Anonymous Referee #1

Specific comments

p.3, line 4: I think eddy covariance is usually employed to measure fluxes over areas with lateral dimensions of 100 m and areas of thousands of m^2 rather than tens
Statement modified as suggested.

p.7, lines 12-19: The bLs model of atmospheric dispersion (Flesch et al., 2004. Deducing ground to air emissions from observed trace gas concentrations: a field trial. J. Appl. Meteorol. 48: 487-502) could give a more accurate estimate of the surface flux than footprinting, although I have to say that Fig. 7 indicates that the footprint-corrected fluxes were very close to reality.
As indicated in the manuscript title, our main aim was to evaluate the suitability of the measurement systems for eddy covariance flux measurements. For this purpose, we needed a footprint tool to relate the measured flux to the limited artificial source area (with known flux). We agree with the reviewer, that the emission rate of such a limited source area could also be experimentally determined by accurate concentration measurements in combination with the mentioned bLs model. However, this was not the purpose of this study as mentioned above. Apart from the validation experiment presented here, the EC method is often used for flux measurements with less spatially confined sources. In such cases an application of the bLs model approach would be problematic. The main reason that we used our footprint calculation is that we are doing the link between source area and the flux on the flux level. The bLs model as e.g. available in the Windtrax software by Flesh is calculating the emissions from concentrations. In this case the background concentration is essential. In our setup we have a good knowledge on the CH_4 flux of the areas (i.e. zero around our source and a known flux from the source) whereas we can not control the CH_4 background flux.

p.8, line 6: A separation of 0.7 m between the sonic anemometer and the air intakes seems rather large and could warrant corrections for sensor separation as described by Moore 1986 (Boundary-Layer Meteorol. 37:17-35). These can be as much as 10%.
This is a misunderstanding probably caused by the somewhat ambiguous formulation. The distance 0.7 m is not related to the separation between sonic anemometer and air intake, but to the distance between the two different EC systems. As indicated in Table 1 the separation of sonic and corresponding air inlet was only 0.25 m (vertical) for both systems. In order to make this clearer, we reformulated the text as follows:
"For the flux measurements, each trace gas analyzer was connected to a separate sonic anemometer (with the air inlet 0.25 m below the center of the sonic head), thus forming two individual eddy flux systems (see Table 1). The two EC systems were mounted close to each other at a lateral distance of 0.7 m."

Regarding the damping from the sensor separation, this we take into consideration with our empirical correction derived from the ogive comparison for the QCLAS system, while the damping in the FMA system was corrected according Eugster and Senn (1995). See Reference list in the manuscript.

p.8, line 10: I wonder too about the height of the sonic anemometer. Fig.1 indicates that what it “sees” is about 60% source area and 40% grassland. Again, the bls technique could be useful here
See answers to the two previous comments.

p.13, line 1: It is also very likely that water vapour concentration is correlated with the direction of the vertical wind; up drafts from a moist surface are higher in water vapour, down drafts bring down drier air

That is exactly what we argue in this part of the text and which is accounted for in the WPL and the additional cross-sensitivity correction (if the sample air is not dried). So we do not see the problem or question in this comment.

”p.17, Section 3.1.4: It would be useful to have a concluding sentence or two for this section. Do temperature and pressure effects present problems or not?

We agree with this suggestion and added a paragraph to address this issue.

“As a concluding remark, the QCLAS is more sensitive to temperature fluctuations than the FMA. However, under well controlled environment this dependence can be efficiently suppressed. For both setups, the drifts due to ambient temperature changes are too slow to have any significant impact on EC flux measurements, but they are likely to affect accuracy of absolute concentration measurements in case of long-term monitoring applications.

p.17-18: The same comment as for temperature and pressure

We agree with this suggestion and added a paragraph to address this issue.

In summary, the response time of both setups was slightly longer than expected from the manufacturers’ specifications, which typically consider the cell volume and the sample flow. However, since this parameter is properly taken into account by the dampening correction, it does not pose any significant limitation on the flux measurements.

p.21, line 1: I wouldn’t say that a correction of 7 or 8% is small. In mitigation studies or greenhouse gas inventories, a real difference of that magnitude is very important; yet we might not be able to measure it

With the term 'relatively small' we did not want to say here that the correction is not important. In contrast we point out in the following sentences, that although the correction is only 7-8% in the present (artificial) case, it may be much larger in many other application cases. We agree with the reviewer that in mitigation studies, where small relative differences have to be detected, specific care must be taken to improve the relative/differential precision of the emission measurements.

p.33, Fig. 7, top panel: According to the units on the right hand axis, the flux is of water vapour, not of latent heat

Obviously the legend was not being updated together with the Figure and this led to a mistake in the units. The necessary correction was applied.

