Authors' response to Referees' comments (Anonymous Referees #1, #2 & #3)

General comments

Referee #2: General Comments - I found this paper to be well written, clear, and thorough. However, it is a bit on the long and dense side. I have only minor comments discussed below. I think the biggest area of improvement would be to make some of the figures clearer. For example, it is difficult to discern the different traces and error bars in figure 3.

Referee #3: I fully agree with the comments from referee 1 concerning the introduction and the conclusion of the MS. Both sections should be shortened substantially. Paragraphs 1 to 3 of the introduction give a lot of information about ozone deposition processes and ozone impact on plants, which are very interesting but not really useful to introduce a technical MS dedicated to the sources of uncertainty in ozone flux measurements.
These three paragraphs should be considerably shortened, and merged into one, in order to get more rapidly to the paragraph which actually introduces the MS. In the conclusion, several parts which specifically concern the analysers used in this study should be moved in the Results section (see comments by referee 1). The last paragraph of the conclusion, as well as Section 3.4, should be removed because the discussion on ozone deposition processes is beyond the scope of the MS. For the same reason, lines 2-7 of the abstract should be removed.

Answer: The revised manuscript has become more compact as the paragraphs in the introduction were shortened and section 3.4 removed altogether, along with further related sections in the 'Conclusions' chapter. The abstract was also shortened as suggested by all reviewers (see also below). The figures were improved to achieve greater clarity.

We are grateful for the detailed comments made by all referees. Revisions to the manuscript have been made on the basis of the comments, and a better & more compact manuscript is the result. Detailed answers are provided below.

Comments on Abstract

Referee #1: There's material not really belonging to an Abstract. Leaving the unnecessary parts away would also make the Abstract more concise.
It is recommended to leave lines 2-7 out of the Abstract. In line 13 the word 'calibration' should be left out, because in its present form the sentence implies that the fluxes were calibrated while in fact only the concentration fluctuations were calibrated. The content of the sentence in lines 14-16 ("It is shown . . .") is in contradiction with the sentence in chapter 3.2 (p. 2255, lines 25-27) – it should be checked what the result actually was.
In line 25 the words 'behaviour of disc' should be left out, because a potential reader unfamiliar with the instruments might not understand that 'disc' is essentially equal to 'analyser'.

Referee #2: The Abstract is too long. The first few sentences are more introductory material and do not belong in the abstract. The abstract should start with the sentence 'We present . . .'

Answer: *The abstract* was shortened by deleting the first few sentences, including lines 2-7, and it starts with the sentence 'We present…'

Line 13 – word ‘calibration’ was deleted to avoid confusion between flux calibration and concentration calibration

Lines 14-16 – The sentence in the abstract is not in contradiction with lines 25-27, p.2255, but refer to a different type of uncertainty: the sentence in the abstract refers to the errors from the regression parameters (for each method) and the sentence on p.2255 refers to the uncertainty as a result of the differences between methods. The latter is mentioned in the abstract in lines 16-18.

Line 25 – words ‘behaviour of disc’ were deleted to avoid misunderstanding

<u>Comments on Introduction</u>

Referee #1: There’s enough material in the Introduction covering the subject matter of the MS, explaining the working principles and presenting earlier results on the operation and performance of the instruments in question and giving good motivation to the work.
However, the MS would gain if the Introduction was shortened.
It is asked to pay special attention to paragraph 1 (p. 2243, lines 2-20), paragraph 2 (p.2243 lines 21-28 – p. 2244 lines 1-16) and paragraph 3 (p. 2244, lines 17-29 – p.
2245, lines 1-6) and consider whether these could be condensed. On page 2245 in line 11 the words ‘and highly sensitive’ should be left out, because it is not a typical requirement – the degree of sensitivity required depends on the measured component.
In the last paragraph the first two sentences are an unnecessary repetition of what has been written earlier in the Introduction, so it is recommended to leave these sentences (in lines 21-25, p. 2247) out.

