Answer to Reviewer 1:

General comments:
I appreciate the idea of developing a lower-cost CO2-flux measurement system and find the paper to be well written and of appropriate length and depth. The methods are well documented, as are the results, although I wish some more details on the calibration were shown. On the one side, I find the lack of reference/calibration data for the CO2 flux measurement part a bit disappointing. I wish the authors could provide a comparison with a commercially available flux measurement system, but I understand that organizing such an experiment is not easy. On the other side, I have doubts regarding the CO2 measurement calibration model and the necessity of a relatively complex calibration setup for the particular application. I will discuss these points in more details in the following section.

Dear Simone Baffelli,

First of all, we want to thank you for reviewing this manuscript and for your appreciation of the work done. We found your comments really useful and they have helped us to improve the quality of the manuscript.

Regarding the lack of a comparison with a commercially available flux measurement system, we would like to underline that the main aim of this paper was to present the design and calibration of a low cost sensors kit (Air Enquirer) and to show the possibility of using it in dynamic flux chambers to reduce the price of possible CO_2 flux networks. We noted the lack of a robust metrology chain in these type of measurements and we think it is important to describe and proposed it. However, we also believe that it will be really important to test our CO_2 dynamic flux system with other available systems being them commercial or from other research groups. Actually, we are going to submit a new project to, among others goals, buy a new Licor CO_2 flux system (https://www.licor.com/env/products/soil_flux/) and make an intercomparison campaign between different systems. This has been now better clarified in the conclusion of this manuscript as further actions. In addition, we have changed the title of the manuscript to better fit with its content.

Following your comments, we have expanded the calibration section in order to better describe the processes applied for the calibration of the sensors. Moreover, we are now proposing different type of calibrations, we have calculated the errors associated to each one of them and we are doing recommendations for calibration and recalibration of the instruments depending on their specific applications. We think this gives to the study a strong metrology basis.

In the following lines we discussed each one of your comment and we indicate the respective changes, if it is the case, within the manuscript.

1) L17: You wrote that the results were corrected for illumination. This does not appear anywhere in the main body of the article. Please either remove this or edit section 2.2 and the results section to reflect this. Do you think illumination would have an impact on the CO2 calibration? I could only imagine an indirect effect through heating of the sensor, which is already captured by the temperature calibration.

Answer: The reviewer is right. Within this study the data of the illumination sensor was not used. This has been clarified in the revised version of the manuscript (Line 17). This sensor was added

at the kit because we thought it would be interesting for phenology and other ecological studies, but it is not used for the calibration of the CO_2 sensors.

2) L88: In my experience employing very similar sensors (SenseAir LP8), air humidity has a large effect on the data quality, particularly for RH> 80, causing a highly nonlinear saturation in the measured concentration. Could you please report the RH range used for calibration? As you measure directly above soil, I would expect the RH in the chamber to routinely reach these values. Did you experience this? This could have an impact on data quality, particularly if you experience a sharp increase in RH within the chamber during the measurement time. Could you show the graph of RH for the linear accumulation experiment of Figure 6?

Answer: As asked by the reviewer we have added both temperature and relative humidity values measured within the NSS-NTF chamber during the measurement examples shown in Figure 6 (in the revised version is Figure 7). In the NSS-NTF there is always an increase in temperature and relative humidity, but this increase is not sharp, probably thanks to the really short period of time of the measurements. In addition, for the reviewer we added here in Figure A1 a plot with the temperature and RH data from the SSTF experiment of Figure 5 (now Figure 6). We did not observe RH > 80% neither for NSS-NTF or for SSTF chambers.

