We thank each of the anonymous referees for their considered evaluation of our manuscript. We are pleased that they recognise both the novelty of our approach and the important implications this research has for the community. As the reviewers point out, the manuscript is quite technical in nature and could benefit from specific formulations and definitions. With these comments in mind we have revised the manuscript to clarify the areas highlighted by the reviewers, adding numerical formulations where appropriate.

Reviewers 1 and 2 both advised removing Figures 3 and 10. In our opinion these two figures are quite crucial to the study. Figure 3 clearly demonstrates that no systematic bias is introduced from instrument noise regardless of its distribution. As well as justifying our choice of using Gaussian white noise in our error method, this result is something that, to the best of our knowledge, has not been shown before and is therefore too relevant to hide away in the Supplementary Information. We have revised the discussion of this figure to better emphasise its significance to our study.

In our opinion Figure 10 perfectly summarises the effects of mirroring, the bias it introduces and how these effects can be enhanced or minimised through the choice of time-lag determination. Again, we have revised our discussion of this figure to better reflect its importance. Interestingly, when the work was presented at the 2015 EGU conference earlier this year, this figure drew the most discussion and we are therefore reluctant to confine it to the Supplementary Information.

The comments of each reviewer (RC) are individually addressed below using the suffix (AR).

Anonymous Referee #1

RC: line 270/271: What is the reasoning behind your statement "... the standard deviation approach derives significantly smaller limits of detection, which we believe to be underestimates of the true uncertainty"? What makes you believe that these are underestimates?

AR: As shown in Fig. 2B, the SD approach (shown as a dashed blue line) is approximately half that of the RMSE approach and consequently the majority of points in the cross-covariance exceed the LoD threshold. In theory, the LoD should threshold above the general noise of the cross-covariance which has clearly not been achieved in this example. It should be noted that this only occurs in instances where the cross-covariance is predominately of one sign. In Fig 2A, where the cross-covariance switches between positive and negative fluxes, the SD and RMSE methods give the same LoD.

This sentenced was rephrased to make this point more clear.
"The two approaches agree very closely for periods where the cross-covariance fluctuates regularly around zero as in Fig. 2A, but where the covariance is predominantly of one sign the LoD$_\sigma$ approach derives significantly smaller limits of detection, which, as shown in Fig 2B, do not cross the threshold of the general noise of the cross-covariance function."

RC: line 324 and following lines: The passage about the sensitivity of the covariance and the distribution of the selected noise should be shortened and Figure 3 moved to the Supplemental Material. The reader should be informed about the main findings (similar frequency distributions and mean average flux close to zero) in the text. However, I was distracted by trying to understand Figure 3, which finally did not add much information. Also, the dashed lines in Figure 3B are not fully visible.

AR: We respectfully disagree with the reviewer on this point, and firmly believe this section should remain within the main manuscript. As this paper covers a range of analysers, some affected by Poisson noise and others by Gaussian, it is essential to demonstrate that our technique (in which we use Gaussian white noise) is valid regardless of the type of noise used. To the best of our knowledge such an assessment has never been made in the literature before and therefore this is a very important element of the paper. We have revised the text as follows:

"With this in mind we performed several tests to determine if the covariance between the vertical wind velocity and a time series of white noise differs depending on the distribution of that noise (e.g. whether it is Gaussian, Poisson or log-normally distributed). For a single 30-minute averaging period the covariance between w' and ε' was calculated iteratively, whereby the artificially generated noise (ε') was either Gaussian, Poisson or log normally distributed as seen in Fig. 3A. For each of the 5000 iterations a new time series of white noise was generated and the covariance was re-calculated using a prescribed time-lag. Figure 3B shows a distribution of the resulting white noise fluxes calculated using either Gaussian, Poisson or Log-normally distributed noise. In each case, the fluxes are evenly spread about zero, which demonstrates that unstructured white noise creates a random uncertainty in the flux but does not induce a systematic bias, regardless of its distribution. In our calculation of the random instrumental flux error we have chosen to generate white noise with a Gaussian distribution. This simple test clearly demonstrates that our choice of noise distribution is actually unimportant and that the same result will be obtained regardless of the distribution of the noise. Importantly, these findings provide assurances for eddy covariance systems that induce a Poisson counting noise that flux biases are not created."