Technical corrections

p.3, line 25: suggest “at the surface of” for “above”
Changed

p.13, line 9: “levels” for “level’
Changed

p.13, line 19: suggest move “linearly” to after “depend
Done

p.20, line 25: Insert “to” between “according” and “the”
Done

p.33, Fig. 7, bottom panel: It is difficult to see the separate lines and dotted lines clearly. Can a bigger scale be used?
Changes have been made according to the suggestion (see updated Figure).

p.33, Fig. 7, caption, line 2: delete “of”
Done

Anonymous Referee #2

Minor comments:

Cross-sensitivity, cross-talk, and pressure broadening are all used somewhat inter-changeably, this should be reworded to be more clear "concentration" should be replaced with "mole fraction" (preferred) or "mixing ratio" throughout the manuscript as the latter are what is presented in the figures
We agree with the Referee suggestion and replaced the "concentration" with "mixing ratio" throughout the manuscript.

P 2962 L 22: "at the global scale" and "at the landscape level" are awkwardly used together in this sentence
We changed the wording to have a better formulation.
"Understanding the temporal dynamics of methane emission at the global scale requires continuous and long-term field measurements of CH_4 fluxes at representative sites, where the relationships between landscape-scale flux measurements and their environmental drivers can be investigated (Bartlett and Harriss, 1993; Bubier and Moore, 1994)."

P 2963 L 5: tens of m2 may be appropriate in an agricultural setting, but not when measurements are made higher from the surface
See response to Referee#1 above.

P2963 L 22: omit "propose to" since you have obviously already done this
Changed as suggested.

P 2964 L 6: define FMA and QCLAS here with the analyzers so it's clear
Done

P 2964 L 20: replace "minimum still-detectable" with "minimum detectable"
Changed as suggested.

P2964 and 2965: some of this will be lost on a micrometeorological audience but in my opinion this is probably appropriate amount of detail for the spectroscopy in this paper
We tried to keep the amount of detail for the spectroscopy to the necessary minimum, but to correctly address the issues connected to the optical analyzers it was impossible to completely avoid some mathematical formalism.

P 2966 L 6-7: "While the FMA is stated to be calibrated by the manufacturer, the QCLAS requires calibration by the user." All scientific instruments require calibration, regardless of the ridiculous claims by some of these optical analyzer manufacturers (I have heard this from them as well). I suggest changing the wording to acknowledge this.
We understand the point of view of the Referee, but we have to reject his suggestion. We can only recommend checking any factory set calibration. In fact, we would never do without, as reflected by our manuscript itself. However, the experimental data does indicate an excellent stability and accuracy of the FMA used in our study. These characteristics have been maintained even after transport and under field deployment. Thus, we have no foundation to make any diverging conclusion.

P2966: “intensively managed” and “extensively managed” grassland appears in Figure 1, this should probably be elaborated on in section 2.2
For clarification, we add the following text in Section 2.2:
“The field is divided into two parts and they differ in the management intensity (Ammann et al., 2007). However, during the present experiment in spring, both fields were covered by similar short grass vegetation and did not show a significant methane exchange.”
And at the end of Section 2.3:
“In some cases, the footprint was also influenced by the neighbouring extensive grassland field. This was, however, not problematic since both fields had similar short grass vegetation during the experiment in spring.”

P2911: use SI units for pressure (kPa)
We changed the units 'atm' to kPa.

P 2911 L 4: what brand sonics, and were they the same?
The brand/type of the two sonics is given in Table 1 (as indicated in the text). The two sonics were from the same company (Gill Instruments) but not exactly the same type. They have very similar size and geometry of the 3D sensor head but differ in the mounting geometry. The HS has a horizontal and the R2A a vertical mounting setup.