For the calibration experiments, the RH range was 10%-50%, and the Temperature ranged in the interval 20 ^{0}C -42 ^{0}C. Plots of this data have been added into the revised version of the manuscript (Figure 4).
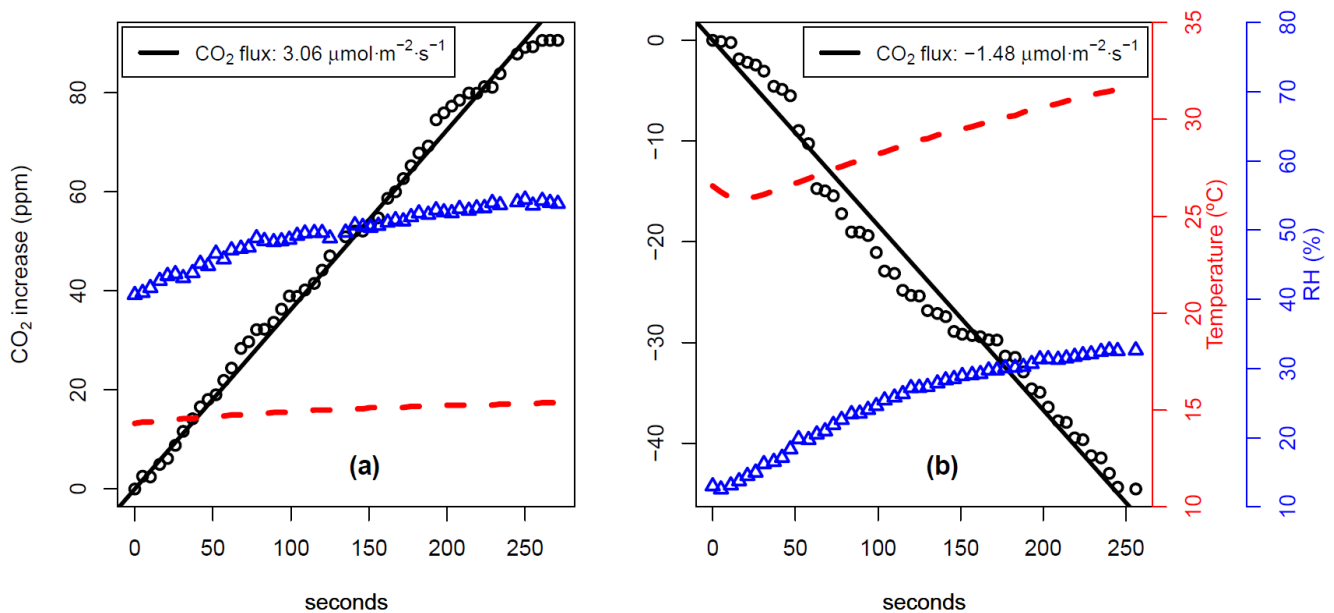

Figure 7. Example of two cases where the linear accumulation method was applied within an NSS-NTF chamber to calculate positive (a) and negative (b) CO_2 fluxes with Air Enquirer Kit #03.
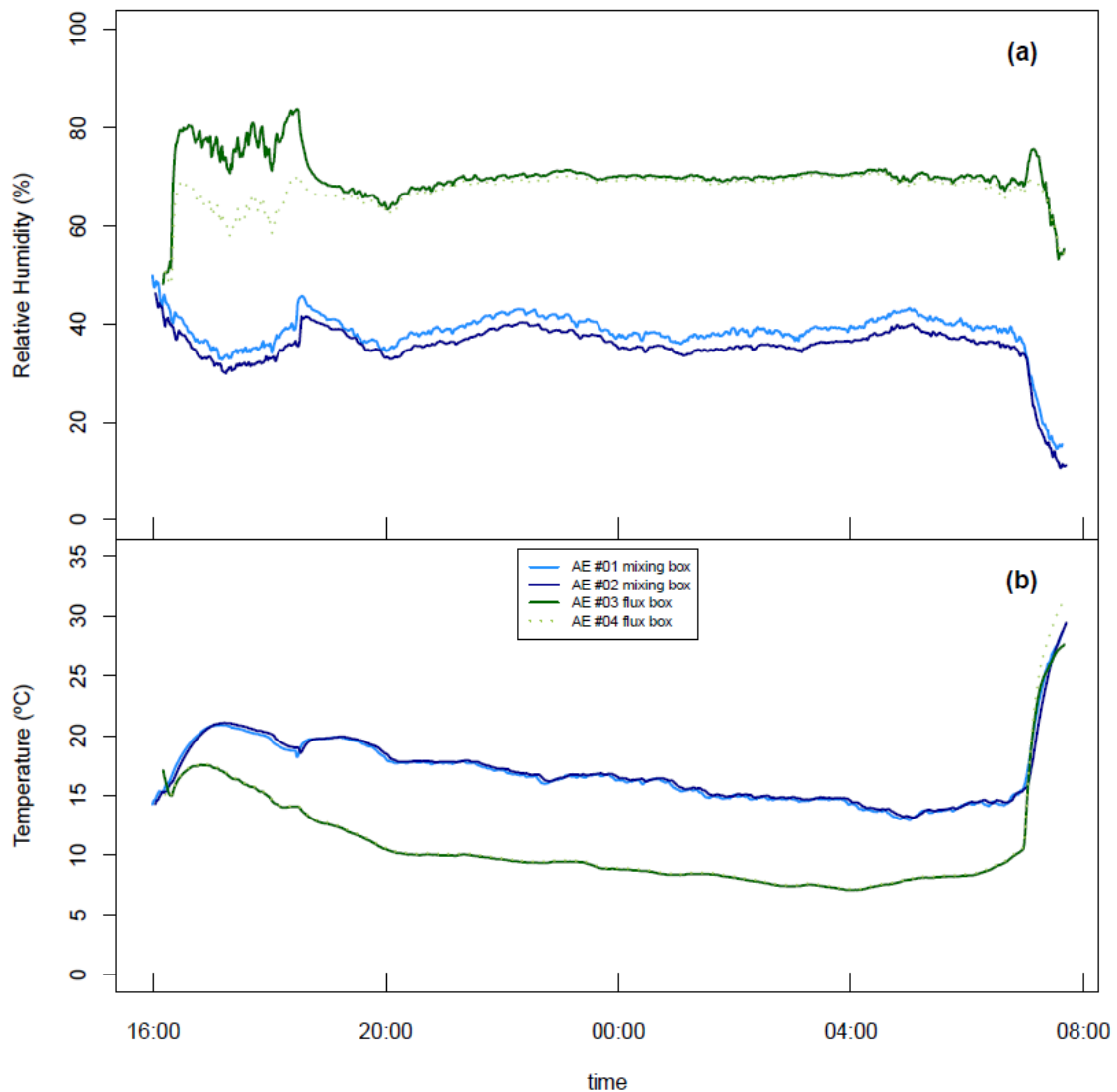
Figure A1: a) %RH from all sensors in the SSTF chamber b) T from all sensors in the SSTF chamber.

