We thank the reviewers for helpful and insightful comments. We believe that after revision our manuscript will be clearer and better organized. The comments are addressed individually below. The referee comments are shown with bold font, and our responses are shown as plain text.

Anonymous Referee #1

It would be better if the authors add some information about the use of pump for the closed path gas analyzers.
Information about the pumps will be added to the text.

P18, line 7, loose?
Text on this line will be changed.

Anonymous Referee #2

<General comments>

In this study, four fast-response methane (CH4) analyzers were tested for determining the surface CH4 flux by applying the eddy covariance technique. It also provided strong evidence that two of the tested closed-path analyzers can work well throughout a growing season for eddy flux measurements. Measurements and data processing/analyses were carried out appropriately. The results will contribute to the flux community, and to thus Biogeosciences, which promotes the progress of full GHGs accounting and CH4 emission/abortion dynamics in terrestrial ecosystems. However, the present version of the manuscript is not well organized. I have found that many descriptions have been repeated several times in the main text. Further, the Discussion and Conclusions sections do not seem insightful. I believe that the authors can improve the manuscript by consciously reorganizing and emphasizing their findings and limitations.
We thank the referee for acknowledging our work. The manuscript will be partly reorganized based on the referee's comments, so that our message will be clearer and more concise.

<Specific comments>

P17653, L25: The word "long-term" seems ambiguous. I suggest using an alternative term such as "season-long" to express and to interpret the presented data.
The word "season-long" will be used instead of "long-term".

P17654, 2.1: Please clarify the mean height of the vegetation. It would help in understanding the spectral analysis results.
Mean vegetation height is low, approximately 10-30 cm. This will be added to the site description.

P17655 2.2: Calibration is the basis of any measurement. Hence, I suggest that the authors describe the manner in which they calibrated each gas analyzer: place (on site/off site), method (static/dynamic), date/frequency, standard gases, and relevance of the coefficients.
Calibration of each gas analyzer was checked in a lab before field deployment and this will be mentioned in the manuscript. The laboratory calibration was done close to ambient CH_4 concentration. All of them showed similar values for CH_4 concentration during the lab test and thus no additional calibration was done. The analyzers were not calibrated on site, since they are really stable and do not need frequent calibration. For instance during the first 15 days of the campaign the CH_4 concentrations (mean ± std) from RMT-200 and G1301-f were 1.911 ± 0.050 ppm and 1.894 ± 0.050 ppm, respectively, and the difference between the concentration time series was on average 0.018 ± 0.004 ppm. The CH_4 concentrations from the same instruments approximately 6 months later during a 15-day-period at the end of the campaign were 1.893 ± 0.041 ppm and 1.882 ± 0.051 ppm, respectively, and the difference between the time series was on average 0.011 ± 0.054 ppm. Therefore the difference between the concentration values stayed almost the same during the measurement campaign which might indicate that the instruments were not significantly drifting.

Hendriks et al. (2008) tested the stability of FMA (Los Gatos Research, USA) which is similar instrument as RMT-200 in our study. They calibrated their instrument daily during a ten day period and saw no change in the calibration coefficients which indicates that the instrument was stable and frequent calibration was not needed. In addition the eddy covariance community has already few years of experience with these instruments based on laser absorption spectroscopy and they are known to be stable.

P17655 2.2.1: Please specify the difference between the LI-COR prototype 7700 used in this study and the commercial version of LI-7700. If the two do not significantly differ with respect to their sensor specifications, the authors could refer to the prototype as LI-7700.
In the case of this specific Prototype-7700, numerous changes were made: (i) re-design and sealing of the housing, (ii) re-design of electro-optical components and interference shielding, (iii) changes in laser signal processing, (iv) structural re-design and spars attachments, (v) motor re-design, (vi) clock and synchronization changes, (vii) major changes in firmware and software.

Essentially, the instrument went through major re-design in all areas, particularly based on the results of the tests from the submitted paper.

P17660, 2.3.2: This subsection seems redundant, and it is difficult to judge whether the authors' calculations are scientifically sound. Please reorganize this subsection.
We agree with the reviewer that most of the information given in this subsection can be found in other papers and thus it is somewhat redundant. However since the WPL and spectroscopic corrections are important and they are carried out differently for different instruments, we think that introduction to the used methods should be included in the manuscript. In any case, based on the referee's comments, this subsection will be modified accordingly.