Referee #2: The Introduction is well written and comprehensive. It does a nice job of explaining the subject matter, i.e. ozone flux measurements, the principles and problems of the instrumentation used, and the motivation behind the analysis performed in the paper. It may be a bit long for those familiar with the material but it is very informative for those with a less comprehensive background in ozone flux measurements and issues.

Referee #3: see comments on Introduction in general comments above.

Answer: *The introduction* was shortened in paragraph 1, 2 & 3 and reduced to half its length, from 762 words to 377 words. Minor rewrites were done as well as the following deletions. In paragraph 1, the sentences in lines 5-8, 11-14 and 16-20 (p.2243) were deleted. In paragraph 2, sentences in lines 20 (p.2243) to line 2 (p.2244) and lines 4-6 (p.2244) were deleted. In paragraph 3, sentences in lines 19 (p.2244) to line 4 (p.2245) were deleted. Paragraph 1 & 2 were combined as one paragraph. Corresponding references were deleted in reference list.

p. 2245, line 11 – We agree that the degree of sensitivity depends on the measured component; however *sufficient* sensitivity is a requirement for eddy covariance. The words “and highly sensitive” have been changed to “and sufficiently sensitive”.

p. 2247, lines 21-25 were removed to avoid unnecessary repetition.

Comments on Experimental

Referee #1: The content of Experimental is thorough describing well enough the measurements, data processing and analysis. A few unclear sentences and probably typing errors should be corrected. Also parts of the analysis methods are not covered till in 'Results and Discussion' and 'Conclusions and Recommendations' sections while they should be given already in 'Experimental'.

In chapter 2.2.1 in line 15 (p. 2249) the description of the co-ordinate rotation is unclear – what is meant by "best plane of fit correction" ? Also a reference to the rotation method should be given. It is also recommended that explanations of the calculation of spectral and co-spectral estimates be given in here with the appropriate references.

In chapter 2.2.2 in line 2 (on page 2250) the term 'flux covariance' is unclear. It is suggested that the whole sentence be re-written. Maybe it would be appropriate to first define flux as the covariance between the vertical wind and concentration and also explain what is basically meant by the symbols w' and X'. In line 7 it is suggested to write 'the mean absolute ozone concentration' instead of just 'ozone concentration'. In chapter 2.2.3 the equation 4 is believed to be incorrectly written in its present form – the fault is made evident by the different units of the terms in the addition. The evaluation of the ROM method compared to RM (sentences in lines 8-10, p. 2251) actually would belong to 'Results'. And in fact it is recommended leave the discussion on the offset voltages (and subsequently Figure 1) totally out of the MS. In line 11 (p. 2251) the term 'error bars' should be replaced by term 'error estimates', because 'error bar' is what one plots in a figure. In chapter 2.2.4 the equations 6 and 8 are believed to contain the same typing errors – the X should be squared in the denominators. In line 11 (p. 2252) the term 'error bars' should (again) be replaced by term 'error estimates', because 'error bar' is what one plots in a figure. In chapter 2.3 the symbol V should be replaced by symbol X in line5 on page 2253.

Referee #2: The Experimental section does a nice job of describing the measurement set-up, data processing, and analysis techniques used in the paper.

In particular I found section 2.3 data quality control very clear and worthwhile. In section 2.2.1, what is meant by de-spiked and how is a spike determined? In my experience some data is disregarded because it does not 'look right' when there is no objective reason to throw it out. I caution the authors to be very careful to not arbitrarily disregarding data. Also, how is the 'best plane of correction' determined? These concepts could be better explained of defined briefly in the text. In section 2.2.2, I am confused by the term raw flux covariance. Is it measured or calculated? What is the difference between w'X' with the bar on top versus the w' and X' described in 2.2.1. In section 2.2.3 line 11, remove word bars. I am a little uncomfortable in applying an offset determined from 15 minute averaged data to 30 minute averaged data. If the fit with the 30 minute data is uncorrelated then doesn't that imply some other processes going suggesting all the analysis be done with 15 minute averages? I think could be better explained and defended in the text.

Referee #3: Section 2.2.1: Page 2249, line 21: How was chosen the threshold value of u* (0.03 ms-1) ? It is rather low compared to what is commonly used to filter turbulence data.

