

Feasibility study of using a “travelling” CO₂ and CH₄ instrument to validate continuous in situ measurement stations

S. Hammer¹, G. Konrad¹, A. T. Vermeulen², O. Laurent³, M. Delmotte³,
A. Jordan⁴, L. Hazan³, S. Conil⁵ and I. Levin¹

[1]{Institut für Umweltphysik (IUP), Heidelberg University, Germany}

[2]{Energy research Center of the Netherlands (ECN), Petten, The Netherlands}

[3]{Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Gif-sur-Yvette, France}

[4]{Max - Planck Institute for Biogeochemistry, Jena, Germany}

[5]{Agence Nationale pour la gestion des Déchets Radioactifs (ANDRA), Bure, France}

Correspondence to: S. Hammer (shammer@iup.uni-heidelberg.de)

Abstract

In the course of the ICOS (Integrated Carbon Observation System) Demonstration Experiment a feasibility study on the usefulness of a Travelling Comparison Instrument (TCI) was conducted in order to evaluate continuous atmospheric CO₂ and CH₄ measurements at two European stations. The aim of the TCI is to independently measure ambient air in parallel to the standard station instrumentation, thus providing a comprehensive comparison that includes the sample intake system, the instrument itself as well as its calibration and data evaluation. Observed differences between the TCI and a gas chromatographic system, which acted as a reference for the TCI, were $-0.02 \pm 0.08 \mu\text{mol mol}^{-1}$ for CO₂ and $-0.3 \pm 2.3 \text{ nmol mol}^{-1}$ for CH₄. Over a period of two weeks each, the continuous CO₂ and CH₄ measurements at two ICOS field stations, Cabauw (CBW), the Netherlands and Houdelaincourt (Observatoire Pérenne de l'Environnement (OPE)), France, were compared to co-located TCI measurements. At Cabauw mean differences of $0.21 \pm 0.06 \mu\text{mol mol}^{-1}$ for CO₂ and $0.41 \pm 0.50 \text{ nmol mol}^{-1}$ for CH₄ were found. For OPE the mean differences were $0.13 \pm 0.07 \mu\text{mol mol}^{-1}$ for CO₂ and $0.44 \pm 0.36 \text{ nmol mol}^{-1}$ for CH₄. 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. For the Cabauw instrument an additional error contribution originates from insufficient flushing of standard gases. Finally a comprehensive quality management strategy for atmospheric monitoring networks is proposed.

1 Introduction

Long-term atmospheric observations of greenhouse gases (GHG) provide the backbone of our current understanding of global and regional GHGs budgets and their changes. This approach requires combining measurements from distributed stations which, in many cases, are run by independent laboratories with various instrumentation and methodologies. To ensure data compatibility the GHG measurement community agreed on relating their calibration standards to international scales (WMO, 2011), which are produced and maintained by dedicated WMO (World Meteorological Organization) Central Calibration Laboratories (CCL). Moreover, to make best use of the global atmospheric GHGs measurements and allow for meaningful source/sink estimates the WMO experts have set Inter Laboratory Compatibility (ILC) targets for each individual GHG species (WMO, 2011).

There are currently two calibration strategies implemented in the atmospheric measurement community: (1) The centralised Working Standard (WS) calibration approach that is maintained e.g. in the AGAGE (Advanced Global Atmospheric Gases Experiment) project, and (2) the individual WS calibration approach generally performed by WMO-GAW (Global Atmosphere Watch) stations. Within the AGAGE network, WSs are provided to all field stations by a central laboratory, the Scripps Institution of Oceanography. Each working standard is calibrated prior to and after usage against the Scripps primary calibration scale (Prinn et al., 2000). This centralised WS calibration approach avoids multiple calibration chains, e.g. at each field station, and ensures good data compatibility within the network. The individual WS calibration approach is often found in networks that constitute nationally operated stations which have all been set up independently. Traditionally, each laboratory maintains its own calibration chain to the WMO scales by hosting, ideally, several secondary laboratory standards provided by the CCLs. This individual calibration approach is recommended by WMO for the GAW stations. However, the individual WS calibration approach requires multiple scale propagations that may introduce inter-station biases.

The compatibility of measurements within a network have been examined in other studies for example, the WMO Round Robins (Zhou et al., 2011). Although an extended comparison campaign in terms of participating labs, these round robins are temporally

1 limited as the gases (prepared and calibrated by the CCLs) are measured at individual
2 stations only once every four years. A comparison of higher frequency (approximately
3 annually) is the CarboEurope-IP “cucumber” project (<http://cucumbers.uea.ac.uk/>)
4 (Manning et al., 2009). This, however, is limited to mainly European stations/laboratories.
5 As round robin programs can only compare the precision and the accuracy of cylinder
6 measurements at the sites, observed offsets in those round robins cannot be directly
7 transferred to ambient air measurements that may potentially also be affected by the intake
8 system, including pumps and the drying unit. A more comprehensive “end-to-end”
9 comparison exercise is that performed at the GAW site Alert in the high Arctic (Worthy et
10 al., 2012). Here ambient air samples are filled for different laboratories at the same time
11 and compared with each other and with the in-situ measurements at the site. Measurements
12 like these can validate the complete chain from sample collection, analysis and data
13 evaluation, but this type of comparison is not suitable for comparing continuous
14 measurement sites.

15 The atmospheric observational network in the new European ICOS infrastructure
16 (<http://www.icos-infrastructure.eu/>) will consist of field stations equipped with continuous
17 analysers. This network aims for the highest possible quality and compatibility of
18 measurements. As such, a feasibility study on the use of a Travelling Comparison
19 Instrument (TCI) within the network was initiated. This TCI will be set up at a monitoring
20 station and run in parallel to the existing monitoring system, sampling the same air for a
21 sufficiently long comparison period. It is vital to this quality control (QC) concept that the
22 TCI is a completely independent instrument that measures precisely enough to determine
23 concentration offsets on the order of the WMO ILC targets (i.e. for CO₂ at the 0.1 μmol
24 mol⁻¹ level and for CH₄ at the level of 2 nmol mol⁻¹). Although common in the reactive
25 gases community (Brunner, 2009), TCI’s are rare for GHG measurement as gas
26 chromatography (GC) the traditional GHG measurement technique has not been considered
27 robust enough for travel. In recent years, however, a new generation of optical techniques
28 like FTIR (Fourier Transform InfraRed) spectrometers or CRDS (Cavity Ring-Down
29 Spectroscopy) analysers have become a standard analysis technique. These approaches are
30 much more robust, easier to use and less demanding in terms of laboratory conditions,
31 making them ideal TCIs for GHG comparisons. The 15th WMO/IAEA Meeting on Carbon
32 Dioxide, Other Greenhouse Gases, and Related Measurement Techniques, encourages the
33 use of a TCI for the World Calibration Centre (WCC) for CH₄ and CO, and to include the

1 sample intake system in the audits (WMO, 2011). The WCC has so far audited WMO-
2 GAW station using travelling standards as described in Zellweger et al. (2011). At the 16th
3 WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related
4 Measurement Techniques the WCC presented first promising results of combining the
5 established travelling standard audit system with a travelling instrument.

6 The in-situ FTIR spectrometer used in the present experiment was calibrated and evaluated
7 against the conventional GC instrumentation in the Institut für Umweltphysik (IUP) carbon
8 cycle laboratory, located in Heidelberg, Germany (Hammer et al., 2012). During summer
9 2011 the FTIR analyser was then used as a TCI at two stations: Cabauw, in the Netherlands
10 and Houdelaincourt (OPE), in France. At both stations and in Heidelberg individual
11 sections of the instrumental setup, including the intake and drying system were assessed. In
12 the following we present the results of these comparison experiments and discuss a possible
13 quality management strategy for in-situ GHG monitoring networks, such as ICOS.

2 Methods and site description

2.1 The Travelling Comparison Instrument (TCI)

We used an in situ multi-species FTIR analyser (built by the University of Wollongong, Australia, (Griffith et al., 2012) as **TCI**. Prior to its usage in this study, the TCI setup was improved and extensively tested in Heidelberg (see Hammer et al., 2012). The ICOS **Demonstration** Experiment (and planned ICOS station network) uses CRDS analysers to continuously monitor CO₂ and CH₄ at the field stations. **Hence the use of an FTIR as the TCI has the advantage that it may identify possible biases inherent in the analytical technique.** In addition, the FTIR is capable of measuring CO₂, CH₄, CO and N₂O simultaneously with a precision to fulfil the WMO ILC targets (Hammer et al., 2012). The in situ FTIR is sufficiently compact and robust to be transported by car and easily moved by two persons. Setting up the FTIR takes about four hours and the instrument stabilises and reaches its full precision after a 12 hour settling-in period.

