

## BG-2913-293

### Author's responses to Referee 1

***3680, 5-9: The statement is probably correct, but do you have published evidence? Who publishes to have ignored something that he/she deems unimportant?***

We do not have direct, citable evidence, but indirect evidence stems from the fact that, to the best of our knowledge, no article concerned with eddy covariance and no popular EC processing software provide or suggest a means to correct absorptance biases, nor has the topic ever been raised and discussed in the community (e.g. in the FLUXNET mailing list). Although not supported by direct published evidence, we think it is important to stress that, currently, this potential source of error is likely not being duly considered at data processing time.

***13680, 19-20: Is not part of the scientific work and shouldn't be part of the abstract.***

Eliminated from the abstract, but added at the end of the conclusions. We think it is useful to many readers to know that the correction is (will be) available in a free, open-source eddy-covariance software where, in addition, all the technical details can be considered and possibly criticized.

***13680, 23-24: Reference is not in the reference list***

Reference added.

***13682, 7: I'm not sure, whether it is optimal to call this effect a 'bias in precision'. This is correct for the concentration measurement, but not for the variance and covariances, which are important here. The altered sensitivity causes purely systematic errors in variance, consequently their accuracy is biased.***

Instruments measure absolute gas concentrations (including fluctuations) and the terms accuracy and precision, as defined previously in the paper, refer to absolute concentrations. However, we do acknowledge that confusion can be ingenerated, and decided to simply remove the texts in parentheses, thereby eliminating the focus on a formal definition of the type of error, which is not essential.

***13682, 8-9: The question is, whether one needs the example of non-dispersive IRGA to understand this? To me the existence of non-linear calibration function is enough to understand the point. We don't need to know the reason for the non-linearity. Or can you state the problem a bit bigger - we also need to understand different causes for offsets in different instruments that have different effects on the flux estimate uncertainty in order to find the appropriate correction?***

We agree with the Referee that, from a purely mathematical point of view, the existence of a nonlinear calibration function is enough to understand the origin of the flux error. However, we think it is important for the readers to understand that:

- (1) The correction flows from an understanding of the theory. This is probably not the last work on this topic. We believe readers will be in a better position to understand new phenomena if they understand the theory. Therefore, even though certain elements of the theory have been described elsewhere, we believe the community will be well served by some brief repetition here in order to give a complete treatment of the problem at hand.
- (2) The relation between broadband absorption and concentration is inherently nonlinear. Otherwise, one may argue why not simply improve IRGAs by making their calibrations curve linear.
- (3) The curvature can vary slightly across different IRGAs of the same model, due to minor differences (13684, 15-16).
- (4) Contamination creates primarily zero-offsets, although small calibration changes can also occur (13693, 14-22).

We think that this material, though not strictly necessary, nicely complements the main message of the paper, allowing an understanding of the topic at an appropriate level of depth without reference to external material. To address the comment of the Referee, we shortened the text, and moved part of the introduction to the beginning of the “Materials and Methods”.

***13683, 1-14: interesting, but already published. Refer rather to the publications (Welles and McDermitt, 2005)***

The topic of this paper is the effect of contamination on fluxes. The effect is illustrated by contamination in an unfiltered LI-7200, but contamination can also affect filtered closed path instruments like the LI-7000 if filters break down or are not changed sufficiently often. Contamination will affect the LI-7000 differently than the LI-7200 because their designs are different. We believe it is relevant, useful, and appropriate to also describe how contamination is likely affect such single wavelength, dual path designs because they are in wide use. As per our previous comment, we addressed the comment by shortening the text.

***13685, 13-14: It complicates life for the reader that you define the meaning of the term ‘concentration’ depending on the gas. Please use the correct terms (e.g. ‘dry mole fraction’ and mole fraction) rather than ‘concentration’ in the text.***

In the revised manuscript, we used dry mole fraction and mole fraction instead of ‘concentration’, as suggested.

***13686, 5: Here you show that the type of the sensor doesn’t matter – the only important issue is the non-linear calibration function.***

Please refer to previous comments.