RC: I suggest removing Figure 7 from the manuscript and adding one example of the disjunct eddy covariance exercise to Figure 6 in direct comparison with the standard eddy covariance result.

AR: Following the reviewer's suggestion we have merged figures 6 and 7 to include the EC example and two DEC examples (Δt 2.5 s and Δt 7.5 s). The omitted panel will be added to the Supplementary Material.

RC: In Figure 8 (Section 3.2.2) you discuss the influence of the peak width of the cross-covariance function on the relative flux error of artificially generated multi-frequency signals. Could you expand a little bit on the discussion how the findings from these simulations can be transferred to real world data? Can the observed increase of the relative flux error with increasing FWHM be generalized, or might this be a peculiarity of the frequency distribution of the artificial signals? With respect to the caption of Figure 8, it suggests to me that the relationship between the average relative bias and the FWHM of the cross-covariance function peak is exponential. Is this really the case?

AR: In simple terms, when using the MAX or AVG methods, the broader your cross-covariance peak the greater the probability of you selecting a false maximum (i.e. the true covariance + maximum random noise). This is an important consideration and one that can be generalised to real world data, but not universally be quantified. The frequency distributions of the underlying signals are unimportant as we are here only interested in the properties of the cross-covariance peak (e.g. the width).

"Broader covariance peaks, reflecting slower turbulence/larger eddies, result in a higher probability of an extreme maximum (i.e. the true cross-covariance plus random noise: $(\overline{w'c'}) + (\overline{w'\varepsilon'})$) being chosen and therefore both the probability of overestimating the flux and the magnitude of the bias are closely linked to the peak width."

The relationship between the relative bias and the FWHM of the cross-covariance function peak is linear. We have amended the figure caption to reflect this.

RC: line 476 and following lines "The logarithmic relationship between turbulence and height ...": Please explain your line of reasoning step by step. It is not obvious how this passage relates to the sentences before it. Also, the last sentence of this passage ("Nonetheless, during more unstable periods...") needs more explanation and integration with the rest of this passage.

AR: Here we were making the point that the cross-covariance peak width, in part, relates to the mean eddy size (i.e. larger eddies result in broader peak widths). Mean eddy size scales logarithmically with height, therefore measuring high above a forest or city, cross-covariance peaks will be broader than when measuring close to the ground. To better reflect this we have removed the line beginning "The logarithmic relationship…" and replaced it with:

"Broader covariance peaks, reflecting slower turbulence/larger eddies, result in a higher probability of an extreme maximum (i.e. the true cross-covariance plus random noise: $\overline{(w'c')} + \overline{(w'\varepsilon')}$) being chosen and therefore both the probability of overestimating the flux and the magnitude of the bias are closely linked to the peak width. The increase of mean eddy size with height means trace gas and aerosol flux measurements at higher elevations above ground are more at risk to systematic bias when the MAX or AVG methods are employed."

In addition we have removed the line *"Nonetheless, during more unstable periods the potential for a greater influence of lower frequencies in the turbulence spectra cannot be overlooked (Horst, 2000)."*

RC: Section 3.3.1: I suggest moving Figure 10 to the Supplemental Material and to focus on Figure 9 in this section. Line 528 may be changed to "... can result in very unnatural bimodal flux distributions (see Supplemental Information)."

AR: We believe figure 10 is essential as it very neatly summarises the bias introduced by the differing time-lag methods. We have clarified that for most compounds this type of plot can actually be used to assess whether mirroring is occurring, which addresses a question of Referee 3 below. As well as clearly demonstrating the "mirroring" effect (i.e. by showing the bi-modal distribution centred around zero), by overlaying the mean flux values the induced bias is very effectively highlighted. We have revised the text to better reflect these points:

"Adopting the MAX or AVG methods exagerates the mirroring by systematically choosing the furthest point away from zero which in the extreme case can result in the very unatural flux distributions shown in Fig. 9. Adopting the AVG method with 5 s running mean limits this effect to a certain extent, but a noticible dip around zero remains. Importantly, the use of a prescribed time-lag eliminates the splitting of data from either side of zero to give a much more natural looking flux distribution. For many compounds an assessment of the frequency distribution of flux data evaluated with the MAX method will highlight whether mirroring occurs and whether this approach is therefore not applicable. Care needs to be taken, however, when making this type of assessment on CO_2 fluxes, as the interplay between strong positive fluxes during night and negative fluxes during day could potentially result in a similar bi-modal distribution. Figure 9 also nicely illustrates the flux bias introduced by using time-lag methods that systematically search for a maximum in the cross-covariance. In this instance the MAX method gives a mean flux 2.3 times larger than the PRES method."

