Reply to Referees #1, #2 and #3

We would like to thank all three referees for their very helpful comments and suggestions which will improve our manuscript. We provide here explicit responses to all reviewer comments, and attach a draft revised manuscript incorporating these responses. In the following, the referee comments are in black, followed by our replies in red.

Anonymous Referee #1

Received and published: 24 October 2012
Review: Feasibility study of using a "travelling" CO2 and CH4 instrument to validate continuous in-situ measurement stations

General Comments The paper outlines an interesting and relevant study in which the feasibility of a travelling comparison instrument is investigated. This is a very pertinent paper considering the current rapid expansion in in situ instrumentation and networks and the necessity of ensuring the comparability of the data collected at these stations. The paper is generally well written, however, the abstract and sections 1 and 2 could benefit from some editing. A particular request of the associate editor was suggestions as to ways to shorten the paper. As such, I've listed areas in the "Technical Corrections" section which I feel could and should be streamlined and given some specific examples of how I would do this. Once these small corrections are made I would be happy to recommend it for publication.

Specific Comments The paper presents a study highly relevant to the greenhouse gas (GHG) measurement community. Although not a new concept the application of a Travelling Comparison Instrument to GHG measurements is an exciting development in the effort to assess the comparability of disparate data streams. The paper contains a sufficient level of detail and the study itself appears to have been particularly thorough. The results of the comparison are presented clearly and discussed in detail with adequate references to related work.

Technical corrections Abstract P7142, ln 16 "their flushing pumps" not "there flushing pumps"
Changed

I found the last few sentences (P7142, ln 15-20) of the abstract a little confusing. The paragraph lists two potential sources of differences (Leakages/contamination in the intake lines and/or pumps and insufficient standard gas flushing at Cabauw only). It then says that differences in working standards and drying systems (neither which were listed as a potential source of error) are too small to explain the differences. So what did cause the differences? Did you mean to say? "Offsets arising from differences in the working standard calibrations or leakages/contaminations in the drying systems are too small to explain the observed differences. Hence the most likely causes of these observed differences are leakages or contaminations in the intake lines and/or their flushing pumps. At Cabauw station an additional error contribution originates from insufficient flushing of standard gases."
Your proposal is better, thanks. We changed the succession of the sentences accordingly.

It would be nice to give more details of the locations of the two measurements stations for those not familiar with them, "... two ICOS field stations, Cabauw, the Neatherlands and OPE, France, were compared..." would

be sufficient.
We included the countries.

1.Introduction Remove the sentence starting “Good results from...” (p7144, ln 2-5) the same point is made in the following sentence.
The sentence was removed.

An example of how I would streamline paragraphs 3 to 6 is given below. This reduces the number of words by _20 %:
The compatibility of measurements within a network have been examined in other studies for example, the WMO Round Robins (RR) (Zhou et al.,2011). Although an extended comparison campaign in terms of participating labs these RRs are temporally limited as the gases (prepared and calibrated by the CCLs) are measured at individual stations only once every four years. A campaign of higher frequency (approximately annually) is the CarboEurope IP “cucumber” project (http://cucumbers.uea.ac.uk/) (Manning et al., 2009). It, however, is limited to mainly European stations/laboratories. As RR programs can only compare the precision and the accuracy of cylinder measurements at the sites, observed offsets in RRs cannot be directly transferred to ambient air measurements that may potentially also be affected by the intake system, including pumps and the drying unit. A more comprehensive “end-to-end” comparison exercise is that performed at the GAW site Alert in the high Arctic (Worthy et al., 2012). Here ambient air samples are filled for different laboratories at the same time and compared with each other and with the in-situ measurements at the site. Measurements like these can validate the complete chain from sample collection, analysis and data evaluation, but this type of comparison is not suitable for continuous measurement sites. The atmospheric observational network in the new European ICOS infrastructure (http://www.icos-infrastructure.eu/) will consist of field stations equipped with continuous analysers. This network aims for the highest possible quality and compatibility of measurements. As such a feasibility study on the use of a Travelling Comparison Instrument (TCI) within the network was initiated. This TCI will be set it up at a monitoring station and run in parallel to the existing monitoring system, sampling the same air for a sufficiently long comparison period. It is vital to this quality control (QC) concept that the TCI is a completely independent instrument that measures precisely enough to determine concentration offsets on the order of the WMO ILC targets (i.e. for CO2 at the 0.1 μmolmol−1 level and for CH4 at the level of 2 nmolmol−1). Although common in the reactive gases community (Brunner, 2009), TCI’s are rare for GHG measurement as gas chromatography (GC) the traditional GHG measurement technique has not been considered robust enough for travel. In recent years, however, a new generation of optical techniques like FTIR (Fourier Transform InfraRed) spectrometers or CRDS (Cavity Ring-Down Spectroscopy) analysis have become a standard analysis technique. These approaches are much more robust, easier to use and less demanding in terms of laboratory conditions making them ideal TCIs for GHG comparisons. The in-situ FTIR spectrometer used in the present experiment was calibrated and evaluated against the conventional GC instrumentation in the Institut f Â´I ur Umweltphysik (IUP) carbon cycle laboratory, located in Heidelberg, Germany (Hammer et al., 2012). During summer 2011 the FTIR analyser was used as a TCI at two stations: Cabauw, in the Netherlands and Houdelaincourt (OPE), in France. At both stations and in Heidelberg individual sections of the instrumental setup, including the intake and drying system were assessed. In the following we present the results of these comparison experiments and discuss a possible quality management strategy for in-situ GHG monitoring networks, such as ICOS.
Thanks very much for streamlining these paragraphs. We have included your suggestions and tried to streamline the rest of the manuscript accordingly.