**13686, 11-12: The error propagates into 'flux estimates' not the fluxes.**

Agreed and corrected.

**13686, 16: exactly, i.e. the reference doesn't have the same value as the instruments 'assumes'. Does this make any difference compared to the offset from contamination of the sample cell?**

In the case of chemicals going bad, there will simply be CO<sub>2</sub> or water vapor in the optical path outside the sample cell even if the gas concentrations inside the sample cell are zero. This will cause an absorbance offset, but it will have no effect on the reference absorbances because the reference bands are outside the range where CO<sub>2</sub> or water vapor absorb light. Such an absorbance offset should be correctable by subtracting it off and proceeding as we describe in the paper, but it would be important to understand the time evolution of the offset.

**13687, 2: 'we quantify flux errors' -> 'we quantify flux estimation errors'**

Corrected.

**13687, 5: I wouldn't call this a hypothesis, rather a theory. It can be clearly formulated and is a mathematical consequence.**

Agreed and changed 'hypothesis' into 'theory'.

**13687, section 2.1.1 This reads like in a manual. The question is, whether one needs this detail.**

We do think the level of detail is appropriate to mathematically demonstrate the fact that the error caused by contamination is mostly a zero-offset, plus a minor span-offset (Eq. 9). There is no reason to expect so much detail in a manual. This amount of detail is essential to understand the origin of the correction. From the author's experience, we know that less rigorous derivations have inevitably lead to incomplete understanding and analyses of the problem.

**13688, 6: giving tau\_s2 another name 'tau' has confused me -, why is that necessary? Same for 'a' (13689, 3-4)**

Getting rid of that "\_s2" is convenient; otherwise, the subscript would have to be replicated all throughout the paper, making the notation uselessly cumbersome. The "\_s2"

components are the ones of interest (absorption/transmittance through the gas volume), thus it doesn't seem inappropriate to rename them to 'a' and 'tau'.

**13689, 2: Is it possible that a negative offset is caused by contamination?**

Yes, in "dual-wavelength, single-path" instruments, or in general in any case in which contamination can affect transmittance in the reference channel more than in the sample channel. While the offsets observed in this paper are positive, our recent experience shows that negative offsets are as common as positive offsets.

**13689, 16-19: Please add explanations: What do you mean by 'scaled', here? Larger effect than what? In which case? Multiplies?**

By scaled we mean "multiplied". We are simply saying that, due to the typical ranges over which  $a$  varies, in the equation:

$$a_m = 1 - \frac{z}{z_0} + \frac{z}{z_0} a$$

the additive term  $1 - \frac{z}{z_0}$  has a dominant effect over multiplicative factor  $\frac{z}{z_0}$ , in modifying the value of  $a$  into  $a_m$ . This is important to note, because it explains that the total effect of contamination is given by a dominant zero-offset plus a small (largely negligible) span-offset component. In the new text, we replaced 'scaled' with 'multiplied'.

**13689, 25: This sentence doesn't seem to have an end.**

Not sure which sentence the Referee refers to. The short version of the first sentence is "The impact ... can be found by differentiating ...". We completed the second sentence as: "Taking  $z_0$  as constant on the time scale of fluctuations, we can write:"

**13690, 2-3: Help the reader and mention why you 'consider how the bias, ...' - because the different admixture of other gases can also lead to offsets, if I'm correct.**

What we mean to describe in this paragraph is how the bias  $a_0$  in absorbance translates into a bias in gas concentration and its fluctuations. We start from the rigorous expression of the calibration function, the one that needs to be used, and then simplify the notation only for the sake of readability. With regard to "... different admixture of other gases...", adding other absorbing gases at constant pressure would both dilute CO<sub>2</sub> and water vapor, and add an absorption bias; but this phenomenon is irrelevant to the contamination issue being treated by the present paper. The same can be said of the effects of adding a non-absorbing gas.

**13691, 13: Wouldn't it be more precise to replace 'different' by 'larger than'?**

No, because the offset can also be negative (and fluctuations be underestimated), as discussed in a previous comment. In the following text we make the case of a *positive* offset, that's why there "larger" is appropriate.

