of COS cospectrum, the attenuation is now more visible. Probably the instrument random noise was overlapping with sensible heat cospectrum by chance.

l. 350-351: how was this calculated?
Thanks for the comment. We will add more details in Section 2.4.3. In practice the response time τ_s was determined by fitting a sigmoidal function $1/(1+(2\pi f \tau_s)^2)$ to the ratio between ensemble averaged CO_2 and T cospectra.

l. 353: indeed this is as expected . . . remove?
Removed the sentence.

l. 355-357: I think it would be instructive to show at least one characteristic example comparing the experimental and analytical frequency response correction approaches as otherwise this remains a “black box” for the reader

The two correction methods are now fully compared by adding median daytime and night-time fluxes to Table 2, histogram of fluxes to Supplementary material Fig. S2 and stability categorized flux attenuation versus wind speed to Supplementary material Fig. S3. Otherwise, we do not understand what the referee means by saying “to show at least one characteristic example”.

l. 358: while I understand that the experimental frequency response correction approach is part of the standard against which the comparison is made, I think (i) that the no frequency response correction scenario is useless as we know that this leads to a bias and (ii) instead it would be useful to compare the magnitude of the correction between the analytical and experimental approach in order to understand how much of a difference it makes whether one or the other correction is used

The “no correction” option demonstrates how much high frequency corrections affect the final fluxes and is thus left to the analysis. Magnitude of the correction is demonstrated in Table 2 as total median fluxes and daytime and night-time median fluxes. Histogram of the resulting COS fluxes is added to Supplementary material Fig. S2 and flux attenuation versus wind speed to Fig. S3. Final fluxes are also compared as a scatter plot in Fig. S7.

l. 364-369: this section comes a bit as a surprise as it never has been mentioned as a goal before to do this kind of comparison and also lacks a proper discussion – without discussion I rather suggest to remove Fig. 8 and the corresponding text; one point of discussion might be how much the difference in the cospectral reference models contributes to the differences between the analytical and experimental approach; btw., the results in Fig. 8 are qualitatively consistent with Fig. 11 of Wohlfahrt et al. (2005)

As this was not a very important part of our analysis, we have now moved the figure to supplementary material (Fig. S4), as suggested, and removed the corresponding text. Equations 15 and 16 of the former version were moved to Methods-section and are now presented as equations 6 and 7 in the present version of the manuscript.

l. 371-377: here I have the impression that you not describing the actual magnitude of the correction, but rather what we might expect based on one example shown in Fig. 5?!

This correction is based on theoretical transfer functions by Rannik & Vesala (1999), and the range of magnitude of this correction is maximum 15 %. The correction is not related to setup or site specific, it is a general and theoretical correction, and expected to be similar to Rannik & Vesala (1999) study. It does not make sense to only correct for low frequency loss in this study to check the actual magnitude of the correction. We will remove this section from the revised manuscript and move Fig. 5 to supplementary material.

l. 372-374: this expectation would only be justified if the algorithm used for correcting for flux loss at lower frequencies would “know” of the noise, which I expect it does not

Exactly, it does not know of the noise and that’s why cannot make large corrections. We have decided to leave out the result section on low frequency spectral corrections.

Figure 9: to what period do the data shown refer to?

To the whole measurement period. Clarified in the figure caption "*...during the measurement period 26 June to 2 November 2015.*"

l. 400-401: again, this is well established

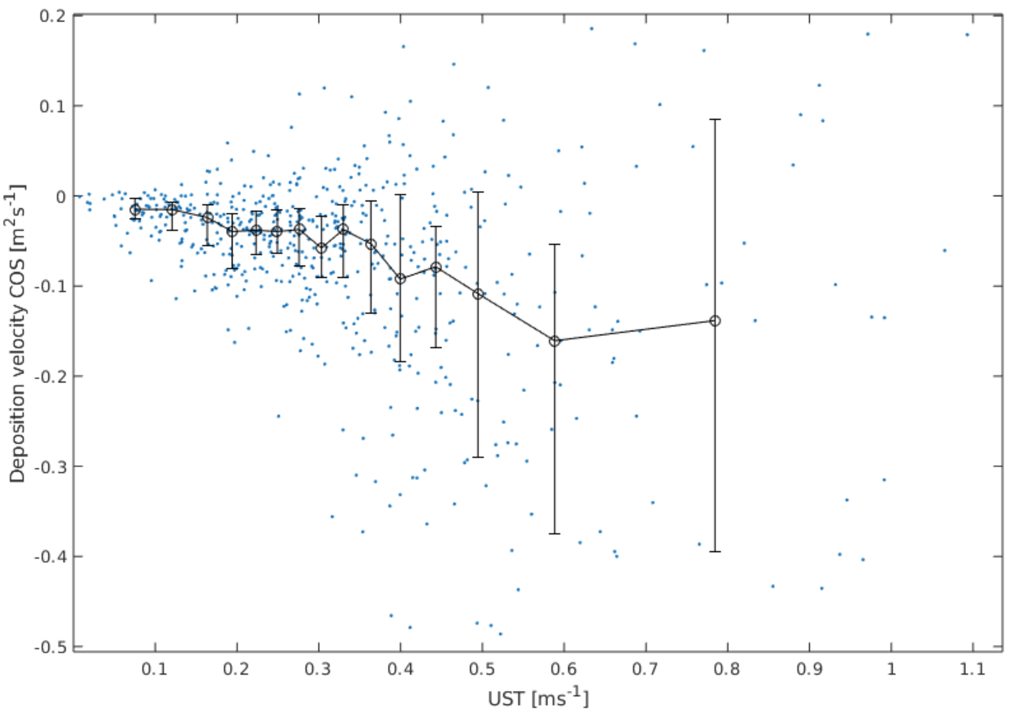
We agree with the Reviewer, yet storage change flux is neglected in most of COS EC studies (Table 1). Thus, we feel it is important to emphasize that for diurnal variation it cannot often be neglected. We revised the sentence to emphasize this point: "*In conclusion, the storage change fluxes are not relevant for budget calculations – as expected – and have not been widely applied in previous COS studies (Table 1). However, storage change fluxes are important in the diurnal scale to account for the delayed capture of fluxes by the EC system under low turbulence conditions.*"