2. Methods and site description P7148, ln 13 Change “smple” to “sample”
Changed
Options for streamlining:

P7145, ln 21 – Replace “travelling comparison instrument” with “TCI”
as you’ve already defined the acronym in the introduction and in the heading above.
Done

P7146, ln 22-24 – Replace “...Heidelberg. For a comprehensive description and performance evaluation of the TCI please refer to Hammer et al. (2012)” with “...Heidelberg (See Hammer et al. (2012)).”
Done

P7146, ln26 to P7147, ln 2 – Replace “...Thus, a reason for using the in-situ FTIR as the TCI was to implement a different analytical technique for comparison. This might prove advantageous in order to detect possible biases that may occur due to the applied analytical technique” with “Hence using an FTIR as the TIC has the advantage of identifying possible biases inherent in the analytical technique”.
Done

3. Experimental
I feel that Table 1 and Figure 1 show the same information. As length is an issue I would remove the table, similarly remove Table 2 and leave Figure 2. If you’d like to give the exact values of the differences then perhaps give them in the text but it is quite easy to estimate them from the Figures.

Anonymous Referee #2
Received and published: 16 November 2012
Review of 'Feasibility study of using a "travelling" CO2 and CH4 instrument to validate continuous in-situ measurement stations' by Hammer et al.

The submitted manuscript presents results from validation experiments at greenhouse gas monitoring sites in France, the Netherlands, and Germany. It provides much useful information and definitely merits publication in Atmospheric Measurement Techniques. It is well written and structured. However, the manuscript is rather long, also a bit wordy and too detailed in some places. To my mind, the main focus of the paper should be on highlighting the general potential of such quality control activities exemplarily corroborated on the basis of the lessons learnt from the two campaigns. Thus, I ask the authors to carefully screen the document (and the tables) if all detailed site-specific information is really needed to come to the conclusions drawn in the manuscript. See also the general comments below for suggestions to shorten the paper.

General comments:
At some places in the manuscript the authors use the bare station name/acronym (e.g. page 7153, lines 22-23: 'At OPE, the intake system test was performed . . .'), at others it is followed by 'station' (e.g. page 7142, lines 16-17: 'At Cabauw station an additional error contribution originates . . .'). Be consistent.
We stick to the bare station name/acronym and removed all instances where those where combined with the word 'station'.

Isn't the official name of the Cabauw research station 'CESAR'?
We decided to stick to the international 3 letter code system for the GHG measurement site since this acronym is widely used in the GHG community

There is an excessive use of acronyms that make the manuscript at some places hard to read. Remove also some acronyms that are introduced but not really used afterwards, such as on page 7143, lines 13 – 14: ' . . . WSs are provided to all field stations by a central laboratory, the Scripps Institution of Oceanography (SIO). Each working standard is calibrated prior to and after usage against the SIO primary calibration . . .'. My suggestion is to remove '(SIO)' and to replace 'SIO' in the second sentence simply by 'Scripps'.
Thanks for your suggestion. We changed the text accordingly.

Same page further below, there is no need to introduce 'RR'.
We have removed the RR abbreviation in the entire manuscript.