Section 2.3: Page 2253, lines 24-26: This 'final visual inspection' seems rather subjective to me. What is the percentage of data removed by this way? Are they punctual 30-min flux values or do they correspond to quite long periods in specific meteorological conditions (e.g. just after rainfall)

Answer: *p.2249, line 15*, Description of co-ordinate rotation was corrected and words 'best plane of fit correction' replaced with 'planar-fit method'. Reference was added: Wilczak, J.M., Oncley, S.P., Stage, S.A.: Sonic anemometer tilt correction algorithms, Bound-Lay. Meteorol., 99, 127–150, 2001.
p.2249 – When we talk about despiking, we talk about the application of an objective criteria to identify individual spikes in the raw data time series, e.g. due to transducer malfunctioning during rain events. We are not talking about arbitrary removal of spurious 30-minute flux results. A spike in the raw time series is defined as an instantaneous value whose absolute difference to the time series' mean value is larger than 3.5 times the standard deviation of the time series, see e.g. Foken, T. Micrometeorlogy, Springer-Verlag, Berlin Heidelberg, 2009. In this de-spiking process, the spike is replaced with the previous data point.

p.2249, line 21 – The threshold limit for friction velocity depends on the roughness, heterogeneity of the underlying surface and measurement capabilities of the anemometer. During the measurement period at this grassland site, canopy height was small at 0.1 m and homogenous throughout the footprint. This produced an aerodynamically smooth surface, and a low friction velocity threshold would be expected. The chosen threshold limit is also well above the measurement capabilities of the sonic anemometer used. Whilst no clear u_*-flux dependence was evident at low friction velocities for this dataset, ozone fluxes at this site were filtered using a friction velocity threshold of 0.08 ms^{-1} in previous studies (Coyle, 2005). If this standard threshold had been used, 10 % more data would have been rejected, decreasing the number of flux values that could have been compared. It was found that the lower threshold used here produced acceptable results by removing the majority of erroneous positive ozone flux values, occurring during very low turbulent conditions, whilst also minimising the number of rejected values (only 2 %).

p.2249, section 2.2.1 – A paragraph has been added to the revised manuscript to define variables and explain the calculation of spectral and co-spectral estimates. An appropriate reference has also been given (Stull, R.B.: An Introduction to Boundary Layer Meteorology, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1988).

p.2250 – The flux and variables have been explained and defined in a separate paragraph that was included in section 2.2.1. (p.2249). The sentence in line 2-4 has been rephrased and confusing terms such as "raw flux covariance" deleted: "The ozone deposition velocity v_d (in units of m s^{-1}) can be calculated by the relative ozone flux ($\overline{w'X'}$ in units of V m s^{-1}) divided by the mean relative ozone concentration (X, in units of V)"

p.2250 line 7- using 'the mean absolute ozone concentration' instead of 'ozone concentration'

Equation 4 – the denominator in equation (4) is believed to be correct as both χ-terms are squared and the units are consistent.

Equation 6 & 8 – denominator X^2: the typo was corrected in both equations; see e.g. equation

$$(6)\ \ c = \frac{N \cdot \Sigma X_{(15)i} \chi_{O3(15)i} - \Sigma X_{(15)i} \cdot \Sigma \chi_{O3(15)i}}{N \cdot \Sigma X_{(15)i}{}^2 - (\Sigma X_{(15)i})^2}$$

p.2251, lines 8-10 – The evaluation of ROM compared to RM has been moved to the Results section (Section 3.2, line 20). We believe that this short evaluation, including the figure, helps illustrate to the reader when ROM might be more suitable to use than RM. The discussion on data quality control and calibration approaches are part of the rationale behind this work and hence the figure, which highlights the importance of QC, has been kept in the revised manuscript.