P2911 L 8 “large overlap” is vague
Changed to “to better match the overlap”

P 2911 L 17: “sL min-1”, presumably you mean standard liters per minute but this notation is a bit vague. Some instrument manufacturers have the company listed (the FMA and QCLAS) but others do not – be consistent and adhere to journal standards – also some part numbers are missing (eg Swagelok sintered filter)
We agree with the Reviewer and changed the “sL min-1” to the more established notation: standard liters per minute (slpm). We also added model and part numbers as well as the manufacturers name to each item we mentioned in the manuscript.

P 2969 L 15-20: I would argue that a 16h test does not characterize “long-term stability” – if you let this run for a month or a year do you feel this is adequate?
The “long-term stability” used in our context refers only to the integration time defined by the Allan variance minimum or sometimes to the time interval between two consecutive calibrations. Since the EC method works on 30 min time-series, the instrumental stability is important in the first instance within this time domain. For real long-term monitoring purposes a systematic calibration is obviously required.
We replaced the “long-term” with “stability over time” to avoid any misleading interpretation.

P 2971 L 16: “CMDL scale” is inappropriate for 2 reasons, first, the WMO is the appropriate organization that defines the scale, and second, the Climate Monitoring and Diagnostics Lab at NOAA has changed it’s name – see http://www.esrl.noaa.gov/gmd/ccl/ccl.html
We appreciate this clarification and made the proper change as suggested.

P 2972: the dilution correction is in fact a part of the WPL considerations, since the mixing ratios are expressed in the WPL paper relative to dry air – as the authors are probably well aware, this correction can be applied to the raw time series or to the fluxes and different WPL formulations apply to these cases – your eqn 6 is appropriate only if the raw time series are with respect to moist air

In eqn. 6, no 'raw' time series is mentioned. All concentration related quantities are given as mass densities (e.g. with units kg/m^3). Therefore a specification like 'with respect to moist air' is not necessary in this formulation. It would be only necessary when using mixing ratios for the trace gases.
Changed page 2975 line 10 as following: "This term is non-negligible if water vapour is not removed from the air stream prior to analysis or the measured concentrations are not individually corrected for the dilution effect. "

P 2974: "small but real" is awkward – even your artificially applied fluxes were "real"
We agree.

P2975: lines 5 and 6: the definition of sigma is awkwardly inserted into the text
Changed to: "ρ_a, ρ_v and ρ_c are the densities of air, water vapour, and a trace gas c, respectively"

P2976 L 10: "are sensitive"
Changed as suggested.

P 2978 L 25-27: but you have just explained the importance of water vapor (WPL and pressure broadening) so this is not acceptable for eddy covariance without appropriate corrections
We modified the sentence as follows to make it clearer that this statement only concern CH_4 concentration but not EC flux measurements (see also following part of the text):
"… which is acceptable for most applications if only methane mixing ratios have to be measured, but not for EC without appropriate corrections.."

P 2980 L 12: "suffering" should be replace with "poor" – also, the verb tense in this section is often present, but most of the manuscript is in past tense (as is appropriate)
We agree with the suggestion of the Reviewer and changed the verb tens to past within this section. The word "suffering" was also replaced with "poor".

P 2982: define "cw-QCL"
Added "continuous wave (cw)"

Fig 2: units needed on top axis
Added.

Fig 3: caption does not indicate which CH4 instrument is shown in top panel (also Fig 5)
Changed caption to make clear statement about the CH_4 instrument: "Instrumental response to varying water vapor content. Top graph shows the methane mixing ratio % changes measured by QCLAS over time induced by the stepwise addition of water vapour."

Fig 4 caption: "when water vapour fluxes were high" (insert the word "vapour")
Done

Fig 5 caption: illustrating two situations (plural)
Changed as suggested.

Fig 7: latent energy (heat) flux would have units of W m-2, the molar units used are a water vapor flux
Changed. See also the answer to Referee #1.

Fig 7 caption: “proportion contributed by” (omit “of”)
Done

Fig 7 caption: explain the meaning of the boxplot features (25%-tile, mean, etc)
Fig.7. Methane fluxes during the fumigation experiment. The top panel shows precipitation and the water vapour flux. Wind direction and the proportion contributed by the fumigated area to the measured flux are displayed in the second panel. In the third panel, the footprint corrected fluxes are plotted for the fumigation periods. Additionally, boxplots illustrate the difference of the measured and the applied flux for each system as well as the difference between the two instruments for each fumigation period. The upper and lower ends of the box in the boxplot are drawn at the quartiles, and the bar through the box at the median. The whiskers extend from the quartiles +/- 1.5 the inter quartile range. The gray solid line represents the applied flux. The stars above the boxplots indicate the significance level, if the values are significantly different from zero (t-test: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, • $p < 0.1$). Prior to the analysis, the lag-1 auto-correlation was removed.