2.2 Site description and site instrumentation

2.2.1 Heidelberg (HEI)

The Heidelberg observational site (49°24'N, 8°42'E, 116 m a.s.l., approx. 130 000 inhabitants) is located in the upper Rhine valley, a polluted region in south-western Germany. The air sampling intake is installed on the top of the Institut für Umweltphysik building located in the western outskirts of the city. Due to the proximity to GHG sources, large concentration variations are observed frequently (Levin et al., 2010). The station is equipped with a **GC** which was designed for simultaneous analysis of CO₂, CH₄, N₂O, SF₆, CO and H₂. The 1/2" stainless steel intake lines, one in the south eastern and one in the south-western corner of the building, are mounted approx. 30 m above local ground, are permanently flushed and are each sampled every 15 minutes. The reproducibility of the individual measurements is $\pm 0.05 \mu\text{mol mol}^{-1}$ for CO₂ and $\pm 2.4 \text{ nmol mol}^{-1}$ for CH₄. A detailed description of the entire GC system can be found in Hammer et al. (2008). Due to structural conditions, no independent intake line could be deployed for the TCI in Heidelberg. Thus, the south eastern intake line and its line flushing pump were used as intake for the TCI in parallel to the GC, but sample pumps and drying systems (GC: cryo cooler -45 °C, TCI: Nafion and

Mg(ClO₄)₂) were still completely separated. To allow for better comparability between the continuous TCI and the discrete GC measurements, a buffer volume was installed in the GC sample line. The buffer volume allows capturing the short-term concentration variations in between the discontinuous GC measurements. Details of the setup are given in Appendix A.

2.2.2 Cabauw (CBW):

Cabauw tower (51°58' N, 4°55' E, -0.7 m a.s.l.) is a steel construction that rises up to 213 m a.g.l.. It is located in the centre of The Netherlands, about 25 km southwest of the city of Utrecht. The direct surroundings of the tower are very flat and homogeneous and have a relatively low population density, although the area within 100 km of the tower houses more than 7 million people. The main land use of the area around Cabauw is a mixture of intensively and extensively managed grassland. Cabauw tower was set up in 1972 as an observational site for boundary layer meteorological observations. Greenhouse gas concentration observations at several levels of the tower started in 1992 and are operated by ECN.

Ambient air is drawn from four heights through Synflex 1300 12 mm o.d. tubing at flow rates of about 12 l/min down to the cellar where the analysers are installed. The sample air is pre-dried at the inlets to a dew-point of about -10 °C to prevent water condensation in the inlet lines using Nafion membrane dryers (Permapure PP-625-72). Two spare air sample lines are installed at the 200 m intake level. From the main sample stream a small by-pass of 500 ml/min for analysis is dried further using cryogenic freeze traps at -50 °C. During the experiment sample air was analysed by an Agilent 6890 GC for CH₄, CO, SF₆ and N₂O, a Licor 7000 NDIR was run for CO₂ analysis and a CRDS (Picarro G2301) for CH₄ and CO₂ measurements. The reproducibility of the CRDS used for the ICOS Demonstration Experiment is 0.04 μmol mol⁻¹ for CO₂ and 0.4 nmol mol⁻¹ for CH₄. The sampling system and analysis setup at Cabauw is described in detail by Vermeulen et al. (2011).

2.2.3 Houdelaincourt, Observatoire Pérenne de l'Environnement (OPE):

OPE (Observatoire Pérenne de l'Environnement) is a long-term multi-disciplinary observational site located near Houdelaincourt (48°33' N, 05°30' E, 392m a.s.l.) in a low population countryside in north-eastern France. This observatory owned and operated by the Agence Nationale pour la gestion des Déchets Radioactifs (ANDRA) is dedicated to monitor the full environmental parameters in the air, soil, river waters, fauna and vegetation. Within this framework, an atmospheric monitoring station equipped with a 120 m tower has been set

up in 2011, in collaboration with the Laboratoire des Sciences du Climat et de l'Environnement (LSCE). Its instrumentation consists of a CRDS analyser (Picarro, G1301) and an off-axis Integrated Cavity Output Spectrometer (Los Gatos, DLT100) for continuous monitoring of CO₂, CH₄, CO and N₂O at 10 m, 50 m and 120 m a.g.l. The sample tubing is made of ½" Synflex 1300. At each sampling height, meteorological sensors (temperature, pressure, wind speed and direction) are installed.

The Picarro G1301 (CO₂, CH₄) and the Los Gatos DLT100 (N₂O, CO) are both integrated in an automated and remotely controlled system of air sampling, conditioning and distribution (called the "ICOS integrated demonstrator prototype" (Delmotte et al., 2011)). The ambient air analysed by both instruments is dried by cryogenic cooling to -45 °C and is slightly pressurised (<100 mbar above atmospheric pressure, using a membrane pump (KNF Neuberger, model: N 86 ATE) to maintain identical inlet conditions for all instruments and sample types (i.e. gas from high pressure cylinders or ambient air). The reproducibility of this measurement system is 0.04 µmol mol⁻¹ for CO₂ and 0.19 nmol mol⁻¹ for CH₄.

3 Experimental results

In order to be able to interpret the comparison results of the parallel ambient air measurements in more detail and to relate potential differences to individual components of the instruments, a series of dedicated experiments have been conducted. First, the calibration gases were compared, then detailed investigations of the entire sample intake systems were performed, and finally the instruments have analysed atmospheric air in parallel for about two weeks at each ICOS Demonstration station.

3.1 Comparison of calibration standards

Comparing the calibration standards of the TCI and the ICOS stations is the most fundamental step in the evaluation chain. In order to minimise potential errors originating from the propagation of the international scales to the working standards, the ICOS stations and the TCI received a dedicated set of working standards (WSs) for the ICOS Demonstration Experiment from the designated ICOS Central Analytical Laboratory (CAL, hosted by MPI-BGC in Jena). This centralised WS calibration approach used in ICOS ensures calibration comparability from the beginning. Thus, the following calibration test assesses the internal consistency of the centralised WS calibration approach. This test does not reveal bias from the WMO-CCL which is independently examined by regular comparison exercises between the ICOS CAL and the CCL. However, at CBW the dedicated ICOS Demonstration Experiment WSs were not in place; instead an earlier set of WSs was used that had been also calibrated by the CAL but with other instrumentation and with less precision.

Figure 1 shows the difference between the CAL-assigned concentrations and the TCI measurements for the individual working standards used at the two ICOS stations. At both stations, the assigned concentrations of all working standards are reproduced well by the TCI within the ILC target range of WMO ($0.1 \mu\text{mol mol}^{-1}$ for CO_2 and 2 nmol mol^{-1} for CH_4), apart from one CBW working standard with a CH_4 mole fraction of more than $2600 \text{ nmol mol}^{-1}$. The calibrated CH_4 range of the TCI extends only to $2490 \text{ nmol mol}^{-1}$. The observed 1σ standard deviation of the differences is in accordance with the CO_2 and CH_4 reproducibility of the TCI (Hammer et al., 2012). The relatively large error bars for the CH_4 differences at CBW are caused by a larger uncertainty of the assigned concentrations. At OPE, the CH_4 calibration comparison exhibits a small concentration dependent difference of $0.1 \text{ nmol mol}^{-1}$ per 100

nmol mol⁻¹. This concentration dependence of CH₄ differences could also explain the observed deviation of 2.2 nmol mol⁻¹ for the highest CBW WS. Thus, a small concentration dependent bias in the CH₄ calibration of the TCI seems to be likely. The blue stars in Figure 1 represent the difference between the CAL assignment and the TCI measurement for the two TCI surveillance/target cylinders regularly used as a repeatability check. The surveillance/target cylinder measurements agree at both stations and in Heidelberg (not shown), within an overall 1σ standard deviation of 0.06 μmol mol⁻¹ for CO₂ and 0.3 nmol mol⁻¹ for CH₄ over the entire 5 months of the ICOS Demonstration Experiment. A detailed presentation of the TCI quality control record, including the station visits at CBW and OPE, can be found in the companion paper by Hammer et al. (2012).

3.2 Sample intake system test (SIS test)

The agreement of the WSs and thus the calibration reference when analysed with one instrument (the TCI) is essential but not in itself sufficient to guarantee compatibility between ambient air datasets from different stations. The ambient air intake system comprises additional hardware, e.g. sample intake tubing, pumps, drying systems etc., which are potential sources for artefacts. At most stations, the sample intake system is not used for routine cylinder analysis and is also not tested regularly by other measures. To overcome this limitation, the entire sampling path of the ambient air was inspected by the sample intake system test (SIS test) where air from a high pressure cylinder is “sampled” via the ambient air inlet. Performing this test simultaneously for the routine atmospheric monitoring intake and the TCI intake allows for direct comparison of the two independent sample intake systems. Additionally, analysing the same high pressure cylinder at each instrument without the inlet system allows quantification of possible intake system offsets. During the test, the ambient air inlets were connected to a large, elastic buffer volume, which was constantly flushed with air from the high pressure cylinder. A 100 litre polyethylene-coated aluminium bag was used as a buffer volume. For more details on the material used and its stability for GHG concentrations we refer to Vogel et al. (2011). The elasticity of the buffer volume ensured that atmospheric pressure conditions were present at the inlet. The humidity of the sample itself was however not comparable to ambient air conditions since pre-dried air from a cylinder was used.