p.2251, line 11 & p.2252, line 11 – word 'bars' replaced with 'estimates'

p.2253, line 5 – symbol V replaced with X

p.2253, lines 24-26 – The final visual inspection is indeed subjective, but the aim is to identify significant outliers. In this study, 2 % of all ROFI data and 0.5 % of all GFAS data were removed through this final inspection. The removal of outliers reduced the standard deviation in the 2 ½ hour running mean by 62 - 99 %. For the 11 cases when outliers were identified, 6 cases were removal of individual 30 min flux values and 4 cases when a few hours worth of data were removed. No correspondence to rainfall events could be identified, but some outlier fluxes (4 out of 11 cases) occurred when surface wetness (as measured by a Campbell Scientific 237 Wetness Sensing Grid) changed from dry to wet. However, no consistent link was found between the outlier fluxes and environmental conditions in this dataset.

p.2253 – comment on using 15 min data for regressions and apply offsets to 30 min flux data: We believe that the main reason for higher regression coefficients being achieved by the 15 min data is a result of more data points that are used for the regression analysis. However, the referee makes a good point and we examined whether the 15 min derived offsets are very different from the ones derived from 30 min data. The typical (median) difference in offsets is 17 % for ROFI and 5 % for GFAS, which shows that mostly there is no large difference in this study. It is important to point out that no hard and fast rule or agreement exists on how to calibrate the relative ozone concentrations for flux applications. The comment made by the referee highlights the need for clear documentation of the calibration approach used, as was done here.

Comments on Results

Referee #1: The results of the study have been analysed thoroughly. At a few points the line of thought was not clear, and it is of course recommended to check them. However, one major revision would probably help make the MS more compact.
In chapter 3.2 in line 21 (p. 2255) it should probably read 'measured ozone fluxes' instead of 'absolute ozone fluxes'. In line 27 it is suggested to replace 'error' by 'error estimate', because

an "error" can generally not be known exactly. Also it is pointed out that the error bars in Figure 3 are really not discernible. Likewise the sentence in lines 16-18 (p. 2257) should be re-written so that the reference to the error is left out. In line 28 it is suggested to leave out the words 'of methods' and 'calibration or', because this would help to make the sentence more compact. In chapter 3.3 selecting date 17th Aug as representing a day when the flux data did not agree doesn't seem to be correct, because at least in Fig.4 (and Fig. 3) the two data appear quite similar the date should be checked. The term 'high frequency lag' in line 1 (p. 2260) is not clear. In fact the purpose of the analysis of the cross-correlation between the ozone signals is unclear – there was a difference of 0.25 s between the lag-times of the analysers and the cross-correlation between analyser signals simply seems to corroborate this. It is recommended to leave the sentence starting 'However, the crosscorrelation. . .' in line 27 (p. 2259) out. Linked to this also the figure 7 should be replotted without the ROFI-GFAS cross-correlation. As mentioned in chapter 2.2.2 no corrections for high-frequency losses were performed in the study. It is recommended, however, to calculate the losses and to correct the flux values accordingly. This might narrow the day time gap observed between the flux results of the two analysers. The whole chapter 3.4 could be removed, because it is considered to go beyond the scope of the MS. The MS will be extensive enough without the discussion on deposition and canopy resistance.

Referee #2: The Results of this study have been thoroughly analyzed and well-explained in the text. However, the sub-sections in the Results consist of very dense text and are a bit difficult to read. Section 3.4 seems out of place in this paper in that it discusses application and interpretation of ozone fluxes and not sources of the uncertainty. I realize it is giving a context using the uncertainty analysis. However, it may be better expanded and in its own paper.

Referee #3: Section 3.1: This section ('Meteorological conditions and analyser performance') belongs more to 'Experimental'.