Fig 7 caption: “the applied flux” is better than “the issued flux”
Changed as suggested.

Anonymous Referee #3

Specific Comments:
Page 2964, line 14. Omit the word "particle" in "particle density". Just use "density" or "number density". (the word "particle" implies aerosol)
Good suggestion, the "n" refers indeed to molecule density and not particle. Changed to "number density" as this is a proper usage.

Page 2966. Lines 1-2. What happens when the mirrors become contaminated to some degree during field studies? Does the calibration of the FMA change? I would be quite skeptical that a manufacturer's calibration will be valid for any instrument during extended field operations.
The question is justified, because contaminated mirror surfaces in principle lead to a change in the mirror reflectivity, thus to reduced effective optical path length and finally to errors in the mixing ratio measurements. The Referee question could be answered at best by the manufacturer itself. However, considering the working principle of the FMA, we anticipate that the continuous monitoring of the cavity ringdown time "τ" would compensate for contaminated mirror surfaces and will not create significant errors in the mixing ratio measurements. In extreme cases, when significant change occurs in the mirror ringdown time (accentuated declined mirror reflectivity) the precision of the instrument may be reduced. The manufacturer recommends in case of 20% reduction in ringdown time an instrument service (mirror cleaning). Being skeptical regarding manufacturer's calibration in any circumstances is the best approach.

Page 2968 – General description of both systems. Although not a critical aspect to this paper, it might be informative to readers to give the power requirements of both CH4 systems which could be reported in Table 1. Scientists interested in using these techniques may work at remote sites that are power-limited. Since both of these systems require fairly large flow rates (i.e., a fairly large pumps), this may preclude their use at some flux sites.
The optical analysers do not have high power requirement (< 400 W), but the accessories needed for reliable EC flux measurements (e.g. large pumps) have significant power consumption (0.9 - 1.5 kW). The high flow rate is critical (for closed path analyzers at least), because the gas samples should have turbulent flow in the tubing system and the instruments should have a fast time response (short exchange time of the absorption cell volume). Therefore, the usage of large pumps is mandatory for closed path analyzers used to measure accurate EC-fluxes. We consider that this issue is well known in the flux-community and since we gave complete references to the pump manufacturers, it is up to the reader to check the power requirements for these equipments.
One may overcome the issue of elevated power consumption by using open-path methane analyzers (see model LI-7700, LI-COR). This type of instrument, although does not require any pump, filter or tubing (and as such consumes much less power), still needs to be tested and characterized under field conditions regarding its ability to measure CH_4 fluxes correctly and accurately for prolonged time. We would be happy to see published work on this topic.

Page 2968, Line 3-4. How did the sonic anemometers compare in terms of measured turbulence parameters, momentum and heat flux?
See below

Page 2968, Line24. What material is the "custom-made sampling tube" constructed of? Although wall losses of CH4 would not be expected to be a problem, were any tests performed to test this?

We have added the material information. Regarding wall losses of CH_4, although not shown in the manuscript, the experimental setup was designed such that the standard gas was added at the inlet. Thus, any significant error associated to tubing would have been included or detected during the calibration step.

Page 2971, line 7. I do not know what "cross coupling effects" are?
This is an unfortunate terminology used for the sensitivity to pressure broadening. Changed "cross coupling effects" to "pressure broadening effects".

Page 2971, line 16. The scale is set by the World Meteorological Organization (WMO), not NOAA/CMDL.
Changed. See also the reply to Referee #2.

Page 2973, line 14 (and many other places in the manuscript). The authors do a very nice job of describing the water vapor effects on the absorption methods described here. However, the word "cross-talk" is colloquial and should not be used. It is not accurate of the process being described. Cross-sensitivity is a more accurate description or, better, call it what it is: a "collisional broadening effect."
We agree with this constructive comment and changed all instances of "cross-talk" to "cross-sensitivity".