At most stations the sampling lines are constantly flushed at a high flow rate, i.e. several tens of litres per minute, to minimise the residence time of the sample in the inlet line. The actual sample is then tapped off and directed to the analyser using a smaller sample pump. High flow rates constitute a principal problem to the SIS- test. Large drainage rates from a high pressure cylinder cause expansion cooling at the regulator of up to 10 K. This cooling can lead to small concentration changes in the decanted gas, most noticeable for CO₂ and N₂O. We quantified this drainage effect to be $-0.005 \pm 0.001 \text{ } \mu\text{mol mol}^{-1} \text{ slpm}^{-1}$ for CO₂ and $0.010 \pm 0.007 \text{ nmol mol}^{-1} \text{ slpm}^{-1}$ for CH₄. If these observed concentration changes are caused by thermal fractionation as described by Keeling et al. (2007) or whether they result from sorption effects due to different surface conditions on the cold regulator walls is not clear. However, it is a known and now quantified phenomenon which needs to be considered for large cylinder drainage rates. To minimise the cylinder drainage effect (and save gas for analysis), the main line flushing rates had to be reduced (CBW) or line flushing pumps had to be switched off (OPE) during the sample intake test and the intake lines were flushed by the smaller sample pumps only. For all SIS tests the cylinder drainage rate varied between 3 slpm (HEI) and 7 slpm (OPE). The conditions during the SIS tests were thus not entirely comparable to ambient sampling conditions but the tests allow for detecting contamination or leaks in the inlet systems and examine the entire sample path.

3.2.1 Sample intake system test Heidelberg

Figure 2a and 2b show the results of the sample intake test performed in Heidelberg. Open symbols represent the measurements via the sample intake systems. Filled symbols denote the average and the standard deviation of the direct measurements of the same high pressure cylinder. For CO₂ the test measurements of both instruments, the TCI and the GC agree within $0.02 \pm 0.05 \text{ } \mu\text{mol mol}^{-1}$. Moreover, for both instruments, the measurements via the sample intake system agree with the direct measurement of the cylinder. From the latter we can conclude that the Heidelberg intake systems do not introduce significant artefacts. The CH₄ results of both instruments led to the same conclusions, although the GC precision for CH₄ is much worse compared to the TCI (see Figure 2b).

3.2.2 Sample intake system test CBW

Since the CBW tower is equipped with an elevator that provides easy access to the air intakes, it was possible to attach the buffer volume directly to the inlets of the 200m sampling lines. For the SIS test the line flushing rates of the main pumps were reduced to less than 1 slpm by adjusting the flow controller (CRDS) or closing the respective needle valve (TCI) in the intake system, however the line flushing pumps were kept running. Figure 2c and 2d show very clearly that the SIS test did fail for the CRDS instrument. CO₂ measured by the CRDS is lower by 3 to 4 $\mu\text{mol mol}^{-1}$ compared to the direct measurement and shows large variability of $\pm 0.6 \mu\text{mol mol}^{-1}$. CH₄ at the same time makes step changes with a standard deviation of $\pm 0.6 \text{ nmol mol}^{-1}$. These differences are too large to be explained by contamination or leaks in the intake system itself and are not in accordance with the ambient air comparison results for CBW (see below). Unfortunately, data evaluation was performed only after the campaign so the test could not be repeated. The most likely explanation for the failure is a leakage in the connection of the CRDS sampling line to the buffer volume.

The simultaneously recorded TCI results agree much better to the direct tank measurements, indicating that the experimental setup worked as expected for the TCI. The CO₂ measurement via the intake system and the direct cylinder measurement differ by $0.09 \pm 0.03 \mu\text{mol mol}^{-1}$. This difference is too large and of opposite sign to be explained by the cylinder drainage effect. At the applied drainage rate of 7 slpm, the CO₂ mole fraction should be depleted by $0.035 \mu\text{mol mol}^{-1}$ and not enriched. At the same time the CH₄ mol fraction is $0.5 \text{ nmol mol}^{-1}$ higher when measured through the TCI intake system. Since the TCI uses the identical drying and sample pumping system as in Heidelberg and OPE, the results suggest that the observed offsets are associated either with the TCI 200 m intake line, their connections or the intake flushing pump. To further examine the origin of this offset an SIS test excluding the long intake line would have been helpful.

The direct cylinder measurements performed with the CRDS are for CO₂ $0.1 \mu\text{mol mol}^{-1}$ and for CH₄ $0.2 \text{ nmol mol}^{-1}$ lower compared to the direct TCI measurements. This is not in accordance with the results of the calibration comparison presented in section 3.1 and will be discussed in detail in section 4.1.

3.2.3 Sample intake system test OPE

The results of the SIS - test performed at OPE are shown in Fig. 2e and f. At OPE, the intake system test was performed without the entire 120 m intake line since it was not possible to climb the tower with a heavy high pressure cylinder. Nevertheless, the test included the sampling pumps and the drying systems. The flexible buffer volume of the SIS test was attached at the bottom of the 120 m sampling lines. For both sampling systems the line flush pumps had been switched off and the respective connections were plugged.

For CO₂ the agreement between the two intake systems and instruments is nearly perfect with a $0.01 \pm 0.06 \text{ } \mu\text{mol mol}^{-1}$ difference. The direct measurements and the measurements via the intake system agree for both instruments well within their respective errors. This implies that neither the drying systems nor the pumps create any significant CO₂ artefacts. At the same time the results highlight that the experimental set up of the (restricted) SIS test worked well.

The CH₄ difference of $1.6 \pm 0.4 \text{ nmol mol}^{-1}$ between the two instruments is however significant. Despite the good agreement of the calibration the comparison of the direct cylinder measurements yields already a difference of $0.9 \pm 0.4 \text{ nmol mol}^{-1}$. The difference between direct and intake system measurement is $0.4 \pm 0.4 \text{ nmol mol}^{-1}$ for the CRDS and $-0.3 \pm 0.4 \text{ nmol mol}^{-1}$ for the TCI. The observed CH₄ differences cannot be explained by calibration offset between the two instruments (see section 3.1). The cylinder used for the sample intake test was measured at the Heidelberg GC system as well, yielding a CH₄ mole fraction of $1868.8 \pm 2.4 \text{ nmol mol}^{-1}$, in better accordance with the TCI results. No explanation can be given for the lower CRDS values in the CH₄ intake sample test.

3.3 Comparison of ambient air measurements

The comparisons of the calibration gases as well as the investigation of the sample intake systems at each station were performed in order to gather diagnostic information for the subsequent evaluation of the ambient air comparison. The results of the comparison of the TCI against the GC at the reference station in Heidelberg are shown in detail in Appendix A. Here, we want to discuss only the results from the comparison experiments at CBW and OPE. These are displayed in a unified manner in Figures 3 and 4.

3.3.1 Ambient air comparison CBW

Figure 3 is divided into two similar comparison plots, one for CO₂ (3a) and one for CH₄ (3b). The upper panels of Figures 3a and 3b show the three minute mean values from the TCI in blue, respectively in green, and the one minute means from the CRDS in black. For CO₂, fast changing variations are superimposed on regular diurnal cycles. The CH₄ record is dominated by large fluctuations, presumably caused by changing catchment areas. Strong disturbances in the CH₄ signal correlate with those in the CO₂ record. To account for the different temporal resolutions of the two instruments, the CRDS measurements were convoluted with an exponential smoothing kernel representing a three minute turn-over time. The convolution accounts for the TCI-inherent smoothing in the sample cell, and improves the comparability of the two data sets. The second panels of Figures 3a and 3b show the differences between the measurements of the TCI and the CRDS. No data selection was applied. Each open symbol represents one three minute difference. The filled symbols represent a four hour running median of the individual differences.

For CO₂ the average difference between the two instruments varied around 0.2 μmol mol⁻¹; this difference is larger than the WMO-requested ILC target of 0.1 μmol mol⁻¹. The difference is not concentration-dependent. The variability of the three minute differences is generally smaller than ±0.2 μmol mol⁻¹. From June, 6 onwards the CO₂ variability is larger. This is caused by faster concentration changes in the ambient air CO₂ concentration. In Table 1 the descriptive statistics for the differences is given in more detail. A few outliers are larger than the chosen plot scale. Thus, Table 1 gives the 5th to 95th percentile in order to reflect the range of the typical differences. Potential reasons for the large CO₂ offset of 0.21 μmol mol⁻¹ between the two instruments will be discussed in section 4.