RC: In addition, I cannot follow the discussion of the Gaussian white noise flux data with respect to acetone and benzene. To me, the range of acetone and benzene fluxes in relation to the Gaussian white noise fluxes look very similar.

AR: We have revised the text here to pick out the specific periods in the benzene time series to which we refer:

"In contrast, the range of benzene and particle number fluxes both at least partially exceeds the Gaussian white noise flux and show sustained period of emission fluxes (e.g. 17th to 20th of September) indicating the presence of a "genuine" flux which is, for certain periods, distinguishable from the random sensor noise flux."

RC: While Figure 12 is very illustrative, I found the discussion of down weighting of data points in the figure caption distracting. If there is no further discussion of this fit, I suggest removing these pieces of information altogether. Otherwise, the authors may add a brief discussion of the fit in the main text.

AR: We agree with the reviewer and have removed this discussion from the figure caption. We have also revised the figure to show the limit of detection as error bars. Points at which the error bars intersect the zero line are not statistically different from zero and should be down weighted accordingly in any fitting. We have added the following to our discussion of that figure:

"In this example, the ensemble LoD is shown as individual error bars and thus where those bars intersect the zero line the flux is not significantly different from zero. Rather than eliminating these data from the fit shown in Fig. 11 which might bias the fit, our recommendation is to weight the fit parameters according to the LoD, thus down weighting these points."

Technical comments:

RC: lines 55 and 305: reference should read "Mauder et al. (2013)"

AR: This was changed

RC: line 98: usage of "and or" seems awkward to me

AR: This was changed to "and"

RC: line 135: reference should read "Mahrt, 1998"

AR: This was changed

RC: lines 223 and following lines: I recommend moving the passage starting in line 232 "It should be noted that..." until line 236 "... would still be affected by such noise." To the end of line 222 before you start with the discussion of alternatives in the frequency domain.

AR: We preferred to keep the current structure. We have corrected the sentence following this section in response to Referee 2 and suspect that this has improved the readability anyway.

RC: line 244: add "Wienhold et al.," before "1995"

AR: This was added

RC: line 260: remove "(LoD"

AR: This was removed

RC: lines 302 and 304: "epsilon" should be subscript

AR: This was corrected

RC: line 335: reference should read "Lenschow and Kristensen (1985)"

AR: This was corrected

RC: lines 373 and 417: "e.g." seems to be out of place

AR: "e.g." removed on line 373 and replaced with "(Eq. 3)" on line 417

RC: line 473: remove "maximum" after "FWHM"

AR: This was removed

RC: line 477: "means" instead of "mean"

AR: This was corrected

RC: line 529: should read "to a certain extent"

AR: This was corrected

RC: line 570: for the acetone time series, refer specifically to "Fig. 9C"

AR: This change was made

RC: line 595: replace "we have carefully examined the key factors" with "we have carefully examined several key factors"

AR: This was changed accordingly

RC: line 633: Do you mean "adsorption" instead of "absorption"?

AR: Yes, this was corrected

RC: line 852, Figure 1: Why is the double-arrow, which indicates "Noise" in 1A going beyond the circle indicating AC(0)? According to lines 208/209, an estimate of random instrument noise is given by the difference between AC(0) and AC(1).

AR: This is the effect of resizing the figure. The arrow has been resized to span between AC(0) and AC(1).

RC: line 854: should read "through"

AR: This was corrected

RC: line 881, Figure 5: The units of the isoprene and acetone fluxes (top panels B and C) and the random instrument noise (lower panels B and C) should be the same for better comparison. What is the advantage of plotting the error contributions of instrument noise and variability in surface plots? I'd prefer a simple line plot.

AR: Agreed. The error units will be changed to match the flux units. We feel the stacked area plots nicely demonstrate the partitioning of the total random error between instrument and random variability components.

RC: line 885: add "into" between "can be divided" and "errors"

AR: This change was made

RC: Figures 6, 7 and 10: For consistency with other figures and the manuscript text, use "PRES" instead of "FIXED" in the figure legends.

AR: These changes were made

RC: line 913, Figure 9D: The time axis of the lowest panel must be revised.