Page 2973, line 14 again. Doesn't the "cross-talk" coefficient depend not only on the device, but of the interfering species. The discussion here concerns H2O, but other molecules have the potential to also collisionally affect line widths.
Exactly this statement is defined in the Eq.4, where we also state that "the transition dependent broadening coefficient quantifying the ability of a molecular species *j* (e.g. N_2, O_2, and H_2O) from the air sample to cause pressure broadening due to the collisions with the absorbing molecules." However, we focus our discussion on the water vapour only, because this has the highest concentration within the trace gases and it also undergoes large variations. This later property is the most significant term in considering the pressure broadening effect.

Page 2974. Lines 20-26. If you were unable to repeat the dilution experiment in the field – how do you know that your laboratory-derived water corrections are valid towards your field measurements using the FMA? If the filters were a problem with this experiment, how do they affect the measured CH4 fluctuations during flux measurements? Do these filters affect CH4 fluctuations indirectly by adversely affecting H2O fluctuations within the system?
Based on a recent review of all our information and further laboratory tests we concluded that the main limitation in the field experiment - and thus the reason for erroneous data - was the missing water vapour measurement in the FMA. Dampening of the water signal by the filters should have negligible effect on the CH_4 flux measurements and, thus we do not expect that the CH_4 flux is affected by the filters. We have changed the text accordingly. Furthermore, as shown later in the paper, the corresponding relative correction is relatively small under the experimental conditions.

It is noteworthy, that most recent analyzers (Fast Greenhouse Gas Analyzer, FGGA) include water vapour measurements, which greatly facilitates the corresponding experiments that are relevant at low CH_4 flux and/or high water flux as stated in the conclusions.

Page 2976., Lines 5-6. If the H2O collision l broadening effects were applied improperly, would not the CH4 fluxes still show a correlation with the H2O flux? I am not sure that the correlation with H2O flux necessarily proves anything in this instance.
We have slightly modified the formulation of the paragraph to better explain that the results are indeed affected by the uncertainties in the applied corrections:
"Applying the density correction reduces these negative fluxes by about a factor of 5. However, this still exceeds plausible limits of CH_4 uptake fluxes (Neftel et al., 2010; Smith et al., 2000). The second correction brings the CH_4 fluxes to slightly positive values that are individually not significantly different from zero, but do show a tendency to emission fluxes and positive correlation with the water vapour flux. These positive fluxes are on average higher than potential emissions from grassland vegetation reported by Keppler et al. (2006). Yet the results are strongly dominated by the uncertainties in the applied correction factors (water calibration of the QCLAS, cross interferences of the water vapour on the CH_4 mixing ratio) and sensitive to any slight change in these factors."

Page 2979. Lines 10-18. Although they seem to give similar results (within 6%), why use 2 different spectral corrections methods on the 2 separate instruments? How large are the spectral corrections relative to the total measured flux? What about corrections for sensor separation? These could be substantial at such a low measurement height.
The aim of this paper was validation and comparison of two different and fully independent EC systems for the detection of methane fluxes. The two different measurement systems, as they will be applied in the future, also include some differences in the data evaluation, e.g. the high frequency correction method. As mentioned above they had been compared and gave very similar results. The quantitative effect of the high-frequency correction was typically 12-20% of the flux for the FMA and 25-35% for the QCLAS, respectively, depending on the wind speed.

In both EC measurements the sensor separation effect has been taken into account by the applied correction methods either explicitly (Eugster and Senn, 1995) or implicitly (Ammann et al., 2006). It should be noted that we only have a sensor separation in the vertical distance, no lateral separation, and hence the high-frequency loss correction for this separation is a small component with respect to the correction needed for damping due to imperfect sampling cell air renewal.

Page 2980, 1st paragraph. Much of this section suddenly shifts to the present tense (whereas the majority of the paper is in the past tense).
Changed. See also the reply to Referee #2.