The smoothed CH₄ differences vary between 0 and 1 nmol mol⁻¹, with a slightly increasing trend. The median difference is 0.41 nmol mol⁻¹. The variability of the three minute differences increases from less than ±1 nmol mol⁻¹ on the first two days to ±2 nmol mol⁻¹. The pattern of the increasing variability is comparable to the one from the CO₂ comparison. The 5th to 95th percentile of the three minute differences spreads from -0.8 to 1.8 nmol mol⁻¹, highlighting that more than 90% of the individual three minute differences are within the WMO-requested ILC target of 2 nmol mol⁻¹.

The lowest panels of Figures 3a and 3b show the target/surveillance gas measurements of both instruments. The deviations of the averaged target/surveillance gas concentration minus

the one minute mean CRDS target measurements are shown as open red symbols. For the TCI the representation is vice versa: three minute mean TCI target measurements minus the averaged target/surveillance gas concentration are shown as black open symbols. This representation of the target/surveillance gas variations was chosen to allow for direct comparison between the trends in the target/surveillance gas records to the ambient air difference. The short term spread of the target/surveillance gas measurements represents the repeatability of the instruments (both instruments: $0.03 \mu\text{mol mol}^{-1}$ for CO_2 and $0.25 \text{ nmol mol}^{-1}$ for CH_4), whereas the consistency over the entire comparison period is a measure for reproducibility. For CO_2 as well as for CH_4 both instruments measure well within the requested $0.1 \mu\text{mol mol}^{-1}$ respectively 2 nmol mol^{-1} reproducibility targets for CO_2 and CH_4 for northern hemispheric sites. From the repeatability of the two instruments we can estimate the lower limit of the 1σ variability of the differences to $0.05 \mu\text{mol mol}^{-1}$ for CO_2 and $0.4 \text{ nmol mol}^{-1}$ for CH_4 . Any potential variability originating from the sample intake system, as well as temporal misalignments adds to this. None of the target/surveillance gas records holds evidence for instrumental drifts during the comparison period. The CRDS target/surveillance gas records show small diurnal patterns which are more distinct for CO_2 . The inter-diurnal variability in the smoothed CH_4 differences is to some extent present in the CRDS target/surveillance gas variability, however, no explanation for the slight temporal increase of the CH_4 differences between the TCI and the CRDS can be drawn from the target/surveillance gas records and thus from the analyser performance and calibration.

3.3.2 Ambient air comparison OPE

Figure 4 shows the results for the OPE comparison campaign and is structured identically to Figure 3. The upper panels of Figures 4a and 4b present the three minute mean CO_2 (blue), resp. CH_4 (green) measurements of the TCI and compares them to the one minute mean measurements of the CRDS (black). During the first three days of the two weeks comparison campaign no adequate CRDS data is available due to experimental problems with the ICOS prototype. As for CBW the CRDS measurements have been convolved with the exponential kernel before calculating the differences between the two instruments shown in the middle panels of Figures 4a and 4b. The individual three minute differences are shown as open symbols and have been smoothed using a four hour running median filter shown as filled symbols. For CO_2 the smoothed differences vary between 0.1 and $0.2 \mu\text{mol mol}^{-1}$ with a median difference of $0.13 \mu\text{mol mol}^{-1}$. The scatter of the individual differences is on the order

1 of $\pm 0.15 \mu\text{mol mol}^{-1}$. This is also expressed by the 5th and 95th percentile given in Table 1. The
2 observed median CO₂ difference between the TCI and the CRDS is again larger than the
3 WMO-requested ILC target and discussed in more detail in section 4.2. The smoothed CH₄
4 differences decrease from initially $0.7 \text{ nmol mol}^{-1}$ to $0.1 \text{ nmol mol}^{-1}$. The median difference is
5 $0.4 \text{ nmol mol}^{-1}$. Despite this trend the individual differences are generally well in line with the
6 WMO-requested ILC target. Even the 5th and 95th percentile of the three minute differences,
7 as given in Table 1, is well below the requested compatibility ranges.

8 The individual target/surveillance gas measurements of both instruments are shown in the
9 lower panels of Figure 4a and 4b. As already mentioned in section 3.1 the TCI performance
10 for CO₂ was worse at OPE than at CBW. This was due to large temperature variations (10 K)
11 within the measurement container and a slowly deteriorating flow controller in the TCI
12 instrument, decreasing the CO₂ reproducibility to $0.07 \mu\text{mol mol}^{-1}$. For more details on the
13 performance of the FTIR during this period see the companion paper by Hammer et al.
14 (2012). Also, the CO₂ repeatability of the CRDS system in OPE was worse ($1\sigma = 0.08 \mu\text{mol}$
15 mol^{-1}) than that at CBW, which is explained by the earlier CRDS version running at OPE
16 compared to CBW (see section 2.2). Both effects can be seen in the scatter of
17 target/surveillance gas measurements displayed in Figure 4a and 4b. The TCIs repeatability as
18 well as its reproducibility are affected. For CH₄ the TCI target measurements show a
19 decreasing trend of $0.6 \text{ nmol mol}^{-1}$ over the two weeks period. This trend is in accordance to
20 the observed decrease of the CH₄ differences and attributes it entirely to the performance of
21 the TCI at OPE. For both CRDS target/surveillance gas records an initial settling in effect is
22 detectable which disappears after the first three days when the actual comparison
23 measurements start. The TCI CO₂ target shows variations that are larger than the WMO ILC
24 ranges; this indicates poor reproducibility. In contrast to CH₄ no direct relation to the ambient
25 air difference of CO₂ is visible. The variability in the TCI CO₂ target/surveillance gas
26 measurements is even larger than the observed variability in the ambient air differences. This
27 is related to the deteriorating flow controller of the TCI which caused longer stabilization
28 times. These increased settling in effects are more severe for limited 30 min
29 target/surveillance gas measurements than for the continuous ambient air measurements. It
30 was generally observed that the TCI performance of CO₂ is much more affected from non-
31 stable conditions than CH₄ (Hammer et al., 2012).

3.4 Statistical evaluation of ambient air measurements

Since for all sites the CO₂ and CH₄ differences are not Gaussian distributed the results are presented in separate histograms in Figure 5 (for individual measurements of Heidelberg see Appendix A). A non-Gaussian distribution results if systematic errors are present. Sources of systematic errors could be numerous, e.g. drifts in one of the instruments, temporal misalignments among the instruments or incorrect calibrations/response functions. The WMO ILC targets are stated as standard deviations without specifying the distribution of the population (WMO, 2011), with the tacit understanding of the Gaussian distribution. However, the range covered by the standard deviation is distribution dependent. For example one standard deviation of the CO₂ difference at CBW (Figure 3a) includes 94% of all data. Thus we chose the Inter Quartile Range (IQR), a distribution independent measure, to quantify the spread of the data (Borradaile, 2003). Since it is common to report the standard deviation, we also report the standard deviation of a Gaussian fit to the histogram for comparison reasons. The IQR is far less sensitive to outliers, thus no data flagging needs to be applied, reducing randomness in selecting the flagging criteria and the effort in data pre-processing. The graphical representation of the histogram in combination with a fitted Gaussian distribution, as given in Figure 5, acts as an easy check of normality and highlights the presence and the extent of potential systematic errors.

Table 1 summarises the results of the TCI comparison to the standard instrumentation at the Heidelberg reference site and the two ICOS Demonstration Experiment sites. In addition to the median and the IQR, the parameters of the Gauss fit of the distributions as well as the 5th and 95th percentile are given. The median differences of the TCI relative to the Heidelberg reference GC are -0.02 µmol mol⁻¹ for CO₂ and -0.3 nmol mol⁻¹ for CH₄; the compatibility is thus well within the WMO-requested ILC targets for both gases. The large IQR value for CO₂ and the obvious mismatch between the CO₂ histogram and its Gauss fit are probably caused by the large variability of ambient air concentrations in combination with a temporal misalignment of the two instruments in Heidelberg (see Fig. A1). The differences between the TCI and the GC are largest and most distinct for periods with fast concentration changes. For a more detailed discussion on this topic see Appendix A. The same temporal misalignment problem is present for the CH₄ differences. However the temporal misalignment problem is masked by the CH₄ precision of the GC system (±2.3 nmol mol⁻¹), that constitutes the lower

limit for the variability of the differences which is exemplified in Figure 5. Since this statistical error of the CH₄ difference is already large it masks the systematic error. Thus the deviation of the CH₄ histogram from the Gauss fit is much smaller. From the good absolute agreement between the GC and the TCI we can conclude that it is possible to use the TCI in its current setup to assess the performance of a field station.