AR: This has been revised

*   *   *

Anonymous Referee #2

Specific comments:

RC: Section 1.2: Equation 1 and 2, the symbol "RE" is not explicitly defined (although it is implicit in the text that this represents "random error").

AR: We have now changed the first sentence of Section 1.2 to read: "*Assuming the time-lag is known, the random error (RE) of an eddy covariance flux*"

RC: Section 2.1: The authors state their focus is on unstructured white noise only, and here offer a definition of SNR (Equation 4). However, throughout the paper there is also discussion of "total" random error which includes structured noise. I would find it useful to see short comment on how the authors expect structured noise to affect the calculation/definition of SNR, and what the authors expect to see in the cross-covariance results and/or in the cospectra when structured noise is present. Furthermore, when is signal-to-noise in the manuscript determined from total random error, and not just instrumental white noise, if ever?

AR: Throughout the manuscript the signal-to-noise ratio of the eddy covariance data shown is always determined on the basis of the instrument white noise. Although we are unable to quantify an instrument error specifically for structured noise, it still contributes to the total random error as quantified by the method of Wienhold (1995) (Eq. 6 in the revised manuscript) and especially in our modification (see Eq. 7 in the revised manuscript). Our inability to isolate the structured noise means that the signal-to-noise ratios reported are often an underestimate of the "true" signal-to-noise ratio and that our estimates of the random error due to variability is likely somewhat overestimated (i.e. the partitioning between instrument and turbulence statistics may be biased towards the latter). Nonetheless, the estimate of total random error is unaffected and it is this parameter that is used for LoD calculations.

We have amended the manuscript to make this point clear both within Section 2.3 and also in the conclusions:

"As mentioned previously, the auto-covariance technique, also used by Mauder et al. (2013), is not sensitive to the effects of structured instrument noise. Consequently, the instrumental random error reported by both methods is likely underestimated. Nonetheless, the contribution from structured noise is included in our estimate of the total random error and it is this parameter that is used to define the flux limit of detection."

Conclusions:

"Both these methods can be easily implemented into eddy-covariance processing software and share the key advantage over the more traditional experimental approach of Shurpali et al. (1993) that measurements do not need to be interrupted for the assessment to take place. Nonetheless, it is important to reiterate that both of these approaches are not sensitive to the effects of structured noise, which cannot be quantified using the auto-covariance method (e.g. Eq. 5) upon which both of these approaches are based."

RC: p.2922, lines 5-7: The authors state that it is necessary to extrapolate the autocovariance function back to the zero point, "typically using the first few points". According to whom? Is this specific to flux applications, or is this used generally by others to establish any instrumental noise?

AR: This method is not specific to flux measurements and can be applied to any analyser to establish the contribution of white noise. We have modified this section slightly to include references to other studies where this method has been used.

"The auto-covariance should decrease following a 2/3 power law (Wulfmeyer et al., 2010), but we find a linear extrapolation to be more appropriate. Lenschow et al. (2000) and Mauder et al. (2013) also adopted a linear extrapolation and suggest the deviation from the 2/3 power law to result from the averaging effects of the analysers."

RC: p.2923, lines 1: The authors state that the departure from the -5/3 slope at high frequencies may not show up in the auto-covariance approach. Do they mean in the frequency domain approach (Fig. 1b)? Or is the auto-covariance approach (Fig 1a) the same as the frequency domain approach (Fig 1b)? If so, I am confused at the distinction made in this paragraph (since the frequency domain approach is presented as an alternative to the auto-covariance approach).

AR: Yes, we actually meant the frequency domain approach and have changed this accordingly in the revised manuscript. Thank you for pointing out our mistake.

RC: Section 2.2: I believe this is where the authors first define a limit of detection "LoD" (multiplying the SD of fsub_w'c'(tau) by 3). However, if I understand correctly, the authors settle on using a LoD based on the RMSE approach for the rest of their manuscript, which is slightly different from the SD of fw'c'(tau) approach. The definition and calculation of this "new" or "alternative" LoD (used throughout the rest of the manuscript) is not really explicitly defined nor formulated for the readers (including, if necessary, an equation for the RMSE deviation of fw'c' from zero). Later, the "total random error" is defined as 3XRMSE – but I didn't find RMSE to be adequately formulated itself.