l. 418: I would not call data erroneous – they simply do not fulfil the assumptions underlying our simplified model of surface-atmosphere exchange

Modified the sentence to "*...reliable tool to filter out data that is not representative of the surface-atmosphere exchange under low turbulence (Aubinet et al., 2010).*"

l. 421-423: with this reasoning, wouldn't it make sense to get rid of the concentration dependency by using the deposition velocity, i.e. the flux normalized with the concentration, instead of the flux itself? Or perform the analysis with data stratified by COS concentration to minimize the issue?

Thank you for this suggestion! We made the u* plot using the deposition velocity (EC flux + storage change flux normalized with concentration gradient), but the u* dependency did not disappear. We added the following text to the manuscript: "*However, we did not see u_* dependency disappearing even with a concentration gradient-normalized flux, so the u_* filtering is applied here normally to overcome the EC measurement limitations under low turbulence conditions.*"



l. 428-437: knowing that the u*-filtering needs to be applied on the sum of the storage and EC flux, why do you still give the numbers for the EC flux without storage?

There are quite many COS studies that use *u** filtering even though they neglect (or don't mention) the storage change flux: Asaf et al. 2013, Billesbach et al. 2014, Maseyk et al. 2014, Commane et al. 2015, Yang et al. 2018 (and not sure how it was done in Spielmann et al. 2019, as it is not mentioned). So even though this is somewhat trivial for EC measurements, it has not been implemented widely in COS studies, and is worth mentioning here.

l. 434-435: circular argument - without applying the storage correction it is pretty clear that the storage flux is clear that the storage flux is ignored
Reformatted the sentence to "*If fluxes are not corrected for storage before deriving the u* threshold, there is a risk of flux overestimation due to double accounting. The flux data filtered for low turbulence would be gap-filled, thereby accounting for storage by the canopy, but then accounted for again when the storage is released and measured by the EC system during the flushing hours in the morning (Papale et al. 2006).*"

Figure 11: also shows the vertical advection without being mentioned in the figure legend?!
Vertical advection was mentioned in the figure legend, but was missing from the caption. We have decided to leave out the section regarding vertical advection, according to the reviewer's suggestions, and thus removed also from this figure.

l. 442-444: is this a useful comparison? The cumulative COS flux must be less negative if missing data (which are generally negative) are not gap-filled! In order to evaluate the skill of the gap-filling algorithm the authors need to create artificial (but realistic) gaps in their time series (see CO2 flux literature to that end) and then compare the gap-filled against the measured (during the artificially created gaps) fluxes!
This comparison is just demonstrating how much change there is when using gap-filled versus non gap-filled fluxes in a cumulative sum, and the magnitude of course depends on the number of gaps in the data. We will add a diurnal plot of the gap-filling algorithm and measured fluxes (Fig. 11), a plot of the residuals of different gap-filling algorithms (Fig. S6) and a time series of the gap-filling parameters (Fig. S5) to the revised manuscript and supplementary material.

l. 474-475: why do you recommend the experimental high-frequency correction approach? Is it more accurate? If so this needs to be demonstrated! What about the performance of the experimental approach in situations with low/noisy sensible heat cospectra and what about the effect of the QCL on the ratio central to the experimental approach?
We did not use single 30 min T cospectra for the model cospectra. Only the averaged cospectra (when the cospectra was good) were used for creating the model (see Sabbatini et al., 2018 for reference). A site specific cospectral model is suggested to be used instead of analytical ones in several studies (Aubinet et al. 2000, De Ligne et al. 2010), and especially in the new eddy covariance data processing protocols (Sabbatini et al., 2018; Nemitz et al., 2018). We will provide both Horst (1997) and experimental cospectral models in the cospectrum figure, together with the mean cospectrum.