The variability of the differences is much smaller at the two field stations compared to that at the Heidelberg reference station. This has two reasons; first, the temporal resolution of the CRDS is 1 min and can, after proper smoothing, be more accurately compared with the 3 min data of the TCI, and secondly, the variability of the ambient concentrations in CBW and OPE are much smaller than in Heidelberg. For CBW only small deviations from the Gauss fit, mainly for positive differences are found. These deviations originate from a small increasing trend in the CO₂ and CH₄ difference (compare Fig. 3), accompanied with an increasing variability of the differences. Still, at both field stations, the variability of the differences, either expressed as IQR or as standard deviation of the Gauss fit, is smaller than the WMO ILC target. This highlights the potential of the parallel measurement approach, even without flagging any data points and without restricting the comparison to clean air or baseline conditions. The agreement between the instruments in terms of precision is impressive. For CH₄ also the accuracy between the different measurement systems is sufficient to fulfil the WMO ILC target. However, for CO₂ none of the measurements at the two field stations agrees with those of the TCI within the required limits.

4 Discussion of comparison results and suggestions for improvements

In the following section we will compare all our results, i.e. of direct standard measurements, SIS tests, if available, and parallel ambient air measurements at the ICOS **Demonstration** Experiment sites and try to understand the observed differences of ambient CO₂ and CH₄ measurements. For this purpose Table 2 combines the results of all tests performed at the field stations. In addition, we will discuss some of the misconceptions during this inter-comparison exercise that should be avoided in the future.

4.1 CBW comparison

The sample intake test at CBW was only successful for the TCI and did show a 0.09 ± 0.03 $\mu\text{mol mol}^{-1}$ higher CO_2 values when the cylinder gas passed the complete intake system of the TCI compared to the direct measurement. If the same bias would occur for real ambient air measurements at the TCI, this would explain about half of the observed ambient air difference of $0.21 \mu\text{mol mol}^{-1}$. The TCI measured the SIS test gas higher by $0.10 \mu\text{mol mol}^{-1}$ if compared to the CRDS instrument. Both effects together would account for the difference observed between the ambient CO_2 measurements at Cabauw. However, the result of the direct SIS test cylinder appears to contradict the results of the calibration gas comparison. Compared with the CAL-assigned values of the CBW working standards (Table 2) the TCI values were lower by $0.05 \mu\text{mol mol}^{-1}$. This contradiction points to some bias of the instrument calibration at Cabauw. In-depth investigation of the one minute CRDS measurements of the test gas at CBW did show that the CO_2 concentration does not reach a steady value within the measurement interval of five minutes which was chosen for cylinder gas analysis (Vermeulen et al., 2011). This may be caused by settling in effects of regulators and/or insufficient flushing of dead volumes in the calibration gas inlet system. The same problem occurs for the WS which are also flushed for only five minutes every 25 hours. The insufficient flushing of the WS affects the Licor NDIR CO_2 measurements as well. In the meantime the measurement duration for the WS has been increased to 15 min.

For CH_4 , the observed calibration offsets and the bias in the TCI SIS test are in accordance to the observed mean ambient air difference between the instruments. In addition, the calibration gas differences derived from the CBW WS measurements on the TCI and the offset between the direct measurements of the intake test cylinder agree within their uncertainties. This result is not in contradiction with the insufficient WS flushing which we observed for CO_2 , since pressure regulator effects are known to be much smaller for CH_4 than for CO_2 . This is also supported by the fact that CH_4 reached its equilibrium value within the 5 minute flushing time.

The TCI SIS test results for both species are higher if measured via the intake system. The ambient CO_2 and CH_4 concentrations measured directly prior to the SIS test have been lower compared to those of the SIS test cylinder, thus potential leaks in the SIS test setup are no likely explanation for the observed difference. However for both species elevated concentration levels are very likely to be present within the station building. Therefore

leakages at the connection of the sample line to the TCI or at the TCI line flushing pump are possible explanations. Retrospectively, a SIS test without the sample intake line would have been helpful to further narrow down the source of the discrepancies.

4.2 OPE comparison

The median CO₂ offset at OPE between the two instruments (0.13 μmol mol⁻¹) is also larger than the WMO ILC target. The calibration difference between the instruments can only account for 0.03 μmol mol⁻¹. The “restricted” sample intake system test did not show any differences between the two instruments for CO₂. This leads to the conclusion that the pumps and the drying systems of both instruments probably do not cause any systematic offset. Thus, 0.10 μmol mol⁻¹ of the CO₂ difference remains unexplained. Due to logistical problems it was not possible to test the intake lines at OPE. Thus the untested lines and potentially the line flushing pumps might have caused the unexplained differences. The differences are of similar size compared to the 0.09 μmol mol⁻¹ difference which was found during the TCI SIS test at CBW. Since the TCI and the CRDS sample line are made of identical ½” Synflex 1300 material and the lines have been flushed at similar rates, different contaminations of the two lines are not very likely. It would be in accordance with all our test results if the TCI line flushing pump would have caused this 0.10 μmol mol⁻¹ offset. The pump (Becker, VT 4.4) was, however, never tested for contamination since it was placed downstream of the entire sample intake system.

The median ambient air CH₄ offset at OPE is 0.44 nmol mol⁻¹ and thus well within the WMO ILC target. However, the results of the SIS test and the ambient air comparison are not in accordance. From the sample intake system test we would have expected a difference of 1.6 nmol mol⁻¹ between the two systems (see Fig. 2f). Detailed investigation of both the CRDS and the TCI results of the CH₄ SIS test could not explain this discrepancy.

4.3 Summary of differences

Figure 6 summarises the observed differences between the TCI and instrumentation at its reference station Heidelberg and at the two field stations CBW and OPE. Before and after each field trip, the TCI was checked against the Heidelberg GC system to reassure that its performance did not change. For both species, CO₂ and CH₄, the TCI and the Heidelberg GC did agree before, in between, and after the field trips to CBW and OPE. The averaged differences between HEI-GC and TCI did not change significantly. This finding, in combination with a smooth TCI target/surveillance gas record (Hammer et al., 2012), gives confidence that the TCI remained stable over the time of the **Demonstration** Experiment and can thus serve as reference instrument for both station visits. The generally larger and increasing variability of the CO₂ differences relative to the GC system are discussed in Appendix A and can be explained by temporal misalignment between the two instruments.

The CO₂ deviations at CBW can possibly be partly explained by the offset in the TCI intake line or the line flushing pump and the systematic error of the CRDS calibration caused by insufficient flushing of the working standards. At OPE, the CO₂ deviations are smaller than at CBW and close to the WMO target; however, no explanation for the significant difference could be found. The only parts which remain essentially untested at OPE are the sample intake lines. Dedicated tests of the intake lines have to be carried out at both stations in order to clarify, i.e. if the 0.09 μmol mol⁻¹ contamination which we have found for the TCI sampling line in CBW is indeed reproducible and of the same size when ambient air measurements are performed and if the CO₂ offset observed between the two instruments at OPE is actually caused by a similar line effect.

The CH₄ deviations at CBW and OPE are of the same size and both smaller than the limit of the ILC target.

4.4 Critical assessment of the comparison experiment and suggestions for improvements

The feasibility study of using a travelling instrument to detect differences in the order of the WMO ILC target was successful. The **high** precision of the two optical instruments used in the comparison, i.e. the FTIR and the CRDS, allowed us to even detect small differences in the co-located measurements over a short comparison period of one to two weeks. Significant offsets can also be detected on even smaller time scales like e.g. hours. The approach is truly

comprehensive since ambient air collected from the same inlet point is compared. In contrast to earlier studies, our data evaluation method allows comparing the entire ambient concentration range so that we are not restricted to background conditions, only. With the dedicated component tests, like the SIS test and the calibration check, the travelling instrument approach represents a true diagnostic evaluation directly at the field station under its potentially unique conditions. The travelling instrument approach is, however, very labour intensive, since the TCI has to be referenced to independent instrumentation at its home base before and after each field trip. This reduces the comparison capacity of one TCI to only four or five station visits per year. From our experience, conducting the TCI approach, including preparation and data evaluation, would need the capacity of one full scientifically trained person, at least during the first year. In addition, as long as the FTIR is used as TCI, the campaigns have to be supported by a technician or a student helper. A smaller instrument also with less calibration gas consumption would allow to be installed by one person only. The financial requirements for an ongoing, institutional TCI quality control approach are thus to cover the consumables, gases, drying agent etc. for the instrument, the salaries of the personnel and the travel and subsistence costs.