P17660, L11-15: Pressure fluctuations can be neglected only when their effect on the WPL correction is considerably smaller than the other effects. Did the authors confirm this for their system (closed-path measurements) or environment (open-path measurements)?
The referee is right; nothing should be assumed negligible without first testing if the assumption is valid. During the measurement campaign covariance between vertical wind speed and pressure was not measured and thus the pressure term in WPL-equation (last term on the right hand side in equation (1) in Zhang et al., 2011) cannot be calculated. However rough estimate for the term can be obtained by using results from other studies.

The pressure term in WPL-equation can be written as (e.g. Zhang et al., 2011):

$$F_p = \frac{-\overline{\rho_c}(1+\overline{\chi_v})\overline{w'p_a'}}{\overline{p_a}}$$

where ρ_c is mass density of methane, χ_v is the volume mixing ratio of water vapor, $p_a$is the ambient pressure and w is vertical wind component. According to Zhang et al. (2011) $\overline{w'p_a'}$ can be estimated as a function of stability parameter $\zeta = (z-d)/L$, total air density $\overline{\rho_a}$ and friction velocity u_*:

$$\overline{w'p_a'} = F\overline{\rho_a}{u_*}^3$$

where

$$F = 2.3\zeta - 0.2, \text{ when } -1 \le \zeta \le 0$$

Parameterization for F was adopted from a grassland study (McBean and Elliott, 1975). Zhang et al. (2011) stated that F is a function of surface roughness and thus parameterization for F was adopted from McBean and Elliott (1975) who had similar surface roughness as in our study site. Using these equations F_p was estimated for periods when $-1 \le \zeta \le 0$. Median value of $3.5 \cdot 10^{-4}$ mg m^{-2}h^{-1} was obtained for F_p. 25th and 75th percentiles were $1.4 \cdot 10^{-4}$ mg m^{-2}h^{-1} and $7.7 \cdot 10^{-4}$ mg m^{-2}h^{-1}, respectively. During the same period for instance H_2O-term in WPL-equation for RMT-200 was approximately $8.4 \cdot 10^{-2}$ mg m^{-2}h^{-1}, which is over two orders of magnitude larger than the pressure term F_p. The obtained results support our initial assumption that the pressure term in WPL-equation is negligible and it can be ignored.

The obtained value for F_p is a rough estimate on how pressure fluctuations affect measurements done with an open-path instrument. For closed-path F_p may be slightly different, due to use of sampling line and the fact that the closed-path instruments control internal pressure. According to Lee and Massman (2011) static pressure fluctuations should pass through the sampling line almost without any time lag, while CH_4 had a delay time of several seconds in our system. Zhang et al. (2011) found that $\overline{w'p_a'}$ decreases by 70 % in magnitude if p_a' was delayed by 10 s and thus we can assume that the effect of pressure fluctuations to our closed-path measurements are smaller than the rough estimate obtained above. From these results we can conclude that also for fluxes measured with closed-path gas analyzers F_p should be negligible. Short discussion presenting the latest results with this issue (Zhang et al., 2011; Nakai et al., 2011; Burba et al., 2012) will be added to the text.

P17663, L9-29: I did not completely understand this paragraph, particularly the calculation of the water vapor flux that is used for the correction of each analyzer. The authors mentioned that “WPL terms were calculated using RMT-200 methane time lag” in L20-21, but according to my understanding, the LI-7000

analyzer was connected to a different line with a different flow rate. Therefore, I suggest that the authors explain the lag calculation in further detail.
We agree with the referee that this part needs some clarification. RMT-200 does not measure H_2O and thus H_2O needs to be measured with some other gas analyzer in order to make all the necessary corrections to RMT-200 CH_4 measurements. When doing WPL-correction to closed-path instrument data the H_2O-term in the correction should correspond to the circumstances in the measurement cell (Massman, 2004; Ibrom et al., 2007). That is w'H_2O' covariance used in calculating H_2O-term in WPL-correction should be calculated with RMT-200 CH_4 time lag and, moreover, it should be attenuated to the degree that RMT-200 sampling line attenuates H_2O fluctuations. Then it describes the density effects that H_2O fluctuations have on RMT-200 CH_4 measurements.

As previously stated, RMT-200 and LI-7000 had different sampling lines with different flow rates and filters. The referee is right; this should be taken into account when doing WPL-correction to RMT-200 CH_4-fluxes. LI-7000 H2O measurements are most likely attenuated differently than H_2O in RMT-200 measurement cell and the lag time is different. Moreover, these parameters probably respond differently to changes in relative humidity, since LI-7000 sampling line was heated, while RMT-200 sampling line was not.