Section 3.2: In this section, the authors test the performances of RM, ROM and DCM as if they were three alternative methods for calibrating any fast ozone sensors. They write (p 2255, lines 22-23): 'All three methods presented here are in principle viable options to calculate ozone fluxes'. This is not true. If a sensor has no offset, ROM is strictly equivalent to RM. If there is an offset, then RM is in principle not valid and one must (!) use ROM or DCM. Since it was shown (Fig. 1) that the ROFI analyser has a large offset (changing with each disc), and it is told (p 2255, line 17) that 'the offsets on the GFAS are larger than on the ROFI', what is the point of using RM in this study ? Instead of writing (p 2255, lines 23-25) 'it could be argued that the offsets in this study are large enough to preferably choose the DCM or ROM over the RM', the authors should clearly state, at the beginning of Section 3.2, that 'since the offsets in this study are large, it is necessary to choose the DCM or ROM over the RM', and they should leave all the results from RM totally out of the MS. This would also allow the reader to compare more easily the results from ROM and DCM in Fig. 3. Concerning the Disc Calibration Method, the authors write (Section 2.2.4, p 2252, lines 13-14) that this method 'has an implicit assumption that there is no degradation of sensitivity over the time period considered'. Fig. 2 shows that this assumption is not fulfilled for the Disc Period 6 (24/08/07-28/08/07) of the GFAS analyser. Therefore, I don't understand why, in Fig. 3(d), results from DCM for the GFAS analyser are presented during this period of 4 days when the DCM is in principle not valid. I also wonder if the loss of sensitivity observed in Fig. 2 for the Disc Period

6 is exceptional or a common feature of these discs. In the latter case, DCM would not be valid most of the time. On this point, the authors write (p 2256, lines 19-22): ‘It is possible to minimise such errors in DCM and ROM by choosing shorter periods over which to calculate calibration factors, however there is a trade-off with potentially larger errors in the fitted parameters with fewer points used in the regression calculation’. Even with fewer points used in the regression calculation, this would be much better than using a constant calibration factor over the whole disc period if the disc sensitivity strongly decreases (as for Disc Period 6 in Fig. 2). Is it what the authors actually did for Fig. 3 ? In this case, what was the time scale of the regression ? Else, results from DCM in this figure would be very uncertain and the authors could not conclude that DCM is the most appropriate method for comparing the two analysers.
Section 3.3: There are two significant issues I have with this section. First, I am not really convinced that DCM is the best method for comparing the two analysers (see above). For example, based on Fig. 3, it seems that the difference between them is lower on 22nd August when using ROM. Therefore it would be helpful to complete Figs. 4 and 5 with chronological series of fluxes obtained with ROM. It may be redundant with Fig. 3 if the latter is made clearer by removing the RM results, but it would make Section 3.3 more easily readable in itself. Second, the comparison of the analysers is handicapped by the choice of 17th August as an example of day when the flux data do not agree (see comments by referee 1). I think a better choice would be: (i) for Fig. 6, 12th and 21st August (two days with good agreement between GFAS and ROFI), 13th and 22nd August (two days with significant differences); (ii) for Fig. 7: 12th and 13th August or 21st and 22nd August. I also think it would be useful to present meteorological conditions in Fig. 4, or in Fig.5, or at least add information in the text on this point. For example, I wonder if the disagreement between the two analysers is larger or smaller for wet conditions, as the authors point out in the introduction that ‘the sensitivity of the discs depend on humidity’ (Guesten et al., 1992). Lastly, I agree with referee 1 that the presentation of the cross-correlation between the ozone signals (in Fig. 7) brings nothing to the discussion about the time lags. I also agree with referee 1 on the necessity to perform corrections for high-frequency losses, especially if the difference between raw fluxes from the GFAS and the ROFI is correlated with wind velocity (since high frequency losses are expected to increase with wind velocity and to be larger for the analyser which has the higher time lag, i.e. the ROFI). The latter could be plotted in Fig. 4 or Fig. 5, as well as air relative humidity and solar radiation, or the authors could give in a table the mean diurnal values of these meteorological variables for the days presented in Figs. 4 to 7, and precise if and when rainfall occurred during the experiment.
Section 3.4: This section (including Fig. 9) should be removed because the discussion on ozone deposition processes is beyond the scope of the MS.