Page 2980, Lines 5-6. The use of an emission grid to generate a known flux is quite interesting, but it does raise some issues. It is unfortunate that the authors chose to have 2 completely separate EC systems instead of a single sonic anemometer and co-located inlets. This would have eliminated 2 sources of uncertainty in their flux comparisons: (1) variability between the anemometers and (2) the need for footprint corrections, which, as the authors point out, can cause significant problems during periods with variable wind directions (the footprint corrections would, of course, still be needed to compare to the calculated emission rate of the grid). However, given the described experimental setup, one needs to know how large the footprint corrections are for each system on the measured fluxes, or how the magnitude of these corrections compare between the systems (at least – on an average basis)?
It is true that the use of a single sonic would have allowed a more direct comparison of the performance of the two methane analysers. But on the other hand, a single sonic would have

increased the sources of uncertainty for the comparison with the grid emission (because a potential error introduced by the single sonic could not have been identified)!
It was an important aim of this study to validate the performance of both individual systems (as they will be used independently in the future) against the absolute reference of the emission grid flux.

To address the question of how the difference in sonic anemometers could have influenced the fluxes we aggregated the vertical wind speed of the period shown in Figure 6 and tested for differences in mean and differences in variances. The aggregation was of course necessary because of the separation distance between the two sonic anemometers. The test results show that there was no significant difference in mean velocity, but a significant difference in variances (p=0.03; FMA system has 0.89 the variance of the QCLAS system). For standard deviation this would be a ratio of 0.94, and hence the fluxes from the FMA system should roughly be 6% lower than the ones from QCLAS system. The sample statistics computed with R version 2.11.0 is:

```
>       t.test(eth.w.10s, art.w.10s)

        Welch Two Sample t-test

data:  eth.w.10s and art.w.10s
t = 0.0034, df = 2610.386, p-value = 0.9973
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.006758631  0.006782317
sample estimates:
   mean of x    mean of y
1.497849e-05 3.135465e-06

>       var.test(eth.w.10s, art.w.10s)

        F test to compare two variances

data:  eth.w.10s and art.w.10s
F = 0.8856, num df = 1310, denom df = 1310, p-value = 0.02794
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7946315 0.9869108
sample estimates:
ratio of variances
         0.8855678
```

Concerning the difference in the footprint, we added the following sentence to p2980, line 6/7: "…as the towers were slightly apart from each other (0.7 m). Yet the footprint correction factors for the two flux systems differed mostly by less than 10%."

Page 2980. Lines 28-29. Again I would point out that the author's are applying a labderived correction that could not be duplicated with the field setup. This lessens one's confidence in their results.
See reply to comment on Page 2974. Lines 20-26.

Figure 1. What is the difference between "intensive grassland" and "extensive grassland"?
They differ in the management intensity (see Ammann et al., 2007). But during the present experiment in spring, both fields were covered by similar short grass vegetation. Also see reply to Referee #2.

Figure 4. Remove the word "cross talk" in the legend – use cross sensitivity.
Changed as suggested.

Figure 6. Why not present the power spectrum in the top panel in similar fashion to the cospectrum in the bottom panel? The cospectrum shows the y-axis normalized for frequency and total covariance. If the power spectrum is normalized similarly, one can see the expected fall-off region for the 2 analyzers more clearly.

There are two reasons: (1) the display we chose is the best to show how the signal of the sensor exceeds the noise; in a display where the spectral density is multiplied with f the white noise line is increasing with increasing frequency, and the signal is seen LEFT of the white noise line (not above); this confuses most readers who are not experienced with such plots and they mistake this for a low signal-to-noise ratio; (2) if the reviewer actually wanted to say that not only the spectral density should be multiplied with f, but also a linear y axis should be chosen (as in the cospectrum), then we would agree, but such a graph barely allows to see the signal with respect to noise because the white noise level becomes exceedingly dominant at high frequencies. We therefore decided to keep the representation of the spectra as we suggested in our discussion paper.

Figure 7. The addition of the boxplots into the time series plot of the bottom panel is confusing. They should be moved to a separate panel below the time series of the measured fluxes. Furthermore, there is no description of what the boxes, error bars, etc. represent in the caption.

We moved the boxplots into a new panel. As it is not possible any more to directly compare the 30-min fluxes with the boxplots, we have chosen to illustrate the difference between the measured and applied flux for the two instruments, as well as the difference between the two instruments. Further, we added a description of the boxes, error bars, etc. in the caption.