In order to estimate WPL-correction properly, difference between H_2O and CH_4 lag time and attenuation of H_2O signal in RMT-200 sampling line should be known. If they are known, then H_2O in RMT-200 measurement cell can be reconstructed from LI-7000 measurements and the correction can be done properly. However, estimating H_2O signal attenuation in a sampling line based on theoretical derivations does not work, since no theoretical transfer function describe the adsorption and desorption of H_2O in a sampling line properly (e.g. Mammarella et al., 2009). Also difference between H_2O and CH_4 lag time cannot be estimated, due to the fact that the effect of adsorption and desorption mechanisms to H_2O lag time in a sampling line which consist of tubes and filters cannot be modeled properly. In essence, due to the fact that RMT-200 and LI-7000 had different sampling lines, the WPL-correction to RMT-200 CH_4 measurements cannot be done in a correct manner. This will be briefly explained in the revised version of the manuscript.

If we assume that crosscorrelation function between vertical wind *w* and water vapor concentration follows exponential shape and it is symmetrical around the maximum value, then the relative error in the correction caused by wrong lag time can be estimated as

$$\sigma_{lag} = \begin{cases} 1 - e^{-\frac{\Delta t}{T_{wc}}}, & when\ \Delta t \geq \delta t \\ 1 - e^{-\frac{2\delta t - \Delta t}{T_{wc}}}, & when\ \Delta t < \delta t \end{cases}$$

where T_{wc} is the integral time scale of w'H_2O', which describes how long time series *w* and H_2O are correlated with each other, δt is the difference between H_2O and CH_4 lag times and Δt is the error in the difference (see also Figure 1). For instance if CH_4 lag time is 3 seconds and the correction is done assuming that H_2O lag time is 4.5 seconds, even though the correct value for H_2O lag time would be 4 seconds, then Δt is 0.5 seconds and $\delta t = -1s$. On the other hand, if $\Delta t = \delta t$, in this case $\Delta t = -1s$, we would get an error estimate for the case that H2O-term in WPL-correction would be calculated with H_2O lag time, not CH_4.

Values for T_{wc} were estimated by fitting exponential curve to crosscorrelation function between w and H_2O. Median value of 0.99 seconds (25^{th} and 75^{th} percentiles were 0.49 s and 1.77 s, respectively) was obtained for daytime situations in June. We can estimate H_2O lag time in RMT-200 sampling system by assuming that ratio between CO_2 and H_2O lag times in LI-7000 sampling line is the same as ratio between CH_4 and H_2O lag times in RMT-200 sampling line:

$$\left(\frac{CO_2\ lag}{H_2O\ lag}\right)_{LI-7000} = \left(\frac{CH_4\ lag}{H_2O\ lag}\right)_{RMT-200}$$

By doing this assumption it is assumed that absorption and desorption of H_2O is the same in both sampling lines. At daytime in June the lag times for CO_2, H_2O in LI-7000 and CH_4 in RMT-200 were approximately 1.65 s, 2.20 s and 2.96 s, respectively. By using the expression above H_2O lag time in RMT-200 sampling line can be estimated to be 3.76 s and the difference between H_2O and CH_4 lag times, δt, in RMT-200 sampling line to be $\delta t = 2.96s - 3.76\ s = -0.80\ s$. In this kind of situation the error in the difference, Δt, can be obtained by $\Delta t = \delta t - \delta t_{used} = (2.96s - 3.76s) - (2.96s - 2.20s) = -1.56s$. A value of -0.04 is obtained for σ_{lag}, meaning that the correction is overestimated by 4 %.

However, LI-7000 sampling line was heated, while RMT-200 sampling line was not and thus the error is probably underestimated, because H_2O lag time in RMT-200 sampling line is likely longer than the estimated 3.76 s and therefore the difference between CH_4 and H_2O lag times, δt, in RMT-200 sampling line is longer than the estimated -0.80 s. A better approximation for the error is obtained using values 3.8 s and 3.9 s for H_2O lag time in RMT-200 sampling line, which correspond to values $-0.84\ s$ and $-0.94\ s$ for δt and $-1.6s$ and $-1.7s$ for Δt, respectively. With these values, σ_{lag} is approximately -0.08 and -0.20, respectively. During the investigated period, the estimated WPL-correction to RMT-200 CH_4 fluxes was on average 0.206 mg m^{-2} h^{-1} and 4, 8 and 20 percentages of this value are 0.008 mg m^{-2} h^{-1}, 0.017 mg m^{-2} h^{-1} and 0.041 mg m^{-2} h^{-1}, respectively. Thus during this period the overestimation in final fully corrected RMT-200 CH_4 fluxes due to erroneous difference between H_2O and CH_4 lag times is approximately 0.01-0.05 mg m^{-2} h^{-1}, which is 0.3-2 % of the raw uncorrected RMT-200 CH_4 flux.