The evaluation of the feasibility study has demonstrated the potential of the approach but at the same time improvements are needed:

- First and foremost, near-real time data evaluation is a key requirement. Calibrated data from both instruments must be available within 24h during an inter-comparison campaign. This will allow for quickly reacting to experimental problems, e.g. realise the necessity to repeat the sample intake test in CBW. Due to delays in the start of the ICOS Demonstration Experiment, the travelling instrument campaign took place right at the beginning of this ICOS phase where the data processing chains between the field stations and the ICOS Atmospheric Thematic Centre, responsible for processing the continuous data, had not yet been fully operational. Thus, CRDS data was only available a few weeks after the station visits.
- The example of insufficient flushing times in the CBW setup has shown that a check of the station's calibration standards with the travelling instrument is not sufficient. Measuring the WS of the TCI at the station instrument will give additional information regarding the transfer of the WS calibration to the station calibration. In our comparison

1 approach we had only one test gas measurement from the SIS test which already
2 highlighted a respective problem.

- 3 • For future TCIs it will be mandatory for the TCI host laboratory to have a separate intake
4 line for the TCI instrument and use the same line flushing pump as during the field
5 comparisons.

6

7

5 Possible quality management strategies for a network of continuous analysers

5.1 Requirements for a comprehensive quality management strategy

The travelling instrument approach evaluated in this study could be part of an overall quality management of continuous atmospheric monitoring stations. The requirements for such a comprehensive quality management (QM) of a network are complex. Thus, multiple tools will be needed to cover various aspects. In the following, we will briefly outline the key points that should be addressed when designing such a quality management system:

- **Precision:** the precision of the quality management tool defines the needed statistics to detect differences on the order of the WMO ILC targets. Preferably the precision of a QM tool should be at least twice the requested ILC targets (WMO, 2012).
- **Frequency:** comparison frequency must allow to quickly detecting potential problems.
- **Comprehensiveness:** it must describe which parts of the analytical setup are tested
- **Concentration range coverage:** it should be suitable for the investigated station and slightly exceed the stations ambient concentration range.
- **External station validation:** this quality control item is mandatory to create credibility of the network data.

Table 3 summarises quality control (QC) techniques which could be applied for field stations and classifies them according to the previously defined quality management characteristics. Each QC technique mentioned has its own strengths and weaknesses.

In the last decades, regular target/surveillance gas measurements have proven invaluable to monitor and quantify instrument performance. Target gases can be measured very precisely and at a high frequency. Thus target/surveillance gas measurements will form the backbone of the QM to be able to react timely on any instrumental malfunction including drifts in WS. The major shortcoming of the target/surveillance gases is that they are not comprehensive and test only the instrument response but not the air intake system. To overcome this problem, we

1 propose to use multiple targets and multiple insertion points for one of these targets. In
2 addition to the usual insertion point, e.g. at the sample selection valve, a second insertion
3 point at the end of the ambient air intake line is proposed. Such an alternative
4 target/surveillance gas insertion point was already proposed by Stephens et al. (2011). In our
5 setup we propose to alternate one of the target/surveillance gases between the two insertion
6 points and thus direct the gas through parts of the sample intake system, such as e.g. the
7 drying system and/or the sample pumps. Performing such tests on a weekly to monthly basis
8 in addition to regular direct target measurements will improve the comprehensiveness of the
9 target/surveillance gas measurements enormously. We propose to use yet an additional
10 target/surveillance gas cylinder to check the intake system since gas consumption will be
11 high. However, even this extended target/surveillance gas concept can still not provide an
12 adequate external validation of the compatibility of the station measurements. Target gases
13 can be calibrated by an external body prior and after usage and thus represent some external
14 validation but with very little temporal coverage.

15 Concurrent flask sampling could fill this gap. Regular, e.g., weekly flask samples which are
16 analysed at a different laboratory and compared to the in situ measurements provide the
17 needed external validation. Up to now, the 1σ variability of flask versus in situ comparisons
18 are on the order of 0.3 to 0.8 $\mu\text{mol mol}^{-1}$ for CO_2 (Masarie et al., 2011a, Masarie et al.,
19 2011b), therefore, many single comparisons are needed to achieve the desired precision of
20 inter-laboratory biases. However, the variability of the flask in situ comparisons is expected to
21 improve, once flasks can be filled over an extended, well defined, period (e.g. one hour by
22 using a buffer volume) as proposed recently by Turnbull et al. (2012). This should make the
23 flask sample more representative and will ease the comparison to continuous ambient
24 measurements. The flask in situ comparison will cover the entire concentration range at the
25 station and will thus allow for investigating concentration dependencies of biases.

26 Especially if the flask measurements and the in situ measurements are performed within the
27 same measurement network, it is strongly recommended to also introduce a network external
28 QC measure. Such comparisons could be similar to today's WMO Round Robin (Zhou et al.,
29 2011) exercise or the European Cucumber ICP (Manning et al., 2009) and can be performed
30 on a lower frequency. The important aspect is that they have to be completely independent
31 from the network they are monitoring.

True comprehensiveness is only established via a completely independent measurement that includes the air sampling system, such as the flask in situ comparison. In order to have a second comprehensive check of the whole system we propose to perform a sample intake system test, at least after setting up a new station, and regularly e.g. once per year.

The TCI is an alternative, completely independent and comprehensive QC measure. In contrast to concurrent flask sampling analysis, this approach is capable of detecting small discrepancies, even on a sub hourly time scale. We propose thus to use the TCI as a diagnostic tool within a monitoring network. Its travel route should however be flexible and be decided based on the results of the other QM measures.

5.2 Proposed quality management strategy for a monitoring network with continuous high precision analysers

- 1) Entire check of all intake lines before a station becomes operational. If the intake line of a station is made of Synflex and has connectors which are exposed to the weather this line test should be repeated at a regular interval.
- 2) High frequency instrument target/surveillance gas measurements at the station to be able to quickly detect malfunctioning of the instrument, insertion point: selection valve (at least daily).
- 3) Low frequency instrument target/surveillance gas measurements to be able to quantify system stability over decades, insertion point: selection valve.
- 4) Low frequency (e.g. weekly) intake target/surveillance gas measurements with *two* insertion points: a) selection valve, b) prior to drying system and pump.
- 5) Regular flask – in situ comparison: weekly
- 6) Travelling cylinders (highest hierarchy, ideally calibrated by the WMO CCL) to check calibration scales. This should be carried out as a blind test every 2 years.
- 7) The travelling instrument serving as a diagnostic tool particularly for stations where systematic biases in the flask vs. in situ comparison occur.

The QM strategy as proposed above does fulfil all needs in terms of precision, frequency, concentration range coverage, external validation as well as comprehensiveness. The diversity

1 of the applied QC measures complement one another and offer sufficient redundancies to act
2 as a defensible QM system. Technical restrictions at the monitoring site or other constraints
3 may impede the implementation of the entire scheme. In any case we would advise to have at
4 least the high frequency target/surveillance gas measurements and the regular flask – in situ
5 comparison.
6

6 Conclusions

The approach of using a travelling comparison instrument as a quality control measure was successfully tested at two field stations for CO₂ and CH₄. The TCI approach has proven to be sufficiently precise to detect differences between the measurement systems, which are well below the WMO ILC targets. Even on a three minute temporal resolution the spread, measured either as IQR or fitted standard deviation, of the differences between the optical instruments is smaller than the ILC target. This allows for very detailed (e.g. hourly) investigation of the differences. For such high resolution comparisons an extended definition of the ILC target would be desirable, adding variance limits to the current offset limits. These variance limits should be defined in close cooperation with the inverse modelling community and should anticipate the scientific needs and the capacity of future fine grid models. Ideally, these variance limits should be defined in a distribution independent manner.

The combination of the TCI with the dedicated test of the instrument components like the calibration check and the SIS test allow for subdivided testing of individual components of the measurement device. The proposed SIS test constitutes, besides co-located flask sampling, the only comprehensive test of the entire instrument system. Despite the experimental problems which we encountered in this feasibility study the SIS test has a high diagnostic potential. However, it also showed that more work and *on-site* evaluation of the results by the station team in cooperation with the TCI team is required to eliminate observed biases after they have been diagnosed. The results of this study highlight the demand for truly comprehensive QC. The proposed QM strategy could fulfil the diverse requirements and provide sufficient redundancy to establish traceable and defensible data quality standards.