I recommend to carefully consider if all tables are really needed, especially since nearly all numbers are also mentioned in the text and Table 4 again contains a good deal of numbers already presented before.
We have removed Tables 1 and 2.

Chapter 5 on quality management hardly refers to the actual site validation made with the travelling instrument. Even if it provides useful information, it could be completely removed and could be considered to be part of another more general paper on quality management in air composition measurements.
We agree that section 5 has a more general character. It was our attempt to suggest a comprehensive quality control approach for continuous stations. This section highlights that a TCI is by far not sufficient and a comprehensive quality management system has thus to include other quality control measures. Since QC is the major topic of this manuscript and we are not aware of any other comprehensive QC discussion in the recent GHG measurement literature we decided to include and keep this general section.

Appendix A is a rather Heidelberg system-specific issue and might also considered to be skipped.
This section has only be included, since we feel the need to demonstrate that the approach of the TCI works at this home base laboratory. Without this prerequisite no conclusions from the comparisons at the field stations can be drawn. This section is already placed in the Appendix and is thus regarded as auxiliary information.

Specific comments:
Page 7142, lines 8 – 9: 'the Heidelberg gas chromatographic system . . .' sounds somehow odd. Why not simply: 'Observed differences between the TCI and a gas chromatographic system . . .'
Changed accordingly

Page 7142, line 12: explain acronym 'OPE'.
Done

Page 7144, line 12: what are 'end-to-end' QC measurements? Explain acronym 'QC'.
This sentence has been rephrased. It now reads: 'end-to-end' comparisons

Page 7144, line 27: add a reference that it was also agreed during the last Meetings of Experts on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques that parallel measurements with a travelling instrument during audits of the World Calibration Centre are preferable. See e.g. WMO-GAW Report #194.
We included that the WCC is encouraged to use a TCI and to test the sample intake system as it is stated in the proceedings of the GGMT 2011 conference.

Page 7147, lines 12 – 13: 'Remove 'and the observatory takes part in the ICOS Demo Experiment.' Move 'operated by ECN' to the previous sentence.
Done

Page 7148, lines 9 – 10: what is a level 1 site? Remove the sentence. The ICOS Demo Experiment is already mentioned above.
We removed the sentence

Page 7148, lines 19 – 21: How is the ambient air pressurized?
The sample is pressurized using a diaphragm pump manufactured by KNF Neuberger, Germany.

Page 7148, line 22: what is an 'ICOS integrated demonstrator prototype'? Is it simply the G1301 from Picarro?
This is the terminology used in ICOS for this entire setup. We replaced it in the manuscript by "This system"

Page 7148, lines 23 – 25: Do you show some results of the second Picarro analyser? If not, skip this sentence.
Agreed, we skipped this sentence. In the original manuscript we compared our results to both Picarro analysers but this made the manuscript even longer.

Page 7149, line 27 f.: Is the FTIR non-linear above 2500 nmol mol-1 CH4?
It is not necessarily the FTIR which is non-linear above 2500nmol mol-1. As we state in the manuscript we suggest a small concentration dependent bias in the TCI CH4 calibration. This would explain the deviation of the 2600 nmol mol-1 cylinder in CBW as well.
We have included a comment that the deviation of the large CH4 concentration at CBW is consistent with a concentration dependent bias in the TCI calibration.

Page 7150, line 21: I suggest covering all disturbing effects (contamination, leaks, but also depletion or memory effects) by the term 'artefacts'.
Agreed

Page 7153, lines 2 – 4: I don't understand the link between the length of the tubing and the quality of the connection.
Yes this sentence is unclear and does not provide useful information. We thus removed it from the manuscript.
FYI: The link between the length of the tubing and the quality of the connection was given by the fact that the sample intake lines, the gas supply as well as an additional piece of tubing, acting as an open end, were directly inserted in the polyethylene/alu bag through one opening. The open end was only inserted a few cm while the sample lines should normally reach 1m into the bag. In CBW this was not possible. Since the open end as well as imperfect sealing of the bag allows back diffusion the localization of the sample intake line in the bag becomes important.

Page 7155, lines 2 – 3: Skip sentence 'Ambient mole fractions varied . . .'. Sentence doesn't provide any additional information. The ranges are easily observable in the figure itself. Same for the OPE (page 7157, lines 2 – 4).
Done

Page 7156, line 3: what is a target gas?
We replaced all instances of target gas with target/surveillance gas

Page 7157, line 23: the temperature fluctuations of 10K: is that the observed day-nightdifference?
No, the only location available for the TCI was under the air condition directly beside the container door. So the 10 K is opening the door on a hot day and cooling the container down afterwards.