**13691, 16: *'the error of too large fluctuation estimates'*.**

Changed according to suggestion.

**13691, 18-21: *As you ignore this, put it in a footnote. What is a differential effect?***

By 'differential' we actually mean 'different'. The contamination does not act as a neutral density filter attenuating the light equally in the sample and reference channels. Instead, there is a greater attenuation in either the sample channel or reference channel, which means the light spectral quality has changed in one or both channels to some degree. Such a change in spectral quality can be presumed to have an effect on the shape of the calibration curve. We didn't have data to treat that issue so we mentioned it as a possibility and explicitly neglected it. It is a relevant and important phenomenon and we don't see why it should be a footnote.

**13692, 2: *I'm not happy with an unnecessary multi letter symbol including even a unit. I recommend using a simple symbol with an index, e.g.  $\epsilon_f$ , for error in the fluctuation.***

Agreed and changed throughout the manuscript.

**13692, 9: *Is curvature measured in degrees?***

No, curvature was quantified as a departure from perfect linearity. We cannot attach an actual curvature measure to the curves, as the curvature itself changes with absorbance (i.e., the curve is not an arc of a circle). For our purposes, it isn't critical to rank curves for curvature, rather to provide an idea of the ranges over which the error varies, which is to some degree also a function of the curvature of the calibration function.

**13692, 17: *A correlation analysis is problematic, because it depends on the range chosen. Why don't you use the value of the second derivative at the point of interest?***

As per the previous comment, the method of quantification of the curvature is not critical, insofar as it is not necessary for the correction. We quantified the curvatures of the calibration curves of all LI-7200 manufactured within a period of 3 years and ranked them accordingly, only as a means to select representative curves to be used for the numerical simulations. Furthermore, defining a "point of interest" could be relatively feasible for CO<sub>2</sub>, but definitely more difficult for H<sub>2</sub>O which in typical environmental conditions spans over a much larger portion of its calibration curve.

**13692, 19-21: This has been said before (13685, 13-14)**

Eliminated.

**13695, 3: 'measure' is rather 'prediction'**

Changed according to suggestion.

**13695, 5-7: This is interesting, I wouldn't have expected the flux error lower than the  $\Delta_c$  error. Why is that so? I didn't quite digest the explanation.**

We are not sure what the Referee is referring to with ' $\Delta_c$ ', and this actually reflects an ambiguity in the manuscript, where we talk about 'concentration variations' (13695, 6-7). There are three different errors involved, and they all turn out to be different from one another:

1. Error in estimation of mean concentrations
2. Error in estimation of concentration fluctuations
3. Error in estimation of fluxes

The reason why errors 1. and 2. are different is described in the manuscript (13695, 7-10). The reason why errors 2. and 3. are different has to do the computation of EC fluxes from gas concentration fluctuations. In the specific case of our simulations, the flux error is smaller than the error on fluctuations because of the partially compensating effect of an artificially increasing water vapor density, which implies an artificially decreasing air density. For each half-hour the biased flux is the result of an increased covariance and a reduced air density, multiplied as per Eq. 1.

In the original manuscript we are ourselves victims of this ambiguity. In the revised manuscript we make a better distinction among the different errors and explain why they are different.

**13695, 10: partially compensating mechanism.**

Corrected.

**13695, 18-19: see above suggestion reg. 13692, 17**

See reply to that comment.

**13698, 12: It is a little bit funny that the manuscript continues rather naturally describing the results of the before mentioned experiment, but switching from the Methods section to the 'Results and discussion' section. Above, results were also presented analysed and**

***discussed in the same method section. I suggest a clearer distinction between theory and results sections.***

The numerical simulation and the analysis of its results were functional to the development and refinement of the correction. In fact, as we explain in the paper, the correction is merely the inverse process of the simulation. We consider the analysis of simulation results as integral part of the Methods and not as objectives of the paper. Deferring the presentation of these results to the Results section, would make it odd to present the correction method (in the previous section) as simply the ‘inverse’ of the simulation, because the simulation wasn’t yet proved to provide a satisfying model of the phenomena involved. Schematically, our logic was:

(Introduction) → (Description of the simulation → Results of the simulation → Description of the Correction) → (Results of the correction and discussion).

