Anonymous Referee #1

The authors would like to thank the Referee for taking the time to comment on our manuscript. His suggestions and valuable comments gave us the opportunity to improve and clarify the manuscript. The authors' replies (in dark blue) are given below individually with the Referee's comments.

This paper builds upon a number of previous papers of the same group and associated partners (Nelson 2008, Tuzson 2008a and b, Tuzson 2011). To some extent, significant overlap exists with previously published paper by the same group, in particular in the instrument discussion.

Our aim was to highlight our experiences and findings based on the four years of spectroscopic measurement at a remote site and present solutions, improvements that we found necessary and important for long-term and reliable monitoring. We believe that the existing overlap in the instrumental description is well identified and helpful to the reader. Our intention was to focus on relevant new developments and results. Nevertheless, we will strengthen this aspect in our revised paper.

1) In addition to stating the precision (0.02 per mil) in the abstract, the authors should include the replicate precision or accuracy of their measurement. A 10 min averaging time is stated to obtain 0.02 per mil precision.

We added to the abstract the information regarding the reproducibility of our spectroscopic measurements.

Given the series of drift compensation described in the paper, the authors should clarify the complete measurement time and duty cycle of measurement. Although this performance is outstanding it is easily misunderstood that such performance is obtained continuously (100% duty cycle). The statement of being able to acquire with a high time resolution (10 Hz) and stating a precision that is obtained for long averaging times is inappropriate and misleading. Further, the authors mention that the instrument has been deployed since 2008 measuring at 1 Hz. No data is shown for 1 Hz and it is not helpful to state in such way as a measurement cycle is much longer. Mixing in statements of time and precision out of appropriate context or with more specificity should be omitted.

We consider it relevant to state that the instrument is capable to retrieve isotope ratio data with 10 Hz resolution, which is decisive for some applications (e.g. eddy covariance technique). However, for background air monitoring purposes, where the changes of interest are much slower (hourly or diurnal scale), the best achievable precision is more important than the speed of the data acquisition. Therefore, the main objective is to minimize the instrumental drift, keep its operation stable as long as possible, and average the acquired data to improve the measurement precision. In general, as long as the two-sample variance of the measurements is white noise dominated, it is fully appropriate to select any time resolution (*t*) for the measurements to best match the experimental needs. In fact, the full spectral information is available within 0.2 ms (i.e. at a rate of 4.8 kHz). Our effort was focused in extending the stability domain of the instrument, which now allows averaging almost 3 million spectra (600 s) for one data point. For all experiments shown, the raw spectroscopic data and all the other instrumental parameters are recorded at 1 Hz and are averaged in a post-processing step.

The 1 Hz data is shown in our manuscript (see Fig. 2, top graph).

2) It would be helpful to state how long it takes for a flask to be filled.

This information is given in the referenced work by van der Laan-Luijkx et al. (2012), and they state that the flasks are purged for 30 min starting at 7:00 AM.

3) Clarify the temperature stability of the cell and associated limit of precision

This is already given in the manuscript: see pg. 428 lines 17-23.

4) Data processing and Performance: When mentioning the spectroscopic performance the authors state an improvement to the previous arrangement and papers. This is hard to follow given that different averaging times from this with previous papers are compared. Looking at the Allan variance plots and precisions mentioned in Nelson 2008, Tuzson 2008a, b, and Tuzson 2011, it seems that only the 1 s performance has actually degraded from ~0.2 to 0.4 per mil (13C). This paper, shows 0.02 (0.04) per mil for 600 s (400 s) of averaging, while Nelson 2008 shows 0.02 per mil for 400 s, Tuzson 2008a 0.035 per mil for 400 s, while Tuzson 2011 only states the performance of 0.046 per mil at 50 s. This is in contrast to the significant improvement shown in Fig. 3. This reviewer suggests to clarify this by comparing only Allan variance derived precisions for identical averaging times (1s and 400 s for example) and better distinguish the precision for a complete measurement cycle involving drift etc. corrections (Fig.3).

Although the Allan variance is an excellent and adequate parameter characterizing instrumental performance, it still can lead to confusions. Simply using the same averaging times, as suggested, does not resolve the dilemma of comparison. Therefore, we added a paragraph to better explain the significance of the presented improvements.

It is important to realize that an instrument in a well air-conditioned laboratory may allow for long averaging times, while the same analyzer under field conditions (temperature variations, mechanical vibrations, etc.) may suffer from significant drifts, which hinder averaging. In this case, it is appropriate and even necessary to state the Allan variance minimum as optimum integration time, regardless of laboratory results. The 1 s precision can also be used as common referencing value, since at these time-scales the measurements are normally in the white-noise domain.

The most significant improvement (Tuzson et al. 2011 vs. the present manuscript) is the extended white-noise domain allowing for longer averaging times (more than 10 times) and the resulting operation stability, which weakens the requirements for frequent calibrations. For long-term monitoring it is of crucial importance to reduce the calibration gas consumption. In the same line, a lower instrumental drift is beneficial for the data post-processing (corrections and calibrations) and largely compensates for the slightly lower 1 s precision in the present manuscript.