See also the reply to Referee #2

Anonymous Referee #4

How sensitive are results of the comparison of measured fluxes to the fumigation to the footprint model applied? This should be at least discussed in the paper.
The fluxes with footprint correction are obviously proportionally sensitive to the uncertainty of the calculated footprint fraction. However, the uncertainty of the footprint calculation itself cannot be assessed individually but only together (in combination) with the uncertainty of the EC flux measurement, as in the present experiment. Thus the results presented here may as well be regarded as a validation for the applied footprint model. The performance of the specific footprint model used here in relation to an alternative Lagrangian type model has been analysed in Kljun et al. (2003).
Kljun, N., Kormann, R., Rotach, M.W., Meixner, F.X., 2003. Comparison of the Lagrangian footprint model LPDM-B with an analytical footprint model. Boundary-Layer Meteorology 106, 349-355.

Page 2969, line 24: How is the precision here defined? Is it based on 95 % confidence interval assuming normally distributed noise, i.e. 1.96 times the standard deviation of one second time resolution measurements?
In this approach, the precision (variance) is given as 1σ, consistent with the cited literature (Werle et al. 1993).

Page 2971, line 18: What does calibration factor mean here? Is it a correction factor for concentration given by analyzer (i.e. multiplicator, with unit of ppm/ppm)?
It is the slope value of the linear calibration function. See also discussion below.

Page 2971, line 19: At what response time the reproducibility measurements were conducted. I believe precision and reproducibility are the same thing, so the result here should correspond to results from previous sub-chapter.
The precision is expressed as reproducibility when the precision of the method as applied in different laboratories is taken into account (e.g. assays carried out according to a specified statistical design by different laboratories applying the same analytical protocol as part of an interlaboratory collaborative study). See also ISO 5725-2:1994 "Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method."
We replaced the term "reproducibility" with "repeatability" to avoid any confusion and to adhere to the clear definitions of these terminologies. Thus, *repeatability* is when the least changes are allowed *(*e.g. assays carried out over a short period of time, by the same analyst using the same instrument, etc.) and is calculated using

$$r = t_{\nu,\alpha} \frac{\sqrt{\hat{s}^2}}{b} \sqrt{2}$$

Where $t_{\nu,\alpha}$ is the two-sided Student t-factor (degrees of freedom ν = 2, i.e. three repeated measurements, α = 0.05), $\hat{s}^2$ is the smoothed variance function, b the slope value of the linear calibration function and while r refers to the difference between two single measurements the factor √2 has to be included. See also: P. Werle *et al.*, Signal processing and calibration procedures for in situ diode-laser absorption spectroscopy, Spectrochimica Acta, A60, (2004), 1685-1705.

Page 2971, lines 25-26: "The following paragraphs are dealing with considerations related to water vapour that are not directly covered by the WPL theory, such as the effect of water vapour on the spectral line shape". Hovewer, on the very beginning of the next page (page

2972, lines 2-6) “This is, because the laser spectrometers measure the mixing ratio of CH4 to total pressure (Pt), including also the vapour partial pressure (Pw=[H2O]Pt), : : :” Isn’t the dilution caused by the water vapor (and temperature) exactly the point of the WPL terms. I found this part a bit confusing.
The first sentence is used to introduce the topic of the Section itself, while the later text is part of the introduction paragraph where we discuss the different aspects of water vapour effect on spectral features. We believe that this becomes clear when reading the whole paragraph.

Page 2975, line 1: “: : :density flux correction equation: : :” I am not sure if the WPL terms should be called “correction” as they really are terms in the flux equation when it is derived physically correctly, not making some simplifications needed to arrive to the form containing only the term I in the Eq. (6).
We adopted the terminology used by Webb, E.K *et al*.
See also Webb, E.K *et al*., Correction of flux measurements for density effects due to heat and water vapour transfer. (1980) Quarterly Journal Royal Meteorological Society, 106 (447), 85-100. The terms 'Webb-correction' or 'WPL-correction' are nowadays very commonly used in eddy covariance related scientific literature.

Page 2976, lines 1-2: Does the small downward flux (-5 nmol m-2 s-1), seen after dilution correction, differ significantly from zero.
See our answer to Referee #3.

Page 2977, lines 15-19: Discussion on temperature sensitivity of measured concentrations. Authors should mention/discuss whether this small sensitivity has any effect on measured fluxes.
See our answer to Referee #1.