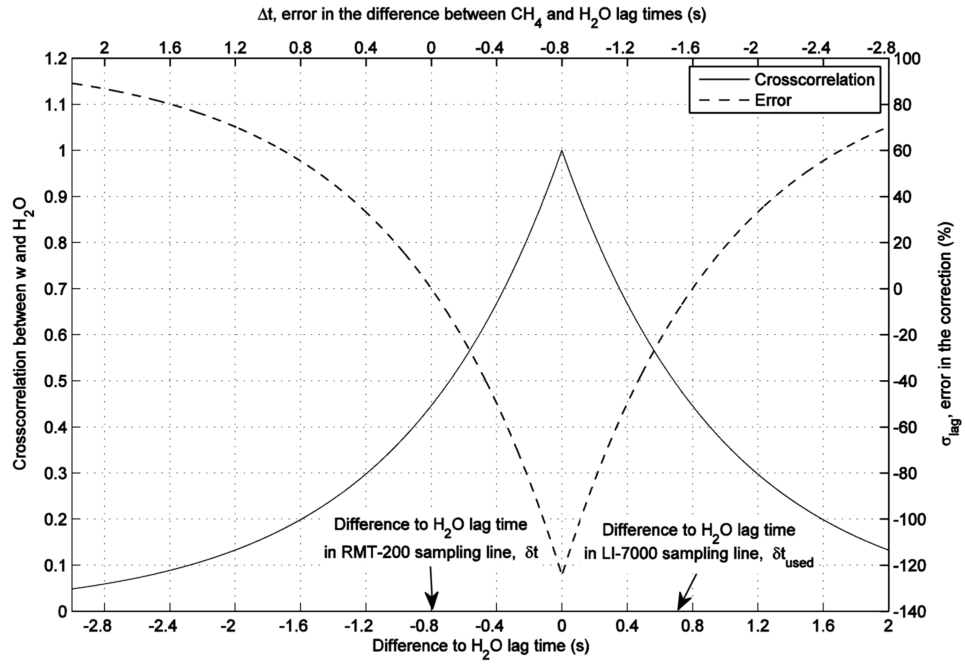

Figure 1: Figure showing an example of crosscorrelation between w and H_2O and estimate for the error caused by an error either in CH_4 or in H_2O lag time. In this case CH_4 lag time is 2.96 s, H_2O lag time in LI-7000 sampling line is 2.20 s and H_2O lag time in RMT-200 sampling line was estimated to be 3.76 s. The dashed line shows the relative error in the correction, σ_{lag}, as a function of Δt (right y-axis and upper x-axis).When making this plot a value of -0.8 s was used for δt and 0.99 s for T_{wc}.

Error caused by different attenuation of H_2O fluctuation in LI-7000 and RMT-200 sampling lines can be estimated by calculating correction factor CF (Eq. (5) in the manuscript) with different values for response time τ. The response time describes how much the sampling system attenuates high frequency fluctuations. Most likely RMT-200 sampling system attenuated more H_2O fluctuations than LI-7000 sampling system, since it was not heated. Therefore the WPL-correction is probably overestimated, when LI-7000 H2O is used.

In June, during periods when relative humidity was between 55% and 65%, the correction factor CF for LI-7000 H_2O flux was on average 1.09 and response time between 0.14 and 0.17 s. If we estimate that in these cases a value of 0.3 s for τ describes how much H_2O fluctuations are attenuated in RMT-200 sampling system, then a mean value of 1.17 is obtained for CF. Thus during this period the WPL-correction is overestimated approximately by 7 %, due to different attenuation of H_2O signal in LI-7000 and RMT-200 sampling systems. However, this approximation for the error depends strongly on what kind of value is estimated for τ. If we use values of 0.2 s and 0.4 s for the response time, we will obtain values of 3 % and 12 % for the overestimation, respectively. During this period the WPL-correction to RMT-200 CH_4 fluxes was approximately 0.160 mg m^{-2} h^{-1} and 3, 7 and 12 percentages of this value are 0.005 mg m^{-2} h^{-1}, 0.011 mg m^{-2} h^{-1} and 0.019 mg m^{-2} h^{-1}, respectively. Thus it can be assumed that during this period, the fully corrected RMT-200 CH_4 flux was overestimated by 0.005-0.02 mg m^{-2} h^{-1} due to this difference in H_2O attenuation. These values correspond to 0.4-1.4 % of the raw uncorrected RMT-200 CH_4 flux.