***13699, 16: Can you estimate how much the differing positions influence the flux measurements? In the comparisons, you assume that there were no, everything was explained by the different sensors.***

The experiment was designed to measure ‘the same fluxes’ with the two setups. Inlets to the two analyzers were placed very close to each other, and wind data came from the same sonic anemometer. The major systematic difference between corresponding fluxes was expected to stem from different spectral attenuations, which were however quite small for both systems, and carefully compensated for.

Nonetheless, sources of random errors are different for any two systems, thus we do not deem it appropriate to compare fluxes on a one-to-one basis, as if fluxes from the LI-7000 could be considered as ideals, to which those of the LI-7200 would tend if totally error-free.

***13699, 19: ‘error creation mechanism’ sounds a bit unusual.***

Replaced with “reflecting a sound understanding and representation of the origin of errors in flux estimates”

***13699, 21: ‘dispersion’ do you mean ‘propagation’?***

No, we meant that the correction shall increase the  $r^2$  in the scatter plots. We replaced the text with “...to increase the coefficient of determination”.

***13700, 21-22: ‘An improvement is achieved also for  $F_c$ , with the deviation from a perfect correlation reduced by about 39% after the correction (Fig. 8d).’ I needed to read this twice to understand the sentence – please, try reformulating this.***

Replaced with “An improvement is achieved also for  $F_c$ , with the regression slope reducing from 1.096 to 1.059”.

***13703, 18: Isn't this in conflict with the recommendation of calibrating close to application temperatures? To me this seems to be an unresolved issue.***

The apparent conflict arises because we should have written “In addition, we recommend that for the reasons already given, a field calibration *check* should be performed at a temperature comparable to the *last* calibration temperature.” What we mean to say is that, when performing a calibration check, it is important to take into account that observed offsets could be due to temperature effects. Performing the check at a temperature equal/similar to that of the last actual calibration should eliminate/minimize this uncertainty. Actual calibration shall then be performed at a temperature that, on average, minimizes the excursions during the following observation period. We revised the manuscript in this sense.

## Author's responses to Referee 3

***P13682, L10-P13687, L8. In this part of Introduction, the authors describe theoretical background of the issue the paper deals with. I, however, wonder long description in introduction makes readers feel boring, and therefore recommend that the authors separate this part from Introduction and place it prior to Section 2.1.1.***

We accepted the suggestion of the Referee and moved that paragraph to the beginning of Section 2.

***P13687, L9-11. Since both numerical simulations and analysis using field data are important contributions to the paper, it is curious for me that results of numerical simulation were placed in "2. Materials and methods" while results of analysis using field data were placed in "3. Results and Discussion". Both sections should be reorganized so that readers can easily understand the whole structure of the paper***

The numerical simulation and the analysis of its results were functional to the development and refinement of the correction. In fact, as we explain in the paper, the correction is merely the inverse process of the simulation. We consider the analysis of simulation results as integral part of the Methods and not as objectives of the paper. Deferring the presentation of these results to the Results section, would make it odd to present the correction method (in the previous section) as simply the 'inverse' of the simulation, because the simulation wasn't yet proved to provide satisfying modelization of the phenomena involved. Schematically, our logic was:

(Introduction) → (Description of the simulation → Results of the simulation → Description of the Correction) → (Results of the correction and discussion).

***P13696, L15-24. More description on the field experiment is desired, especially on measurement height as well as vegetation, soil moisture conditions, and which kind of field operations were made during the experiment.***

We included the measurement height (3.5 m) and vegetation type in the revised manuscript. We do think, however, that further bio-meteorological information such as soil moisture conditions have little relevance to our discourse.

***P13685, L4. Is "P" here the same as P in equations (2) and (3), the total pressure including water vapor pressure?***

Yes, P is total air pressure, thus including water vapor partial pressure.

***P13692, L12-15. How much the population of calibration curves?***

If the Referees meant to ask how large is the population of calibration curves, the answer more than 200.