3) I appreciate the effort spent to develop a high-quality calibration model. However, in the case of NSS-NTF flux measurements, calibrated data is not strictly necessary as long as the sensor's calibration does not change during the measurement timespan as the flux is determined "differentially" by considering the rate of growth of the CO_2 concentration within the chamber. Could you briefly comment on this in your paper? This could save significant resources for researchers that intend to reproduce your system but do not have the means to perform a comprehensive calibration.

Answer: As commented by the reviewer in the case of NSS-NTF chambers the absolute CO_2 concentration value is not used because the need parameter is the slope of the CO_2 concentration increase during the time interval. Thus it is only important that the calibration factors of the sensors do not change over the timespan of the measurement. We have added this discussion in the paper as explained in detail in the next point.

4) In contrast to the above point, in the case of SS-TF chambers, the effect of calibration is not negligible here. For example, if the sensor in the mixing chamber shows a large bias in its response compared to the sensor in the flux chamber (or vice versa), the bias will induce a similar

bias in the estimated flux. Therefore, it is clear that uncalibrated sensors cannot be used in this configuration. Could you comment on this? Could you suggest a simplified calibration procedure for those who do not have access to calibration facilities? Would a field calibration with a known gas sample be sufficient?

Answer: Considering comments 3 and 4, we decided to expand and to improve our study. Now we apply different calibration approaches and compare the RMSE obtained with them: i) raw data; ii) theoretical correction for RH and P; iii) theoretical correction + bias removal; iV) theoretical correction + simple calibration; v) theoretical correction + multiparametric calibration. The whole chapter 3.1 has been rewritten. The RMSE for all kind of calibrations is now showed in table 2. In figure 5, we also show the fit results from the different calibration approximations. Benefiting of these previous results we have added a section in the results and discussion chapter (chapter 3.3) where we make recommendations on how to proceed with the calibration and recalibration of sensors, differentiating between sensors for SS-TF chambers and for NSS-NTF chambers, and purposing different solutions depending on the laboratories possibilities. We have emphasised and quantified the importance of the bias removal in the SS-TF sensors for the flux calculation, and also quantified the error introduced in the NSS-NTF by not calibrating the sensors.

5) L167 could you provide a summary of the environmental parameters the sensors were exposed to during the calibration cycle? Did you systematically sweep a space of T/RH configurations?

Answer: In lines 117-121 of the manuscripts the intervals are now detailed: *"Both experiments were performed in a temperature range between 20 ºC and 42 ºC and a relative humidity with diurnal cycles between 10% and 50%. Temperature in the calibration box was set to be in increased in slopes of 10ºC, although at low temperatures it fluctuated with room temperature. The pressure ranged between 1004 hPa and 1012 hPa in the calibration at IC3 and between 838 hPa and 850 hPa in the calibration at CRAM."*. In the new Figure 4 the variations in T and RH during the IC3 and CRAM calibrations are shown.