Answer: *p.2255, line 21*- words ‘absolute ozone fluxes’ replaced with ‘measured ozone fluxes’

p.2255, line 27- replaced ‘error’ with ‘error estimate’

p.2257, lines 16-18 – The sentence was rewritten and the word ‘error’ was replaced by ‘error estimate’: “Although ROM fluxes on the whole agree more closely with DCM fluxes, the ROM and DCM flux values are still not generally within the error estimates as obtained by the uncertainty in the regression parameters (Fig. 3)”.

p.2257, line 28 – deleted words 'of methods' and 'calibration or'

Section 3.3 – Point on selecting day 17th as example period when fluxes do not agree: We admit that figure 3 & 4 as presented in the MS is insufficiently clear to determine that 17th August is indeed a suitable period to choose. The figure has been improved and the time series and error bars are now clearly discernable. On the 17th, the disagreement in fluxes occurs during the daytime from 8 am to about 5 pm with the absolute difference varying from a few tens of ng $m^{-2}s^{-1}$ to 100 ng $m^{-2}s^{-1}$. Fluxes agree during night time. The suggested 13th (see Referee #3) shows the same pattern and absolute differences during daytime and hence it proves no more suitable than the 17th. This is supported by the improved figures 3 & 4.

p.2259, line 27 to p.2260, line 1- the cross-correlation of ozone signals has been removed in the revised manuscript and the sentence starting with 'However, the cross-correlation..' has been deleted as suggested. Figure 7 has been re-plotted accordingly.

Section 3.1. As suggested, the section 'Meteorological conditions and analyser performance' has been moved from the 'Results' to the 'Experimental' chapter.

Section 3.2, p.2255, lines 22-23 –Comment made on all methods being viable options: We believe that the statement is correct, and that *in principle* all methods are viable options to calculate fluxes. Depending on the dataset and its appraisal, one of the three is likely to be preferable. The point made on the use of ROM rather than RM when there is an offset: We agree that in cases of very large offsets, ROM is preferable to RM, but in all other cases, it is a matter of degree and a question of setting a threshold, i.e. how big does the offset need to be to prefer ROM over RM? Regarding the follow-on point on using ROM and DCM only and not RM: We believe that the aim of this section is to highlight the potential differences in methods and show that at times (when e.g. the offset is large) RM will not produce an accurate flux. RM is a method that is likely to be used by the research community, and the many imprecise descriptions of calibration methods in the literature preclude from being absolutely certain about this. Hence this section is to highlight the importance of the choice of calculation method and consideration their implied assumptions. We believe that RM needs to remain in the analysis and form part of the discussion, as it illustrates the usefulness of documenting/reporting the calculation method which is particularly important for data analysis in operational flux networks.

The point on DCM method and sensitivity degradation, as e.g. shown by Figure 2: Disc degradation or sensitivity drift is observed in some cases/discs, but not for every disc period. Distinct or short term loss in sensitivity does occur sometimes and that is one reason why the QC procedure and regressions were performed, i.e. to identify periods when non-linearity in sensitivity occurs. Small losses or gradual drift in sensitivity cannot always be clearly identified and it will generally add to scatter around the linear fit in the regression. Certainly the baseline drift as illustrated in figure 2 is very small (-6.21×10^{-7} V ppb^{-1}, provided in the legend of figure 2). As DCM is based on a linear fit of relative to absolute ozone concentrations, the implied assumption is that the linear relationship holds for the whole time period given. This might not be the case in all discs in a very strict sense, i.e. there could be some small sensitivity drift with time. However one gains confidence from a good agreement of the relative signal with absolute ozone concentrations, and in this study, this measure of quality control and confirmation of good data quality has been suggested. Whilst for instance, the RM method does not rely on linearity of sensitivity with time as part of the calculation method, there is a lack of gauging the data quality, as the underlying (untested) assumption is that the disc behaves correctly.

p.2256, lines 19-22 –The point made on using shorter periods for obtaining calibration factors, addressing issue of drift in sensitivity: Disc sensitivity drift varies from disc to disc, and in some discs no drift is seen (during the operational period), in others abrupt, short term changes can be observed. As this occurrence of drift is so variable and not the only aspect affecting sensitivity, it seems incorrect to use a set length (say, 24 hours). Using variable lengths leads to the issue that the data do not necessarily show obvious periods of different sensitivity and the choice on period length would be subjective, which could introduce some additional uncertainty into the flux estimate. If the data allow, shorter periods might be suitable for calibration factors. We have not found this to be the case in this study. In Figure 3, all-disc period calibration factors were used and the time scale of the regression varied from 4-6 days. The point made on uncertainty in DCM and not using DCM for analyser comparison: We disagree with the statement made by referee #3. DCM is not expected to be very uncertain as all regressions have got higher correlation coefficients (R^2) than 0.5 and some regressions show little scatter around the fit. This clearly indicates a sufficiently stable sensitivity over the concentration range and with time. The uncertainty in the DCM flux as associated with the regression is estimated and documented in Table 2 and shown as error bars in the figures. Whilst using the all-disc period approach for obtaining the calibration factors does clearly produce some uncertainty (i.e. non-zero error estimates), it does not produce "results from DCM ..[that are] very uncertain".