AR: We agree with the reviewer here (and also with reviewer 3) that we do not give an adequate formulation of the RMSE approach. In the revised manuscript we now give a numerical formulation for both the sigma and RMSE approaches which can be seen in our response to reviewer 3.

RC: p.2927, line 18: Is the -1 superscript in "(SNR-1)" a typo?

AR: This has been removed

RC: Section 3.1: Here the authors explicitly define "total random error" (3xRMS) and random instrument error (REnoise). When the authors later on refer to flux data points potentially being rejected for being "below conventional limits of detection", is this always referring to the 3xRMS? Are they ever referring to exceeding 3XREnoise (as determined by the Gaussian simulated data) alone? The later discussions of Figure 9 and 10 (e.g. Section 3.3.1) made me think they might only be considering the Fsub_GN (Gaussian simulated flux data) when determining whether a data point might be rejected.

AR: Limits of detection are based on the total random error throughout the manuscript. However, for compounds with very poor signal-to-noise ratios, the random error is dominated by the instrument noise. The purpose of figure 9 is therefore to illustrate that what might be thought of as a flux can in fact be replicated solely by calculating a flux from a time series of Gaussian white noise.

RC: Section 3.2.1: The "signal-to-noise ratio of the analyzer" Does this refer back to Equation 4? I wondered if "signal to noise" here refers to the measured FLUX signal over the determined flux noise – but to me this is different than instrumental signal (i.e. mixing ratio measurement) and instrumental noise. Could the authors clarify?

AR: Yes, it is the flux signal vs the contribution of the instrument noise to the flux noise. For flux measurements, the signal is contained in the variations, for concentration measurements it is contained in the mean. The signal to noise ratio in the manuscript always refers to the time series of mixing ratios from the analyser which is inherently linked to the noise in the cross-covariance function.
For clarity we have revised that sentence to read:

"The simulations applied to sensible heat flux data reveal a distinct relationship between the signal-to-noise ratio of the raw temperature data and the relative flux bias for both the MAX and AVG lag determination methods."

We have also added a clarification to Section 1.1:
"The meaning of SNR is different for flux measurements than for concentration measurements. When measuring concentrations, the information (or "signal") is associated with the mean, while for flux measurements it is associated with the variability in the time-series that reflects fluctuations induced by turbulence rather than by instrument noise. It is in the spirit of this letter definition that SNR is used throughout this paper."

RC: p. 2930, line 14: I believe the "AVE" should be "AVG"

AR: This was changed

RC: p. 2934, line 7: after "the red time traces" please insert "in Figure 9". Moreover, here is where the authors refer to some data points exceeding or not exceeding the Gaussian white noise flux (or, "random sensor noise flux") – but shouldn't the

determination of a true limit of detection be based on the "total random error" not just the Gaussian white noise flux?

AR: Yes, the limit of detection should always be based on the total random error. As discussed above, this figure is used to illustrate that fluxes calculated from noisy data using the MAX method can in fact be replicated by calculating a flux from a time series of Gaussian white noise.

RC: Section 3.3.2: The shaded areas in Figure 11 are said to represent the average LoD. Here again, I am wondering whether this is based on the "total random error" calculation (3 x RMS?) or simply the white noise/Gaussian flux/random sensor noise flux (3XREnoise?).

AR: Following Reviewer 3's comments we now provide a numerical definition of the LoD and make it clear that this is always based on the total random error. In this figure the shaded areas represent the ensemble LoD at the 95th percentile (e.g. α = 1.96). This is clearly stated in the figure caption

Figures:

RC: In some figures, the "prescribed" approach to lag determination is abbreviated as "PRES" and in others it is abbreviated as "FIXED". Please be consistent.

AR: All labels have been changed to "PRES"

RC: In the annotations that report a best-fit line, I am not convinced that 5 significant digits are necessary.

AR: All annotations have been reduced to 3 significant figures

RC: In general, I encourage the authors and editors to make sure that all the annotations in Figure 5, 9, and 11 are legible in the final version of the manuscript – They are currently very small, but it is not clear how legible they might be once finally implemented.

AR: Figures 5, 9 and 11 have been revised accordingly

RC: Figure 3: I agree that Figure 3 and its discussion in the text should be moved to the Supplemental Material, and shortened respectively.

AR: Please see our response to Reviewer #1

RC: Figure 8: Personally I thought this figure was unnecessary, too. The discussion is pretty clear without it.