Acknowledgements

We wish to thank Pim van den Bulk (ECN), Cyrille Vuillemin (LSCE), and Maxime Simon (ANDRA) for their very helpful support at the stations. Dominik Schmithüsen (IUP) is gratefully acknowledged for his help during the OPE campaign. **We thank the three anonymous reviewers for their valuable comments that helped to improve this paper.** This work has been funded by the European Commission under the ICOS Preparatory Phase Project (INFRA-2007-211574)

Appendix A: Discussion on the CO₂ variability for the Heidelberg TCI comparison

Temporal representativeness of the measurements

Depending on the measurement principle and instrumentation, different temporal resolutions of ambient air measurements are achieved. Classical GC systems have a sampling frequency of 5 to 15 min, and the measurement represents a snap-shot of the currently sampled atmospheric condition. The averaging time of the in situ FTIR instrument is three minutes, which corresponds to the turn over time τ of the white cell in the spectrometer (Hammer et al., 2012). The large white cell of the FTIR smoothes the high frequent variability in atmospheric trace gases which might be captured by the snap-shot type measurement of the GC system. Especially for polluted or semi-polluted stations close to GHG sources this leads to a representativeness difference between the measurement systems. To increase the representativeness of the snap-shot like GC measurements, a spherical 10 litre buffer volume was introduced in the GC sampling line. A detailed description of the buffer volume concept can be found in Winderlich et al. (2010). In the Heidelberg GC set up, the turn over time of the buffer was set to 20min. For the comparisons presented in this publication, the FTIR measurements have been aggregated to a 20 min turn over time by convolution with an exponential averaging kernel. This technique should avoid the temporal representativeness difference between the two instruments.

In Heidelberg we put the 10 litre buffer volume after the drying system (cryo traps) of the GC. The cryo traps however change their flow resistance, depending on the amount of ice they have accumulated. As a consequence, the flow rate through the cryo traps is not constant leading to a non constant turn over time of air in the buffer. The impact of the disagreement between assumed and real turn over times is directly dependent on the concentration change that occurs during the sample interval. Since Heidelberg is a polluted site the concentration changes are quite large.

This temporal misalignment problem caused an increased scatter in the HEI concentration differences, however it did not affect the average difference. This can be seen by the symmetric deviations in Figure A1. The uncertainty of the mean CO₂ differences in Heidelberg shown in Figure 6 increased as the ambient air became more humid during summer, increasing the variability of the flow through the cryo trap.

References

- Borradaile, G. J., *Statistics of Earth Science Data: Their Distribution in Time, Space and Orientation*, Springer, Berlin, 2003
- Brunner, J.: Field intercomparison of air quality measurements in Switzerland, *CHIMIA International Journal for Chemistry*, **63**(10), 661–664, 2009.
- Delmotte, M., Laurent, O., Conil, S., Corpacci, O., Simon, M., Baum, S., Hazan, L., Levin, I., Ramonet, M., Schmidt, M., Vuillemin, C., and Ciais, P.: A level one ICOS atmospheric station in Eastern France conference contribution: 16th WMO/IAEA meeting (GGMT-2011), Wellington, New Zealand, 25–28 Oct 2011, available at: <http://www.niwa.co.nz/atmosphere/ggmt-2011/posters>, last access: 31.Aug.2012, 2011.
- Griffith, D. W. T., Deutscher, N. M., Caldwel, C., Kettlewell, G., Rigganback, M., and Hammer, S.: A Fourier Transform Infrared trace gas analyser for atmospheric applications, *Atmospheric Measurements Techniques, Atmos. Meas. Tech. Discuss.*, **5**, 3717-3769, doi:10.5194/amtd-5-3717-2012, 2012.
- Hammer, S., Glatzel-Mattheier, H., Müller, L., Sabasch, M., Schmidt, M., Schmitt, S., Schönherr, C., Vogel, F., Worthy, D. E., and Levin, I.: A gas chromatographic system for high-precision quasi-continuous atmospheric measurements of CO₂, CH₄, N₂O, SF₆, CO and H₂, available at: http://www.iup.uni-heidelberg.de/institut/forschung/groups/kk/GC_html, last access: 2 Aug. 2012, 2008.
- Hammer, S., Griffith, D. W. T., Konrad, G., Vardag, S., Caldwel, C. and Levin, I.: Assessment of a multi-species in situ FTIR for precise atmospheric greenhouse gas observations, *Atmospheric Measurement Techniques Discussions*, **5**(3), 3645–3692, 2012.
- Keeling, R. F., Manning, A. C., Paplawsky, W. J. and Cox, A. C.: On the long-term stability of reference gases for atmospheric O₂/N₂ and CO₂ measurements, *Tellus B* **59**(1), 3–14, 2007.

1 Levin, I., Hammer, S., Eichelmann, E. and Vogel, F. R.: Verification of greenhouse gas
2 emission reductions: The prospect of atmospheric monitoring in polluted areas, *Phil. Trans. R.*
3 *Soc. A* vol. 369 no. 1943 1906-1924. doi:10.1098/rsta.2010.0249, 2010.

4
5 Manning, A. C., Jordan, A., Levin, I., Schmidt, M., Neubert, R. E. M., Etchells, A., Steinberg,
6 B., Ciais, P., Aalto, T., Apadula, F., Brand, W. A., Delmotte, M., Giorgio di Sarra, A., Hall,
7 B., Haszpra, L., Huang, L., Kitzis, D., van der Laan, S., Langenfelds, R. L., Leuenberger, M.,
8 Lindroth, A., Machida, T., Meinhardt, F., Moncrieff, J., Morgu , J. A., Necki, J., Patecki, M.,
9 Popa, E., Ries, L., Rozanski, K., Santaguida, R., Steele, L. P., Strom, J., Tohjima, Y.,
10 Thompson, R. L., Vermeulen, A., Vogel, F., and Worthy, D. E.: Final report on CarboEurope
11 ‘Cucumber’ intercomparison programme, available at:
12 http://cucumbers.uea.ac.uk/documents/CucumberFinalReport_Final.pdf, last access 31. Aug. 2012,
13 2009.

14
15 Masarie, K., Chao, D., Novelli, P. and Tans, P.: New Capabilities for Interpreting Comparison
16 Data, Report of the 15th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other
17 Greenhouse Gases, and Related Tracers Measurement Techniques, 7–10 September 2009,
18 GAW Report No. 194, WMO TD No. 1553, 2011a.

19
20 Masarie, K., Conway, T.J., and Tans, P.: What are comparison results telling us?, conference
21 contribution: 16th WMO/IAEA meeting (GGMT-2011), Wellington, New Zealand, 25–28
22 Oct 2011, available at: <http://www.niwa.co.nz/atmosphere/ggmt-2011/presentations>, last
23 access: 31.Aug.2012, 2011b.

24
25 Prinn, R.G., Weiss, R.F., Fraser, P.J., Simmonds, P.G. , Cunnold, D.M., O’Doherty, S.,
26 Salameh, P., Porter, L., Krummel, P. , Wang, R.H.J., Miller, B. , Harth, C. , Gready, B., Van
27 Woy, F.A. , Steele, L.P. , Muehle, J., Sturrock, G., Alyea, F.N., Huang, J., and Hartley, D.E.:
28 The ALE/GAGE/AGAGE Network, Carbon Dioxide Information and Analysis Center, J.
29 *Geophys. Res.*, 105, D14, 17751—17792, 2000.

1 Stephens, B. B., Miles, N. L., Richardson, S. J., Watt, A. S., and Davis, K. J.: Atmospheric
 2 CO₂ monitoring with single-cell NDIR-based analyzers, *Atmos. Meas. Tech.*, 4, 2737-2748,
 3 doi:10.5194/amt-4-2737-2011, 2011.

4

5 Turnbull, J., Guenther, D., Karion, A., Sweeney, C., Anderson, E., Andrews, A., Kofler, J.,
 6 Miles, N., Newberger, T., Richardson, S., and Tans, P.: An integrated flask sample collection
 7 system for greenhouse gas measurements, *Atmos. Meas. Tech. Discuss.*, 5, 4077-4097,
 8 doi:10.5194/amtd-5-4077-2012, 2012.

9

10 Vermeulen, A. T., Hensen, A., Popa, M. E., van den Bulk, W. C. M., and Jongejan, P. A. C.:
 11 Greenhouse gas observations from Cabauw Tall Tower (1992–2010), *Atmos. Meas. Tech.*, 4,
 12 617-644, doi:10.5194/amt-4-617,2011.

13

14 Vogel F.R., Jeschka J., Kromer B., Steinhof A., Hammer S. and Levin I., A robust setup for
 15 long-term monitoring of fossil fuel CO₂ and other trace gases, and its test against established
 16 conventional monitoring in Heidelberg, Report of the 15th WMO/IAEA Meeting of Experts
 17 on Carbon Dioxide, Other Greenhouse Gases, and Related Tracers Measurement Techniques,
 18 7–10 September 2009, GAW Report No. 194, WMO TD No. 1553, 2011.

19

20 Winderlich, J., Chen, H., Höfer, A., Gerbig, C., Seifert, T., Kolle, O., Kaiser, C., Lavrič, J.V.,
 21 and Heimann, M.: Continuous low-maintenance CO₂/CH₄/H₂O measurements at the Zotino
 22 Tall Tower Observatory (ZOTTO) in Central Siberia, *Atmos. Meas. Tech.*, 3, 1113-1128,
 23 doi:10.5194/amt-3-1113-2010, 2010.