Paragraph 3.4: This part is particularly wordy and has some potential to be shortened (especially the first paragraph of 3.4). Why not simply showing the histograms and stating the 75%-25% percentile difference is used as a quantity to evaluate the compatibility.
We streamlined the first paragraph of 3.4, by removing the general statistic discussion. However, we refrain from skipping the entire paragraph since it highlights the issue of non-Gaussian statistics, which, although common in atmospheric measurement data, is rarely discussed.

Page 7158, line 18: Aren't these simple histogram plots. Why not calling so?
We changed the term frequency distribution to histogram.

Page 7164, line 10: replace 'impressively' by a more scientific term.
Changed to: The high precision of the two optical instruments used in the comparison, i.e. the FTIR and the CRDS, allowed us to

Page 7165, line 4: What's the role of the entity called ICOS Atmospheric Thematic Centre?
It is responsible for processing the continuous data from the entire ICOS atm. station network. We added a respective sentence

Page 7165, lines 13 – 15: Sorry, I do not get this sentence. Do you speak about a separate inlet line at the home base for quality controlling the travelling instrument before and after the audit?
Yes, the TCI has to demonstrate at its home base that it is precise and accurate enough to fulfill the job. We did this in Heidelberg prior, in between and after the two campaigns, however due to structural conditions the sampling line of the TCI and the GC have been the same. This should be different for the future TCI home base lab.

Figure 6: The standard deviations are much larger for Heidelberg than for the actual audited sites. Is the GC system – especially considering the inferior precision for CH4 in comparison to the novel measurement techniques – a suitable system to act as a reference to quality control the FTIR?
Yes, since it is not the scatter which defines the inter station compatibility. Although the scatter is large the agreement between the TCI and the GC is well within the WMO-ILC. To judge this, one has to

compare to the standard error of the mean instead of the standard deviation. Of our comparisons which consist of thousands of individual comparisons the standard error of the mean becomes very small.

Page 7166, lines 2 – 3: this is a statement made by the Greenhouse Gas Experts Meeting and should be properly referenced. (see e.g. GAW Report #206).
Thanks, included.

Minor comments:
Page 7142, line 16: typo: ‘there’ -> ‘their’
Corrected
Page 7148, line 13: typo: ‘sample’
Corrected
Page 7164, line 8: ‘detect’ seems to be printed in italics. Why?
Removed
Page 7170, line 15: typo: ‘snap-shot like’
Corrected

Anonymous Referee #3
Received and published: 16 November 2012

General comments:
The paper 'Feasibility study of using a "travelling" CO2 and CH4 instrument to validate continuous in-situ measurement stations' by S. Hammer et al. describes the results of an intercomparison study between a travelling instrument and the continuous atmospheric measurements at two European locations. The paper is well structured and although long it reads well. The concept of a travelling instrument for CO2 and CH4 measurements is very interesting and new and especially the results of the two first comparisons are important new information. I would recommend publication of this paper.

The concept of travelling audit systems, such as the system described in the paper is not new. WMO audits such as performed by EMPA are a comparable tool to assess the quality of atmospheric measurements. It might be good to include a reference to those audits.
We included the reference to the "travelling standard" audits run by the WCC.

The title of the paper includes the term 'feasibility study'. However, only the feasibility of the measurements themselves is described and explained. It would be good to add information on the feasibility in terms of logistics, finances and available time and people. It might be helpful to include information on the amount of time it takes to perform such a campaign, including preparations and data analysis. Is it feasible to perform many campaigns per year? Do the authors expect that this instrument will be used frequently in the way as described in their paper? Are new campaigns planned? Will e.g. the ICOS project include quality control measures such as the use of a TCI?

Some of these points have been addressed in the first paragraph of section 4.4. We have extended this paragraph and added an assessment on needed manpower and financial requirements. However we refrain from giving any detailed numbers since these will change from country to country and in time.
Also we think that a scientific journal is not the appropriate platform to balance costs versus additional scientific gain.
There is one additional campaign planned, but we do not see the need to include this information in the manuscript.
ICOS plans to include the TCI in their quality management system.

Specific comments and technical corrections:
Page 7142 line 12: OPE has not been explained yet.
We now introduce the abbreviation in the abstract.

Page 7142 line 17: Insufficient flushing of standard gases: does this apply to the TCI or the in-situ instrument?
We included a specific reference to the CBW instrument.