Section 3.3 – Question whether DCM is the correct method to use for the analyser comparison section (rather than ROM) and the point made on the difference between GFAS and ROFI being smaller for ROM: We do not dispute that at times, the difference between GFAS and ROFI is smaller for ROM than for DCM. It is stated in the MS that "an empirical approach was used" to establish which method might be preferable for the analyser comparison. As described in the MS, the reason for choosing DCM in section 3.3 (p.2257 lines 23-27) is that DCM does only slightly better than ROM overall that and either could have been chosen for comparison in 3.3. In fact, using all half-hourly mean DCM and ROM fluxes, the absolute mean difference between GFAS and ROFI is 33 ng m^{-2} s^{-2} for DCM and 52 ng m^{-2} s^{-2} for ROM. (The median differences are 24 and 42 ng m^{-2} s^{-1} for DCM and ROM respectively). This difference is only small, but nevertheless it seems that in this study, DCM produces the least difference between the analysers, and was hence chosen for further analysis. Indeed it was decided to focus on only one method, rather than present both, in order to simplify the presentation (including clear figures) and the argument made about the observed difference between the analysers. Referee #3 clearly is correct in pointing out that the choice of method is debatable but we believe that no further insight is gained by adding ROM in section 3.3.

The point on flux disagreement being larger or smaller for certain meteorological conditions, as discs sensitivity depends on humidity: We have considered this point during the data analysis and interpretation and found no relationship that would link relative humidity or rainfall consistently to periods of disagreeing fluxes. For this reason, it was originally decided not to present meteorological data, however we understand the referee's request for providing ancillary data for the reader's own assessment. We have now included a separate figure with full time series of wind speed, rainfall, relative humidity, solar radiation and difference in mean DCM flux. An additional sentence has been added in the MS (Section 3.3, p.2258): "The relationship of meteorological conditions, such as e.g. relative humidity, with flux differences by the analysers has been investigated and no consistent link across the whole dataset could be found [...]."

General comments on high-frequency losses – We agree that high-frequency losses can make up a considerable fraction in some flux measurements. In this study, however, we do not consider them to be important for the analyser comparison and hence have not corrected for them. The power spectra in Figure 6 confirm that the analysers behave effectively the same at high frequencies. The co-spectra of the ROFI do not show a greater deviation from the theoretical $f^{4/3}$ slope at the high frequencies than the GFAS, which would be the case if the ROFI would not sample the high frequencies as well as the GFAS. We followed up the suggestion by referee #3 and investigated the link of flux and wind speed: the difference in raw flux from both analysers was regressed against both u_* and wind speed u, and only a poor correlation was found (R^2 of 0.19 and 0.03 respectively). This supports our argument that in this study the high frequency flux losses are not significant and cannot help explain the diurnal averaged daytime difference of about 20% in flux between the two analysers (see figure 8).

Section 3.4 (p.2261 and p. 2262) The whole section on deposition velocities and canopy resistances was removed because we agree that the implications flux uncertainties on canopy resistances do merit a more in-depth treatment, and hence its own paper.