***P13702, L17-19. I agree with this statement, but how to?***

We added the following sentence to the manuscript: "For open-path IRGAs, the suggestion is to incline the head about 10-15 degrees from the vertical, so as to facilitate water drainage. This also helps with contamination, because a dryer surface will tend to detain particulate matter less."

## **Author's responses to comments by S. Metzger**

We wish to thank Stefan Metzger for the thorough consideration of our paper and for insightful comments. In particular, the discussion on potential effects of instrumental temperature sensitivity prompted us to review some of our statements and contributed to add rigor to the manuscript. Obviously, more work is due on these aspects.

In the following, we break down Stefan's text to provide comments and clarifications.

***The authors mention a potential "temperature location bias" of LI-7200 measured concentrations as a result of different temperature levels during laboratory calibration and field validation (p. 13701 l. 7.ff.). I agree with the authors that such mean difference cancels out when calculating the eddy-covariance, as long as drift is restricted to instrument zero and does not significantly affect the instrument gain.***

We do not intend to say this. Actually, what we stress in the paper is that even a constant zero drift of the instrument, e.g. due to a "temperature location bias", does create a flux error. Our claim is the portion of this error caused by temperature is usually minor, and can be further minimized with the deployment and maintenance strategies outlined in the paper.

***[...] The authors also state that "The magnitude and even the sign of (a resulting) deviation are largely unpredictable for each given unit, but it is expected to be within the manufacturer's specifications." (p. 13701, l 18 f.) For example for CO<sub>2</sub>, the manufacturer specifies maximum drifts of  $\pm 0.3$  ppm K<sup>-1</sup> (instrument zero) and  $\pm 0.1\%$  K<sup>-1</sup> (instrument gain) for a population of LI-7200 instruments (LI-COR Inc., 2013). However, for an individual LI-7200, Bertelmann et al. (2013) were able to determine the sign and magnitude of the combined effects of zero and gain drifts to  $+0.19 \pm 0.02$  ppm K<sup>-1</sup> ( $R^2=0.97$ ). This suggests that the thorough characterization of each individual LI-7200 in a climate chamber could better constrain instrument drift by one order of magnitude. Where available, the fast LI-7200 data could also be based on simultaneously measured slow, though high-precision, high-accuracy gas analyser readings, e.g. through complementary filtering approaches (e.g., Metzger et al., 2012).***

At each point in time, it is possible to fully characterize the temperature response of each unit. In actuality, this is done also at production time, when temperature sensitivity of each unit are characterized and compensated for via firmware algorithms. However, temperature dependencies are the result of many contributing factors and may be subject to changes in time (e.g. due to aging of components and possibly, due to environmental drivers such as air humidity); thus the factory correction cannot be assured to hold under all conditions and for the entire lifetime of the instrument. What the manufacturer can assure is that such dependency will remain within the limits of the instrument's specifications. What Bertelmann and co-workers observed is the residual temperature sensitivity *beyond* the factory compensation, and it can be seen to fall within specs. The question is, for how long and under which environmental conditions that assessment would hold.

We revised the text to make this important point more clear.

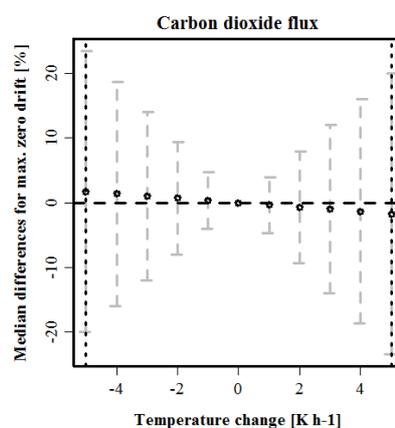
**Where available, the fast LI-7200 data could also be based on simultaneously measured slow, though high-precision, high-accuracy gas analyser readings, e.g. through complementary filtering approaches (e.g., Metzger et al., 2012).**

We make a similar comment in the paper: “Such source of error cannot be corrected using the proposed procedure, unless a paired (temperature-regulated) gas analyser is used to continuously provide a reference for accurate measurement of the gas concentrations”. We referred this statement to the error source of our concern, but it certainly applies also to the type of error Stefan is concerned with, described below.