6) L175 could you also show the response of the sensors as a timeseries of relative deviation from the reference provided by the CRDS? Showing this timeseries alongside T and RH would provide some insights regarding their effect on data quality as mentioned in question 3.

Answer: As suggested by the reviewer we have added a plot (Figure 4) with the differences between the CRDS reference and the kit values for simple calibration and multiparametric fitting. In this plot it can be seen that each kit response differently to the variation of T and RH.

7) L183 you report that pressure has the largest effect of the sensor response and that pressure compensation provides the largest reduction in RMSE. Why did you choose to use a linear model for pressure compensation instead of using the quadratic formula suggested by SenseAir in their application note? (see http://www.co2meters.com/Documentation/AppNotes/AN149-Senseair-Pressure-Dependence.pdf)
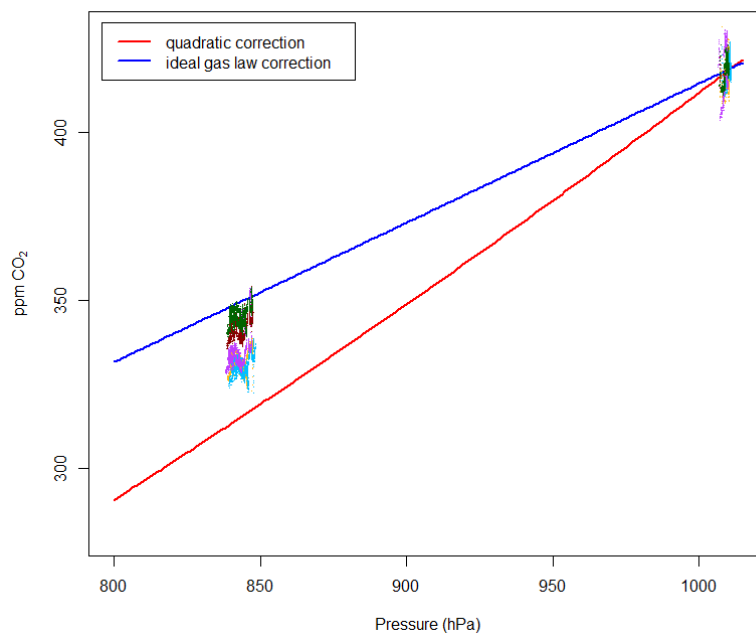
Answer: First of all, we thank the reviewer to indicate us this link. We did not know about this technical sheet. During our study we found out the following documents (https://cdn.shopify.com/s/files/1/0019/5952/files/AN001_Pressure_Compensating_of_a_CO2_Sensor_Rev_1.0_03_May_2021.pdf) where a linear approximation, considering the Ideal Gas Law, was recommended. However, in order to evaluate the reviewer comment, we have made

Pressure tests for all sensors. The plot copied below reports CO_2 sensors data in relation to the environmental pressure during the experiments ($\approx$ 1010 hPa and $\approx$ 845 hPa at IC3 and at CRAM sites, respectively). To skip the CO_2 concentration influence we filtered the data using only those ones when the CO_2 concentration measured by the reference instrument was in the interval 415 ppm - 425 ppm. The plot also shows the quadratic (red line) and lineal (blue line) models calculated using the previous cited references for a concentration of 420 ppm. Bias for CO_2 data is corrected for every kit taking as reference the 1010 hPa data. It can be seen that each sensor seems to have a its own Pressure dependency but, generally, data get closer to the linear ideal gas law fit.

On the light of it, what we did was to apply a three steps correction/calibration:

      a. correction for wet air-dry air;

      b. correction for Pressure influence following the ideal gas law;

      c. calibration of the CO_2 sensor response using a multiparametric fit.

This process has been better explained within the manuscript now.

8) L185 While your statement is true in our experience with the CarboSense network; I think such frequent calibrations could be replaced by in-field calibration for intercept/sensitivity using one respectively two references gases. In the case of NSS-NTF measurements, calibration is not strictly necessary to obtain corrected flux measurements as I pointed out above.

Answer: Authors agree with the reviewer and, actually, a "calibration strategy" paragraph (chapter 3.3) has been now added within the "results and discussion" section where different type of calibrations are recommend in relation to the sensors applications.

9) L202 For the SS-TF measurements you employed multiple sensors on each chamber. Do you think this redundancy could be exploited to improve data quality, for example by averaging the signals or to determine biases?

Answer: Authors agree with the reviewer. Actually, the duplicity of the sensors within the chambers was used for improving the reliability of the measurements, in order to cope with malfunctioning sensors but also for averaging and reduce uncertainty. We have added explanation for this in the revised version of the manuscript.