AR: I believe you referred to Figure 10 here? This was retained (as Fig. 9) for the reasons outlined in response to Referee 1.

*   *   *

<u>Anonymous Referee #3</u>

We thank reviewer 3 for their evaluation of our manuscript. Some of the comments made by this reviewer indicate that they are perhaps less familiar with some of the concepts we discuss relating to the application and use of the cross-covariance and the terminology used. Given that this is a quite technical paper, we would expect most of the readers to be very familiar with the terminology in this field. Whilst we have made an attempt to simplify the wording and introduce concepts as much as possible in response to the referee, the technical nature of the manuscript necessarily remains. We note that Referees #1 and #2 did not raise the same concerns.

RC: There is no formulation of the LoD…please provide the exact formulation for each..

AC: For clarity we now include a numerical formulation of the LoD as calculated using the sigma and RMSE methods:

$$LoD_\sigma = \alpha\,\sigma_{fw\prime c\prime(-\tau,+\tau)}$$

$$LoD_{RMS} = \alpha\sqrt{\left(0.5\left(\sigma_{fw\prime c\prime(-\tau)} + f\overline{w'c'}_{-\tau} + \sigma_{fw\prime c\prime(+\tau)} + f\overline{w'c'}_{+\tau}\right)\right)}$$

Where α is a scaling parameter to give the LoD for a given confidence interval (e.g. 95th percentile $\alpha = 1.96$; 99th percentile $\alpha = 3$), $\sigma_{fw\prime c\prime}$ represents the standard deviation of the cross-covariance function within a defined window, and $f\overline{w'c'}$ is the average of the cross-covariance within a defined time window. The time window (τ) is a user defined parameter typically several times the integral timescale. Here, a negative value of τ represent a time window applied to the left hand tail of the cross-covariance function (e.g. negative time-lags) and a positive value represents a time window applied to the right hand tail of the cross-covariance function (e.g. positive time-lags).

RC: What exactly is $\sigma^2_{Fc.sub}$
AR: This is the variance of the sub-records.

RC: Can you provide a formulation to the MAX, AVG and PRES lag approaches?
We do not believe that numerical formulations of these time-lag methods is necessary. The PRES, MAX and AVG time-lag determination methods are very commonly used throughout the eddy covariance community. We are yet to see these formulated in the literature and believe a written description is adequate. Further, our description of these methods are consistent with those given by Taipale et al. (2010).

RC: Can you provide an objective numerical criterion for the determining that mirroring is occurring and can that criterion be formulated and not only described in word?

AR: Mirroring will occur when the flux is < LoD and the time-lag is determined by searching for a maximum.

E.g. $Fc_{noise} > \alpha\sqrt{\left(0.5\left(\sigma_{fw\prime c\prime(-\tau)} + f\overline{w'c'}_{-\tau} + \sigma_{fw\prime c\prime(+\tau)} + f\overline{w'c'}_{+\tau}\right)\right)}$

The common formulation used for the LoD is scaled by α which sets the LoD at the 95th (α = 1.96) or 99th (α = 3) percentile. However, when applying the MAX method you are by definition looking for the maximum value, which ultimately may lie outside of these confidence intervals. This also explains why the effect is minimised when using the AVG or PRES lag methods, as you are not always searching for the extreme value in these cases. Clearly, α could be

increased to eliminate mirroring, but our focus here is on maximising the information that can be obtained from eddy covariance data with low signal-to-noise ratios.
We have added a concluding sentence to Section 4 to state this.

In addition, as mentioned in response to Referee 1, for most compounds the assessment of the flux frequency distribution (Fig. 9 in the revised manuscript, previously Fig. 10) provides information on whether mirroring is occurring and we have, especially in the light of this comment, decided to leave this figure in the main body of the text.

RC: What is the difference between the cross-covariance and the just covariance?

AR: The covariance is a single value derived from two time-series. The cross-covariance is a series of values which are derived as the covariance of two time series which are shifted relative to each other. For each time shift, a new covariance is calculated.

The manuscript was revised to define this process more clearly.

"Correcting phase shifts between w and c is a key step in the calculation of fluxes and is routinely done by assessing the cross-covariance function (i.e. the covariance as a function of time-lag) between c and w which should reveal a maximum (in absolute terms), when the data are fully synchronised."

RC: What is the difference between the auto-covariance and auto-correlation functions? Can you formulate the autocorrelation function to make that clear?