**If left untreated, as suggested by the authors, the encountered drift (worst case  $0.3 \text{ ppm CO}_2 \text{ K}^{-1} \times 3.6 \text{ K} = 1.1 \text{ ppm}$ ) can reach the same order as noontime  $\text{CO}_2$  fluctuations.**

[...]

**The open round symbols are the median differences (measure of location), and the error bars are the median absolute deviations (measure of dispersion). The results show that the longer-term bias of the observed flux is practically negligible (open round symbols). However, the flux magnitude of individual 1-hourly  $\text{CO}_2$  flux observations is attenuated/amplified by up to  $\approx 20\%$  ( $1 \sigma$  error bars, up to  $30\%$  for  $\text{H}_2\text{O}$ ).**



We think that the design of this simulation suffers from an over-abstraction of the problem. Based on the description of the simulation, we understand that the 186 hourly periods encompass daytime and night-time periods, and the same temperature change rate was applied to all of them, in each given run. However we think that the effect of temperature changes cannot be evaluated without any consideration of the time of the day. If we interpret the simulation correctly, any given temperature change (say,  $2 \text{ K h}^{-1}$ ) will impose the exact same *absolute* variation of  $\text{CO}_2$  variance for all hourly periods. However, the *relative* effect on  $\text{CO}_2$  variance (largely proportional to the quantity shown on the y-axis of the plot above) will be different, depending on the original variance. Thus, for example, at night-time - when  $\text{CO}_2$  variance is small - the effect (i.e., the error) will be larger. However, at night time it is quite unlikely to observe large temperature change rates (this is the core of our objection). Similarly, during midday hours, large temperature changes are not commonly observed over periods of one hour.

The problem gets more relevant during transition periods (early morning, later afternoon), when the largest temperature change rates over hourly periods are to be expected. However, these periods are often flagged for poor micro-meteorological quality, because a steady-state test (e.g. Foken et al., 2004) would show that the turbulence intensity is not constant throughout the period. There appear to be, at least at a conceptual level, an ineliminable correlation between periods of strong temperature change rates extended over a long period (1 hour) and turbulence instationarity. On these grounds, we are concerned with the fact that the simulation design may lead to results that over-estimate the problem.

As a secondary consideration, this discussion appears to fall within the broader (and still largely unresolved) discussion of the interpretation of long-term trends in EC data and the user of high-pass “filters”, such as elimination of linear or nonlinear trends. While it makes sense to retain as much information as possible when calculating fluxes from raw EC data (i.e., no detrending), it is important to consider that that information *may be* irrelevant to turbulent fluxes. Low-frequency contributions can come not only from large/slow eddies and from instrumental drifts, but also from slow non-turbulent variations of the observed quantities. Thus, the strategy of data detrending and use of model cospectra for low-frequency spectral corrections is a perfectly viable alternative to simple block-averaging, which requires spectral corrections anyway (e.g. Rannik and Vesala, 1999). This digression applies to the present discussion, insofar as the simple solution of (non-)linearly detrending and correcting for spectral corrections the data used in the simulation, would identically set the error to zero for all runs, and could hardly be criticized on micro-meteorological grounds.

In conclusion, we concur with Stefan that the issue deserves deeper investigation and cannot simply be dismissed by saying that temperature sensitivity has no relevant effects on fluxes. We thus revised the manuscript in this sense. However we think that, in order to provide realistic indications, the described simulation could be improved by consideration of the following aspects:

- (1) A correlation analysis between temperature change rates and steady-state test results aimed at understanding whether strong gradients are mostly found during instationary periods, which are anyway discarded because the underlying EC hypotheses are violated.
- (2) A refinement of the numerical simulation including only data periods when fluxes are large enough (e.g. beyond the flux detection limit) and flagged as “high quality” by micrometeorological tests (e.g. Foken et al., 2004).
- (3) Extension of the analysis of temperature change rates occurrence to more sites encompassing different ecosystems, to avoid rising a general problem on the basis of data from only one site.