Also, the authors mention a degradation of performance due to the long term operation, but do not offer details. Clearly, the 1 s performance has diminished by a factor of ~2, but overall performance has increased (Fig.3). For laser based systems, spectroscopic degradation can occur by long term misalignment, decrease of transmission through the multi-pass cells (and associated increase of scattering, although pulsed systems may be less affected by this), decrease of spectral purity and power of the QC laser, or degradation of the detectors.

The referee correctly points out that there are many factors that may lead to a degradation of performance over time. We have indeed many elements between the laser and the stored raw spectra that have an impact on the final result. Thus, it would be speculative to point to any single factor (extensively listed by the referee) inducing this change. Nevertheless, we observed a significant aging of the power supply used to drive the laser. Replacing this device almost re-established the previously reported short-term precision performance. In any case, the observed changes in the short-term performance have a negligible effect on the overall precision and we see no reason for further investigations.

Speculating that the QC laser may be emitting with a broader line-width over time may significantly affect the absorption. Based upon the details from Nelson (2008), only 10 % of absorption is yielded in 7.3 m pathlength, which is perhaps a factor of 7 less than with a narrow linewidth laser and suggesting a relatively broad QC laser linewidth.

The laser emission line width did not show any measurable changes during its 5 years of continuous operation. This parameter can easily be estimated during the spectral fitting.

We can't follow the estimation given by the Referee regarding the factor 7 because our calculations, and experimental data with cw-QCLs (laser line width< 10^{-4} cm^{-1}) in a very similar optical setup, show less than 20 % absorption signal. Indeed, with our pulsed QCL we get only 10 % absorption due to the broad (10^{-2} cm^{-1}) laser emission line, which is typical for pulsed QC lasers. This parameter is included into the fitting procedure as an additional Doppler term.

In reference to these comments, it would be helpful for the authors to better tabulate the affects and dependence of precision, drift, and shifts from temperature changes, pressure changes, cal standard drifts, concentration correction and associated determination of CO2 mole fraction. Some details are provided, but are hard to follow. Specifically, it lacks details on how the data is processed (including the many steps of corrections). In particular the 1st and 2nd drift correction is difficult to follow and understand without details.

Untangling the various influencing parameters (gas pressure and temperature, laser, optics and electronics temperature and related drifts) is highly demanding and frequently not possible, mainly because these parameters are often linked and can thus not be observed and analyzed individually (we are looking at very small changes). With respect to the data processing, we have revised the text to give more details where appropriate.

Nomenclature also changes from this to previous papers (cylinder 1, 2, 3 to working, drift, reference, primary etc.) makes it hard to follow if previous papers are referenced for details. A data processing chart in addition to the text description would greatly enhance and add clarity. The correction for the concentration dependence is also difficult to follow.

We agree that the nomenclature for the gases changed, but this is due to the fact that the calibration gases used in the previous publication were replaced and a more elaborate calibration scheme has been developed, which also allows assessing working standards against primary standards for traceability. We believe that giving explicit names to the cylinders (drift, working- and primary-standards) instead of symbolic names (A, B, C) makes the understanding easier. In addition, we briefly summarized in our manuscript the strategy used in the previous setup to facilitate the link between the two configurations (see pg. 429 line 1-17).

Although previous papers from the same group are referenced, Tuzson 2008 (Fig 4 a. and b.) are not well understandable (13CO2 / 12CO2 ratios of ~ 0.8 are what? absorption of respective lines?).

We have added the definition "natural abundance weighted isotopologue ratios R^{13} and R^{18} as retrieved by the spectrometer" which is consistent with the definition given in the referenced paper (Tuzson et al., 2008). These values are calculated by dividing the mixing ratio (in ppm) of the minor (e.g. $[^{13}CO_2]$ or $[^{12}C^{18}O^{16}O]$) and main ($[^{12}CO_2]$) isotopologue determined from the absorption spectra and normalized by their fractional abundances as given in the HITRAN spectral database. The raw spectroscopic ratios are then converted to the international VPDB-CO_2 scale using four-point calibration curve (see Fig. 5 in the manuscript).

5) The FLEXPART section of the paper is somewhat isolated to the instrument performance and improvements discussion. This perhaps could be a separate paper (with fig. 8, 9, 10, 11), yet complements this paper by offering clues on how the acquired dataset can be utilized.

As opposed to increasing the number of publications we opted to combine the analytical aspects with a first approach for the interpretation of the data.

In summary, the paper is by in large well written. However, in current form offers relatively minor value (instrument section) due to a large overlap of content presented in previous papers, and as discussed above lack of clarity, and detail of the discussed advances that would be most useful to understand (without having to read in parallel 4 core reference papers, but then to find many gaps in the actual descriptions therein). The first half of the paper therefore will benefit from incorporating clarification and expansion to minimize the dependence of having to read prior publications.

We acknowledge that the instrumental technique is largely the same as in previous publications. It is exactly for this reason that we stress on the successful efforts towards analytical long-term stability and an elaborate calibration strategy for continuous and reliable monitoring of the tropospheric CO_2 isotopic composition at remote sites. We believe that this is unique and highly beneficial for other research groups. In combination with exemplary data evaluation and modeling, this paper has significant added value to the broader scientific community.

We appreciate the helpful remarks of the referee and will take the opportunity of the final submission to re-phrase and add paragraphs to the manuscript addressing the specific suggestions.