AR: An auto-correlation function quantifies how well correlated a signal is with itself, with values ranging between 0 (uncorrelated) and 1 (perfectly correlated). For our purposes the auto-correlation is not helpful as it is dimensionless. The auto-covariance function looks at the covariance between a signal and itself. Using this approach as opposed to the auto-correlation enables us to express the signal and white noise components of a time series as a standard deviation.

We have amended the text to make this distinction more clear.

"Using the auto-covariance function as opposed to the auto-correlation function (i.e. the normalised auto-covariance) means the calculated signal and noise are variances and retain their original units. Taking the square root gives the standard deviation of both the signal and noise components."

RC: Lines 220-222 make no sense to me
AR: Our response to your previous point should make this now clear.

RC: What do you mean by "visually, the area becomes proportional to the noise variance…"?
AR: We chose to show this figure as a log-log plot as it allows us to clearly demonstrate how the signal and noise portions of the time series can be separated. The disadvantage of plotting the data this way is that "visually" it might appear that the areas attributed to signal and noise are of similar magnitude. However, if shown on a log-linear plot it would be evident that the noise portion of the time series is many times larger than the signal.

RC: What about the middle range – that the correlation is a result of small (turbulence) organised structures at a timescale much smaller than the averaging time that provide

the mixing with the land surface. Effectively the turbulence flux. Why is that not possible?

AR: The correlation between turbulence and concentration at these intermediate time scales should disappear over a Lagrangian time scale (by definition).

RC: Section 2.4.2 - you peak width simulation uses a signal with decaying frequency over time. This is not a fair representation of the real-world turbulence flux signal, as we assume that the turbulence signal is ergodic within the averaging period (typically 30 min) and thus, by definition, has a constant frequency profile over time.

AR: Our goal here was not to replicate real world turbulence. Rather, our aim was to establish if the peak width of a cross-covariance function affects the relative bias when using either MAX or AVG time-lag methods. In order to address this question it was necessary to (i) generate signals for which we could control the signal-to-noise ratio and (ii) apply a cross-covariance analysis to those signals to yield a peak of a pre-determined width. We elected to use multi-frequency signals because we could (i) deteriorate them through the addition of white noise to match a given SNR and (ii) alter the initial frequency to directly control the cross-covariance peak width e.g. lower initial frequencies yield a broader covariance peak width. To be clear, the characteristics which we are interested in relate solely to the width of the cross-covariance function and not the underlying data. The approach was successful in generating cross-covariance functions with varying peak width, which was all we were interested in.

For completeness, the multi-frequency signals (y) used took the form of a sine wave whose frequency varied from an initial frequency (f_0) to a target frequency (f_1) over a given time period (T)

$$y = A \times \sin\big(2\pi(f(t)t)\big)$$

where A is the amplitude, t is the current simulation time and f is equal to:

$$f = f_0 + \frac{(f_1 - f_0)}{2T}t$$

RC: L378-383 – I totally lost you here. Can you provide the formulation for this, so I could trace what you did?

AR: Please see our response to your previous comment.

Minor comments:

RC:L244 1995 - Author name missing

AR: This was added

RC:L250-252 the + or - sign at the second of each pair of numbers is missing. I am not sure if you mean -150 to -180 or -150 to +180. And, I am also confused about this time range, are these the limits of the lag-shift, the potential ranges of the averaging time,...? Having well defined formulation that is explicit about all these things throughout the manuscript will help.

AR: Following your earlier comment we now include a numerical formulation of this method. Figure 2 shows a cross-covariance function between vertical wind velocity and temperature. The cross-covariance shows the sensible heat flux as the two time series are shifted against each other between -200 s to + 200 s. The times you refer to (-150 s to -180 s and +150 s to +180 s) are the period over which the LoD is calculated.

RC:L260 - you open more parentheses than you close.

AR: This has been corrected.

RC:L390 what Gill model did you use?

AR: This was changed to read "(Gill HS-50)"

RC:L532 What are "red time traces"?

AR: This was changed to read "the red time traces in Fig. 8 show"

RC: Fig 2 and 5 - the legends are barely legible.

AR: These have been revised

RC: Fig 11 - please mark each panel with its own index. There are 4 marked with A and 4 with B. You can use A.1 A.2 ..

AR: Each panel is now assigned an individual index