The error estimation revealed that both of the error sources (wrong lag time and different attenuation of H_2O) induce in this case an overestimation of the WPL-correction that is done to RMT-200 CH_4 fluxes. The overestimation caused by wrong lag time is larger than the error caused by different attenuation of H_2O in the RMT-200 and LI-7000 sampling systems.

P17664, 2.4: Thank you for separately estimating the instrumental noises. However, this is not appropriately emphasized in the Results section (P17673, L3-29).
We will analyze instrumental noises in more detail in the subsection 3.3.

P17671, P2-4: Is this true?
The sentence on these lines is slightly unclear and this part of the manuscript will be modified.

P17673, L3-29: These paragraphs are confusing. The AFFE value was determined using Equations (11) and (14), and thus contains both environmental (one-point sampling) and instrumental errors. The authors should use Equation (15), not Equation (11), if they want to conclude the performance of the instruments as in L18-20. In fact, most of the random error resulted from one-point sampling, not from instrumental noise (Fig. 2).
Instrumental errors estimated with Equation (15) will be discussed in more detail in this subsection. However in our opinion also AFFE values should be analyzed since differences in these values should result only from differences in CH_4 analyzer performance. This is because all the other factors affecting AFFE values are the same (measurement site characteristics, sonic anemometer) for all the four methane flux time series.

P17676, 3.4.2: This subsection is also redundant and confusing. For example, the observed diurnal variations of the WPL terms were self-evident according to the WPL theory, and therefore not worth noting in the body.
Subsection 3.4.2 will be modified based on referee's comments. However in our opinion short description of WPL and spectroscopic corrections are needed, since they clearly have a significant effect on the measurements. The differences in these corrections highlight the dissimilarities between the methane instruments used in this study.

P17682, Discussion and Conclusions: I strongly recommend that the authors reorganize this section. Many descriptions given in this section have already been mentioned in Results or elsewhere. The authors could discuss the critical points of this study more consciously by considering the generalization/applicability of the presented findings.
This section will be modified. Overlap between Results section and Discussion and Conclusions section will be removed and applicability of the findings will be discussed in more detail.

<Technical comments/corrections>

Overall: The body is a little confusing and redundant. I found some descriptions in Results that should be moved to Materials and Methods or Discussion and Conclusions. Many sentences in Results are repeated in Discussion and Conclusions. In-depth interpretations of the data that are not mentioned in Results are expected in Discussion and Conclusions.
As addressed in the previous comments we will reorganize the body based on the referee's comments.

Overall: Both British and US English are used in the manuscript.
We will check the grammar and convert British English to US English.

P17658, L2: “L” is used one line before to abbreviate the latent heat of vaporization.
Will be corrected.

P17658, L5: “u” and “v” are not used in Equations (1)–(3).
Sentence on page 17658, line 5-6 “Here *u* and *v* are the two horizontal wind components and *w* is vertical wind component” will be replaced with sentence “*w* is the vertical component of 3D wind vector.”

P17659, L2-14: This paragraph can be understood very well but may not be needed for this manuscript.
Based on the referee’s comment, this paragraph will be removed.

P17661, L2: “different spectroscopic”?
Will be corrected.

P17662, L8: X(chi) is used for denoting the molar mixing ratio of CO2 in other places of the manuscript. It is confusing.
Mole fraction of gas *i* will be denoted by ξ_i instead of x_i, in order to avoid confusion.

P17667, L2: Add the DOY labels to the x axis of Fig. 2.
DOY labels will be added to the figure.

P17669, L11-23: These sentences do not express a result and should be noted elsewhere.
The referee is right, these sentences fit better to Discussion and Conclusions section. They will be moved there with some modification.

P17671, L8-9: I do not understand the syntax of this sentence.
This part of the manuscript will be rephrased.

P17674, L20-22: This should be noted in Materials and Methods.
The referee is right; the cut-off frequency will be mentioned in subsection 2.3.1.

P17675, L10-13: This sentence should be moved to Discussion and Conclusions.
The content of this sentence is already included in Discussion and Conclusions and therefore this sentence is removed.