Chapter 3.1.5. I found this chapter a bit short and weak. The authors could apply a transfer function of e.g. Horst (1997) and find the response time of that function by fitting the observed spectra to the ideal one. The authors could also discuss how large is the high-frequency flux loss derived using the observed response-times.
Horst (1997) used the same approach as Eugster and Senn (1995) but did not refer to it. In his 2000 paper (Horst 2000) he corrected this omission and explicitly referenced previous work. Hence, since this is the same approach with respect to the transfer function. We use this approach in the Eugster & Senn (1995) correction to correct the high frequency damping in almost exactly the way as this reviewer would have expected after having read Horst (1997).
Horst, T. W. On Frequency Response Corrections For Eddy Covariance Flux Measurements Boundary-Layer Meteorol., 2000, 94, 517–520

Page 2979, lines 9-15: Why did the authors decide to use different high-frequency correction methods for the two analyzers?
See our answer to Referee #3.

Page 2980, lines 14-15: “As soon as the wind direction is constant, the scatter of the measured and footprint weighted fluxes is getting very small”. It would make sense to filter the data so that only the half-hourly periods with rather constant wind speed are considered in further analysis. Have the authors considered this?
We consider our investigation a basic study and we are not tempted by the idea to filter away essential information to improve our comparison between fumigation and measurement. We however agree with this reviewer that in studies that address long-term budgets, such filtering is essential to obtain defensible annual sums. In our paper we never attempted to obtain sums

over longer periods. Leaving this information in our paper, we are convinced that this helps anybody aiming at developing an appropriate filtering technique for long-term studies.

Page 2980, lines 16-19. Is the difference between the fluxes measured with two systems statistically significant? Now it is stated that it is "well below the uncertainty that one would expect for eddy covariance flux measurements". The authors should use filtering suggested in the comment above to analyze the significance of the difference.
Our aim was not to offer a complete statistical analysis of the data, but rather to present an objective characterisation of the analysers under field conditions and shed light on issues that might be of common interest for the user community. Follow-up paper focused on flux measurements will contain the requested analysis.
See also answer to the comment above.

Conclusions: The conclusions are very vague when it comes to actual instrument comparison. The authors could at least list the pro's and con's of the two different analyzers is they do not wish to give an explicit recommendation.
We try to adhere to remain neutral as is required by our guidelines for integrity in science. We would consider this a possible conflict of interest – each of the team involved has/had close and fruitful interaction with the company manufacturing the instrument, and it cannot be ruled out that any one of the co-authors has a much deeper insight and knowledge of the instrument he/she owns and uses as compared to the other instruments. The subjective decision which aspects should be considered pros or cons of a specific instrument are therefore not something that we feel we could write down without a personal bias and preference for one or the other instrument. It is therefore clearly the task of the readers, not the authors, to think about pros and cons. In our view there are not even objective criteria around that would allow us to clearly specify the pros and cons for any and all possible applications.

Minor comments

Page 2963, lines 4-5: "typically tens of m2". This is an understatement for many flux measurement sites. The measurement height can easily be 10 meters or more above displacement height in forested sites. I would say rather "typically from tens of m3 to thousands of m2".
See our answer to Referee #1.

Page 2964, 14: "particle density". I would say "number density".
See our answer to Referee #3.

Page 2965 lines 5 and 11, page 2977, line 4; "SNR" the term appears three times in the entire paper. Thus I would not abbreviate it by write signal-to-noise ratio. This would not significantly lengthen the paper but would make it easier to read.

Page 2974, line 27: "..gives us a solid evidence: : :" I would write ": : :gives us solid evidence: : :" However, I am not a native English speaker, so I may be wrong.
We agree. Changed as suggested.

Page 2978, line 14: "(Fig. 6)" It seems that this reference to Figure 6 proceeds the first reference to Figure 5. Please check.
This is correct. We removed the reference from this line.

Page 2979, line 18: “...correction reduction: : :” I would write “: : :reduction of correction: : :”
Changed as suggested.

Figure 5a: Please indicate by which analyzer this concentration time series was measured.
Done

Figure 7: Please explain how the boxplot should be interpreted.
Done. See also our answer to Referee #2

References
Horst, T.W., 1997: A simple formula for attenuation of eddy covariance fluxes measured with first-order-response scalar sensors. Boundary Layer Meteorology, 82, 219-233.
See our reply above.