Comments on Conclusions

Referee #1: A major revision should be performed on the 'Conclusions and recommendations'. The chapter is too long and many of the items would actually be more appropriate in the 'Experimental' and 'Results' chapters.
The sentence in lines 4 – 6 (p. 2263) belongs more to chapter 2.2.1 and it is recommended to move it there. The content of the lines 14 (p. 2263) – 22 (p. 2264) should largely be moved under the 'Results' chapter. The sentence starting in line 14 lets one to expect something general about analysers to follow, when in fact the items 1,2 and
3 treat only the analysers used in the study. In the sentence starting 'The comparison
. . .' in lines 26-28 it is said that the two analysers compare well, but the data in Fig. 8 does not corroborate this – maybe this statement should be left out. The end of the chapter, i.e. the lines 25 (p. 2265) – 10 (p. 2266) should be left out together with the removal of the chapter 3.4.

Referee #2: The Conclusion section is long. The discussion of the three main sources of uncertainty can be moved to the results and discussion section since this is really discussing the results. The paragraph spanning 2264 and 2265 seems to be a conclusion with a few sentences of recommendation at the end. The recommendations should be separated out in a paragraph of their own.

Answer: *p.2263, lines 4-6* – As recommended the sentence with references to uncertainties from eddy covariance method was moved to section 2.2.1 in the revised manuscript.

p.2263, line14 to p.2264 line 22 – Discussion on uncertainties was moved to more appropriate 'Results' section and the starting sentence rephrased.

p.2264, lines 26-28 – The sentence was rephrased and made consistent with data presented in Figure 8: "The comparison of the two analysers in this study shows that on average the diurnal trend and features, but not the absolute values at all times, of ozone deposition are captured well."

p. 2265 line 25 to p.2266 line 10 – Paragraph was removed as it related to the deleted section on canopy resistance/deposition velocities

p.2265 Recommendations sentences were moved to a separate 'Recommendations' section at the end (section 5.)

<u>Comments on references</u>

Referee #1: The references gone through for this MS are valid, well chosen and cover an impressive time span extending also up to the current year. One reference, Derwent et al. (2008), is however missing from the text itself although it's given in the reference list. This should be checked out.

Answer: Reference Derwent et al. (2008) was left in the reference list due to an oversight in the manuscript editing process and has now been removed in the revised manuscript.

<u>Comments on Table and Figures</u>

Referee #1: Tables are justified and clearly written. Regarding Table 1 it is recommended that the two missing values (symbol '-') be replaced by symbol 'n/a'.
Figures need somewhat more editing. A general note on figures is that in the present form the captions and the legends contain the same information. This is unnecessary, and should be corrected. Figure 1 should be left out, because the data in them is not essential for this MS. In figure 3 the error bars are really not discernible. In Figure 6 the axis labels are not clear – it would be good to define all the variables and symbols, preferably by adding their introduction in chapter 2.2.1. In the present form of the MS they just "pop out" from seemingly nowhere – to begin with what is 'z', 'U' etc. The legend (already) gives the symbols for the 'ROFI' and 'GFAS' data and so the corresponding sentence could be removed from the figure caption as unnecessary. What is meant with the '(-ve values)' in the legend? Figure 7 should be re-plotted without the ROFI-GFAS cross-correlation. The legend (already) gives the symbols for the 'ROFI' and 'GFAS' data and so the corresponding sentence could be removed from the figure caption as unnecessary. In Figure 8 the explanations for the line symbols in the figure caption are unnecessary, because they are given in the legend. Figure 9 is recommended to be left out.

Referee #2: Table 1, the dash line in the 2 sigma mean error column is confusing and should be replaced with n/a to be consistent with Table 2. The figures could use work making them clearer. The error bars in Figs. 3, 4, and 5 are barely legible. The longer figure captions unnecessarily repeat what is shown in the legend.

Answer: *Table 1*- symbol '-'was replaced with 'n/a' to make it consistent with missing value symbol in Table 2.

Figures – In all figures where applicable, the duplicated information on symbols was removed from the figure captions.

Figure 1 – comment to leave out this figure, see response in "Experimental'

Figures 3, 4, 5 – Error bars were made clearer and discernable.

Figure 6 – Axis labels were made clearer and the variables & symbols defined in section 2.2.1

Figure 7 – Re-plotted figure without ROFI-GFAS cross correlation.

Figure 9 was deleted as it related to the removed section on canopy resistance/deposition velocities.