P17675, L24: “partly cancel the effect…”?
Yes, the word “cancel” fits better than the word “balance”.

P17676, L15-19: The sentences are already stated in Materials and Methods.

These sentences will be removed.

P17676 L20-25: They should be discussed in Discussion.
These sentences will be moved to Discussion and Conclusions section.

P17681 L16-17 "…, namely…": This is also redundant.
Will be removed.

Fig. 1: Please explain the line colors in the caption.
Line colors were already explained in the caption.

Fig. 2: Add DOY labels to the x axis. Is it possible to combine Fig. 2 with Fig. 10?
DOY labels will be added to the x axis and figures 2 and 10 will be combined.

Fig. 4: This figure seems less informative. Please emphasize the necessity of this figure in the body.
We will emphasize more the instrumental noise estimates in section 3.3 and refer to this figure for estimates of the average values during the two selected periods.

Fig. 6: The fitted line looks different from the expectation from the solid circles for G1301-f, particularly at relatively high frequencies. Is it true?
Yes, the referee is right; the fit in this figure is not the best possible. A better agreement with the measurements was obtained when a value of 0.08 s was used for G1301-f response time. Spectral corrections for G1301-f methane fluxes were recalculated and the results were adjusted to correspond to the new flux values.

Fig. 7: The legend should not overlap.
Good point, the figure will be modified accordingly.

Fig. 9: The legends and dots should not overlap. The body states that the small slope value for the left-middle panel (0.886) resulted from some points located at a relatively high position (P17680, L15-17), but I do not agree, because the number of such irregular data is very small in the entire set of data points.
The figure will be modified so that the legends and dots will not overlap. We agree with the referee that in a case when a linear fit is done to 2935 points, such as in this case, a few irregular points should not affect the fit extensively. That part of the text which states that the small slope value was caused by irregular data was based on inspecting the residual of the fit (data minus fit). The residual was not shown in the manuscript. If the fit describes the data well, then residual values should randomly vary around zero. However, when CH_4 flux is between 6…12 mg m^{-2} h^{-1} the residual is more often positive than negative, meaning that the fit underestimates the flux and it does not describe well the agreement between TGA-100A and G1301-f methane fluxes.

Cited literature in the comments

Burba G, Schmidt A, Scott RL, Nakai T, Kathilankal J, Fratini G, Hanson C, Law B, McDermitt DK, Eckles R, Furtaw M, Velgersdyk M (2012) Calculating CO2 and H2O eddy covariance fluxes from an enclosed gas analyzer using an instantaneous mixing ratio. Glob Change Biol 18:385-399

Hendriks DMD, Dolman AJ, van der Molen MK, van Huissteden J (2008) A compact and stable eddy covariance set-up for methane measurements using off-axis integrated cavity output spectroscopy. Atmos. Chem. Phys. 8:431–443

Ibrom A, Dellwik E, Larsen SE, Pilegaard K (2007) On the use of the Webb - Pearman - Leuning - theory for closed-path eddy correlation measurements. Tellus B 59:937–946

Lee X, Massman WJ (2011) A Perspective on Thirty Years of the Webb, Pearman and Leuning Density Corrections. Boundary-Layer Meteorol 139:37-59

Mammarella I, Launiainen S, Grönholm T, Keronen P, Pumpanen J, Rannik Ü, and Vesala T (2009) Relative humidity effect on the high frequency attenuation of water vapour flux measured by a closed-path eddy covariance system. J. Atmos. Ocean. Tech. 26:1856–1866

Massman WJ (2004) Concerning the measurement of atmospheric trace gas fluxes with open- and closed-path eddy covariance system: the WPL terms and spectral attenuation. In: *Handbook of Micrometeorology* (eds X. Lee, W. J. Massman and B. Law). Atmospheric and Oceanographic Sciences Library. Kluwer Academic Publishers, Dordrecht, Boston, London, 133–160.

McBean GA, Elliott JA (1975) The Vertical Transports of Kinetic Energy by Turbulence and Pressure in the Boundary Layer. J. Atmos. Sci. 32:753-766

Nakai T, Iwata H, Harazono Y (2011) Importance of mixing ratio for a long-term CO2 flux measurement with a closed-path system. Tellus B 63B:302-308

Zhang J, Lee X, Song G, Han S (2011) Pressure correction to the long-term measurement of carbon dioxide flux. Agric For Meteorol 151:70-77