Page 7143 line 10: please explain AGAGE acronym.
Done

Page 7144 line 13: please explain QC.
Done

Page 7144 line 20 and 26 and 7145 line 5 and 21: replace Travelling comparison instrument by TCL.
Done. We have now two places where the abbreviation TCI is introduced, once in the abstract and once in the main text.

Page 7145 last paragraph of section 1 has some overlap with section 2.1, text can be shortened here.
Section 1 has been streamlined and rephrased.

Page 7146 line 15: replace gas chromatographic system with GC.
Done

Page 7148 line 10: please explain 'level 1 site'.
We removed the term 'level 1' site

Page 7148 line 13: replace smple with sample.
Done

Page 7152 line 12: what do the authors mean with similar results?
We replaced this by:
The CH4 results led to the same conclusions,…

Section 3.2: the SIS test at OPE gives a much better result than at Cabauw. However, at Cabauw the inlet lines are included in the test and at OPE they are skipped. In section 3.2.2. the authors suggest a relation between the length of the tubing and the results. Would the results of the SIS test at OPE have been worse when the tubing was included, or would the results at Cabauw have been better if they were not included?
Unfortunately, this is the open question of the SIS test. From the available data we cannot isolate the origin of the discrepancies further. Future comparison campaigns should investigate this in more detail and as we state in the manuscript these tests have to be evaluated while being on site. This is the only possibility to fix this issue.

Page 7153 line 15-16: the authors suggest an additional test, is this planned?
No, currently it is not planned to return the TCI to CBW. It is however possible to examine the potential effect of the intake system with the instrumentation available at CBW. Measuring a tank with and without the different parts of the intake system will provide additional information. It is in the responsibility of the station PI to follow up on this issue.

Figure 4: why do the plots of the differences (middle panels) have shorter time series/ do they start later?
There have been known issues with the sample intake system of the OPE CRDS. Thus the data had been excluded from the comparison. We did mention this in the text, please see 3.3.2.
We have now included an additional note in the figure legend.

Page 7157 line 3: what do the authors want to add by saying: 'mainly driven by the diurnal cycle'? Maybe this sentence can be skipped.
Yes, thanks we skipped the entire sentence.

Figure 5: the CH4 difference at Heidelberg has a different scale than for OPE and CBW (-10 to +10 vs -2 to +2). Using the same scale could be helpful to have a more clear idea of the different results at the different locations.
Expanding the CH4 scale for OPE and CBW would make it impossible to see the difference between the histogram and the fitted Gaussian distribution.
We have instead included a remark in the Figure legend highlighting the different scale.

Section 3.4: the authors explain a lot of general concepts in statistics here. Although it's helpful for readability, this is a section that can be shortened. Perhaps it's good to find a citation to explain the statistical procedures used, such as Inter Quartile Range?
We have shorted this section and included a reference as you propose.

Page 7161 line 13: the system does not reach a stable value within the given time frame. Does this problem occur also in the regular measurements, and to what extent does this affect them?

This problem affected the measurements of the working standards at CBW station as well. Thus all measurements evaluated using these WS are affected. For CO_2 the effect was on the order of 0.05 to 0.10 µmol mol^{-1}.

Page 7162 lines 9-11: another SIS test is suggested here without the intake lines.
Could another test be performed with the CBW system only to check for problems with the intake lines? (See also my comment on Section 3.2).
Sure, see our comments above.

Page 7165 line 20: the TCI approach could in principle also be used to check flask sampling locations. Do the authors see any benefits there?
In principle yes, but this seems to be a bit oversized. For flasks one does not need the high temporal resolution of the TCI. Co-located flask sampling might be sufficient and much cheaper. Performing a modified SIS test for flask samples seems however, appropriate and necessary.

Section 5.2: the suggestions made by the authors to optimize quality control of atmospheric measurements are very thorough and well thought of. However, they are also labor intensive and costly. This approach can most likely not be applied at most stations. Can the authors add something about the feasibility of their suggestions? And if the researchers of a station want to 'upgrade' their quality control strategy, but cannot fulfill all suggestions, which do the authors think is the most useful, i.e. gives the most additional information versus the lowest costs? Is a quality control strategy such as suggested here part of the ICOS program in which OPE is included?

We think a high frequency target/surveillance tank provides definitely the most information for the lowest costs. However, this would not capture any artifacts related to the air intake system. Thus we recommend having regular flask- in situ comparison in addition. We have added a respective sentence in the manuscript.