***Bias vs. drift: From my perspective, using the terminology of “drift” instead of “bias” might even better reflect the phenomenon treated in this manuscript. For example, the International Organization for Standardization defines:***

- ***Drift: Slow change of a metrological characteristic of a measuring instrument.***

• **Systematic error/bias: Mean that would result from an infinite number of measurements of the same measurand carried out under repeatability conditions minus a true value of the measurand.**

We did not use the term 'drift' because in the common usage it indicates a monotonic trend, while the effects of contamination can change direction in time (e.g. contaminants that deposit on the instrument window but that are eventually dislodged by the air flow, or by rain in an open-path analyzer). For this reason, our simplification of a constantly changing effect (the linear interpolation) is to be considered only as a first-order approach to the problem. The above definition of the term 'bias' does fit our usage, if applied to the absorbance error at any given time.

**p. 13681 l. 13 f.: Precision here means that the instrument is able to accurately discriminate small variations of gas concentration.**

**Using the "accuracy" terminology for defining "precision" is ambiguous. Maybe "accurately" could simply be omitted?**

Agreed and term 'accurately' eliminated.

**p. 13684 l. 25 ff.: In the following we will refer to either mole fraction or number density, with the understanding that air density fluctuations are duly taken into account when changing between the two.**

**How specifically are these taken into account? For example, by fast ideal gas law transformations, WPL formulations (Webb et al., 1980), or the WPL formulations plus the pressure covariance term (e.g., Nakai et al., 2011)?**

This is explained in the 5 lines preceding the ones mentioned in the comment. In the specific case of our field experiment, we used point-by-point transformation to mixing ratios using the Ideal Gas Law. However, the actual method chosen does not really matter, as long as all effects (T, P and H<sub>2</sub>O) are duly taken into account. Of course, the WPL formulation is to be intended to include the pressure term, whenever the relevant pressure readings are available.

**p. 13685 l. 9 ff.: Thus, for water vapour, mixing ratio is not proportional to number density or partial pressure, making it an unsuitable unit for gas analyser calibrations.**

**Dry mole fraction (mixing ratio) can be expressed as  $p_{H_2O} / (p_{air} - p_{H_2O})$ , with  $p_{H_2O}$  being the partial pressure of water vapour and  $p_{air}$  the pressure of ambient (moist) air, which is a proportionality.**

It's not a linear proportionality, though. We are not saying that mixing ratio cannot be computed from number density or partial pressure, rather that the more complicated proportionality makes water vapor mixing ratio an unsuitable measurement *for calibration* (and one empty of any physically relevant meaning).

***p. 13686 l. 8 ff.: First, because the slope of the calibration curve increases with CO<sub>2</sub> /H<sub>2</sub>O concentration, any zero offset will have a larger effect at ambient concentrations (Fig. 2) and adjusting zero will remove most or all of apparent drifts.***

***While I agree with this statement, Figs. 3 and 7 show the opposite behaviour, i.e. the errors resulting from a given zero absorptance drift are smaller at higher baseline concentrations. This might suggest that the calibration polynomial becomes more linear (less curved) at higher concentrations?***

The statement cited in the comment refers to absolute concentrations. We are simply saying that a given zero offset (in absorptance) will appear as a larger concentration offset if observed at ambient concentration as opposed to zero concentration. As an example, let's imagine we check the zero of the instrument to find a +10 ppm offset. Without zeroing the instrument, we now check the span at 400 ppm and find +25 ppm offset. We may intuitively think that those extra 15 ppm (25-10) are due to a span drift and so that after zeroing the instrument a span check would show +15 ppm. This is however not the case: most part of those extra 15 ppm are due to the same zero bias, so that, if we zero the instrument, we actually find only a residual +5 ppm offset when checking the span (please note: all numbers here are invented, only to make the case).

The answer to Stefan's question is that yes, calibration curves generally tend to become more linear at higher concentrations and this explains why in Fig. 3 the relative error in the evaluation of fluctuations tends to be smaller at higher concentrations. But this does not contradict the cited statement: our statement concerns the *absolute* effect on *total* gas concentration of a zero bias, observed at either *zero or ambient* concentration. Figure 3 represents *relative* errors in the calculation of concentration *fluctuations* observed at *any given* concentration.

***p. 13693 l. 15 f.: ...the eddy-covariance method involves several computation steps, where mean concentrations and variances of different gases interact in a complex manner.***

***As well as the correlation of gas concentrations with the vertical wind. Applies also to p. 13705 l. 11 ff.***

True of course, but not relevant to our discussion. We are pointing out why errors in fluxes are not exactly equal to errors in gas fluctuations. As long as, for a given half-hour, the bias can be considered constant, and actual concentration does not vary enormously, the gas concentration variance will change, but the correlation coefficient with  $w$  will not change significantly. In other words, if fluxes were simply given by Eq. 1 and air density was unbiased (which is not, with a drifted H<sub>2</sub>O signal), then error in fluxes and errors in fluctuations would be virtually identical.

***p. 13693 l. 25 f.: For example, positive biases in CO<sub>2</sub> and H<sub>2</sub>O concentrations, leading to increased covariances with  $w$ , partially compensate each other in the WPL term for  $F_c$ .***

***Only as long as CO<sub>2</sub> and H<sub>2</sub>O flux are of opposite sign.***

True, but worth mentioning? Again, we are explaining why errors in fluxes are not the same as error in fluctuations. If CO<sub>2</sub> and H<sub>2</sub>O are of the same sign, then the WPL will tend to

exacerbate, rather than compensate the error in fluctuations. Either way, it would tend to make the flux error different from the fluctuation error.

***p. 13699 l. 12 f.: The experimental data and numerical simulation results follow the same pattern suggesting that, to a large extent, the concentration bias explains the observed flux differences.***

***While the analytical derivation of the correction procedure and its simulation is very thorough, it would be interesting to also see how they compare quantitatively to experimental data. This also holds for Fig. 8, where it might be interesting to evaluate the significance of the correction in the light of statistical sampling errors.***

We think that such analysis isn't essential to support our case, and may actually be misleading, for the following reasons.

1. First, the origin of the error and its correction strategy are physically and mathematically based, and – on rigorous theoretical grounds – are expected to improve flux estimates. Thus, an analysis of the random uncertainty does not provide further insights on the quality of the correction.
2. The sampling error is not expected to scale with the concentration drift, as our correction does; therefore, the relative importance of the correction with respect to the random error varies with the amount of drift, making any such evaluation merely contextual to the observed experiment. In this specific sense, the analysis may be misleading.
3. Finally, the correction primarily provides a flux bias correction, while sampling error is random in nature. We do mention that we expect the correction to reduce dispersion but, to be more precise, the dispersion is expected to be reduced with respect to the *true* fluxes, not to the ones estimated from another EC system, itself affected by random sampling error.

***p. 13700 l. 22: In comparison to Fig. 6, the improvement is 3.7% and not 39%.***

We formulated that improvement rather badly in the manuscript. We did revise this sentence after a comment from Referee 1.

***p. 13711 Fig. 2: I think it might be more intuitive for the reader if the effect of the same zero absorptance bias  $\delta a$  would be shown at both, high and low concentrations.***

We agree that it would be more intuitive. However, we tried to represent in the sketch also the fact that the absorptance fluctuations themselves are affected by the bias, being slightly reduced according to Eq. 11, which is rather counterintuitive and interesting, although virtually negligible from a practical standpoint.

## References

Foken, T., Gockede, M., Mauder, M., Mahrt, L., Amiro, B. D., and Munger, J. W.: Post-field data quality control, in: Handbook of micrometeorology: a guide for surface flux measurements, edited by: Lee, X., Massman, W. J., and Law, B. E., Kluwer Academic Publishers, Dordrecht/Boston/London, 81-108, 2004.

Rannik, U., and Vesala, T.: Autoregressive filtering versus linear detrending in estimation of fluxes by the eddy covariance method, *Boundary-Layer Meteorol.*, 91, 259-280, 1999.