24

25 WMO: Report of the 15th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other
 26 Greenhouse Gases, and Related Tracers Measurement Techniques, 7–10 September 2009,
 27 GAW Report No. 194, WMO TD No. 1553, available at:
 28 [http://www.wmo.int/pages/prog/arep/gaw/5 documents/GAW 194 WMO TD No 1553 web](http://www.wmo.int/pages/prog/arep/gaw/5_documents/GAW_194_WMO_TD_No_1553_web_low_resol.pdf)
 29 [low resol.pdf](http://www.wmo.int/pages/prog/arep/gaw/5_documents/GAW_194_WMO_TD_No_1553_web_low_resol.pdf), last access: 2 Feb 2012, Jena, Germany, 2011.

30

1 WMO: Report of the 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse
2 Gases, and Related Measurement Techniques (GGMT-2011), Wellington, New Zealand, 25-
3 28 October 2011, 67 pp. October 2011 GAW Report No. 206, available at:
4 <http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html>, last access: 26 Jan 2013,
5 Wellington, New Zealand, 2012.

6
7
8 Worthy D. E., Huang L., Jordan A., Brand W., Schmidt M., Ramonet M., Levin I., Hammer
9 S., Langenfelds R. L., C.E. Allison C.E.: Flask Air Sampling Programs at the Alert GHG
10 Comparison Site: Presentation of Results, conference contribution 16th WMO/IAEA meeting
11 (GGMT-2011), Wellington, New Zealand, 25–28 Oct 2011, available at:
12 <http://www.niwa.co.nz/atmosphere/ggmt-2011/presentations>, last access: 31. Aug. 2012,
13 2011.

14
15 Zellweger C., Klausen J., and Buchmann B., Summing Up 14 Years of Intercomparison
16 Activities of the World Calibration Centre at Empa (WCC-Empa): Methane, Carbon
17 Monoxide and Ozone, Report of the 15th WMO/IAEA Meeting of Experts on Carbon
18 Dioxide, Other Greenhouse Gases, and Related Tracers Measurement Techniques, 7–10
19 September 2009, GAW Report No. 194, WMO TD No. 1553, 2011.

20
21 Zhou, L.X., Kitzis, D. R., Tans, P., Masarie, K. and Chao, D.: WMO Round-Robin inter-
22 comparison: Progress and a new website, Report of the 15th WMO/IAEA Meeting of Experts
23 on Carbon Dioxide, Other Greenhouse Gases, and Related Tracers Measurement Techniques,
24 7–10 September 2009, GAW Report No. 194, WMO TD No. 1553, 2011.

1 *Table 1: Descriptive statistics of the individual CO₂ and CH₄ differences for the Heidelberg*
2 *reference station and the two ICOS **Demonstration** Experiment stations. The median and the*
3 *Inter Quartile Range (IQR) of the distributions are given in the first two columns followed by*
4 *the 5th and 95th percentile. The last two columns state the results of the fitted Gaussian*
5 *distribution.*

	median	IQR	5 th percentile	95 th percentile	Gauss fit centre	Gauss fit std.dev.
HEI (TCI – GC)						
ΔCO ₂ [μmol mol ⁻¹]	-0.02	0.27	-1.13	1.49	-0.02	0.08
ΔCH ₄ [nmol mol ⁻¹]	-0.3	3.6	-5.1	5.1	-0.3	2.3
CBW (TCI-CRDS)						
ΔCO ₂ [μmol mol ⁻¹]	0.21	0.09	0.08	0.40	0.21	0.06
ΔCH ₄ [nmol mol ⁻¹]	0.41	0.76	-0.77	1.78	0.38	0.50
OPE (TCI – CRDS)						
ΔCO ₂ [μmol mol ⁻¹]	0.13	0.10	-0.02	0.28	0.13	0.07
ΔCH ₄ [nmol mol ⁻¹]	0.44	0.51	-0.28	1.15	0.43	0.36

6

7

1 *Table 2: Overview of the different comparison tests performed at the ICOS field stations. The*
2 *representation of the test results is presented with respect to their sign, to be directly*
3 *comparable to the observed ambient air difference.*

	Differences				
	Calibration: TCI- assigned	SIS cylinders: TCI-CRDS	SIS test CRDS: direct- line	SIS test TCI: line- direct	ambient air: TCI-CRDS
CBW					
ΔCO_2 [$\mu\text{mol mol}^{-1}$]	-0.05±0.05	0.10±0.02	-	0.09±0.03	0.21±0.06
ΔCH_4 [nmol mol^{-1}]	-0.4±0.4	0.2±0.3	-	0.5±0.2	0.4±0.5
OPE					
ΔCO_2 [$\mu\text{mol mol}^{-1}$]	0.03±0.04	0.03±0.04	0.02±0.05	0.00±0.06	0.13±0.07
ΔCH_4 [nmol mol^{-1}]	-0.2±0.3	0.9±0.4	0.4±0.4	0.2±0.3	0.4±0.4

4

5

1 *Table 3: Classification of existing quality management approaches according to the pre-*
2 *defined quality management characteristics*

3

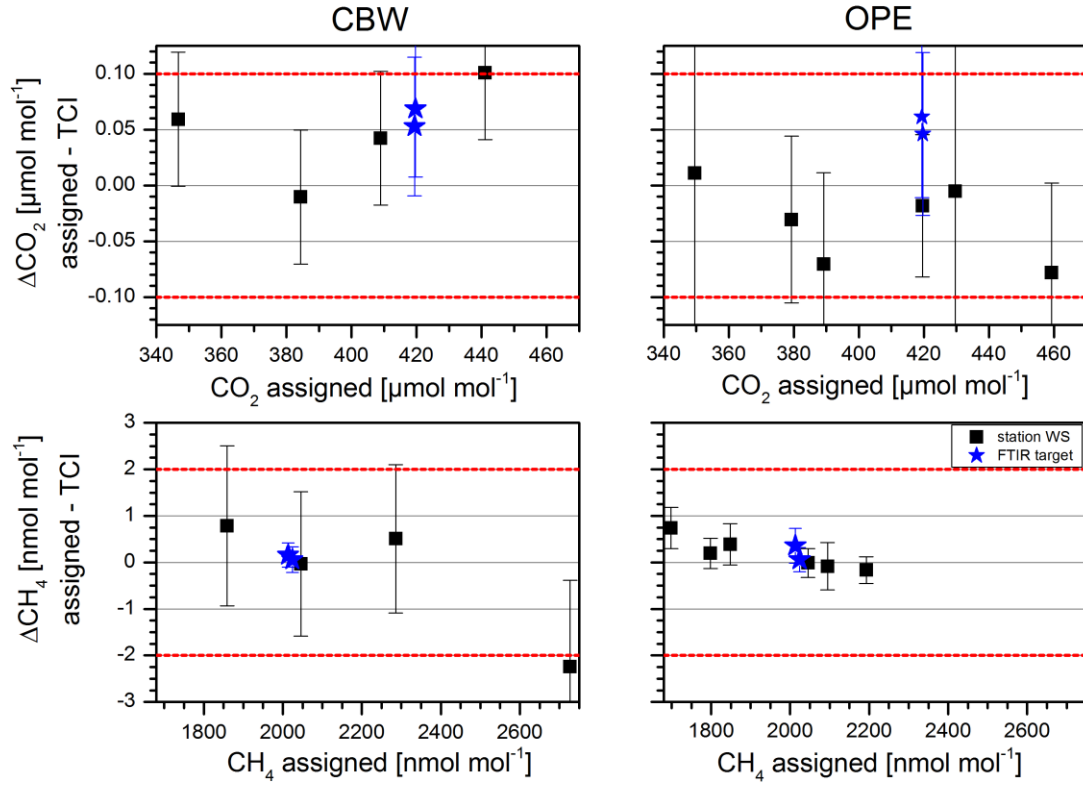
	Precision	Frequency	Comprehensive ness	Conc. Range	External validation
Station targets	High	Sub-daily	No	Limited	Limited
Flask vs. in situ	Low	Weekly	Full	Entire	Yes
Cylinders	High	1-2 per year	No	Limited	Yes
Travelling inst	High	Years	Full	Entire	Yes

4

5

6

1



2

3 *Figure 1: Measurement of the CBW (left) and OPE (right) working standards with the TCI at*
4 *the respective station (black symbols). CO₂ results are shown in the upper and CH₄ in the*
5 *lower panels. The measured means are averages of two times five individual (3 min) TCI*
6 *measurements; the error bars give the 1σ uncertainty of the difference accounting for the*
7 *uncertainties in the assignment as well as the repeatability or the reproducibility (which ever*
8 *is larger) for the TCI measurements. Blue stars denote differences of the averaged daily TCI*
9 *target measurements from the CAL-assigned values and their uncertainties at the respective*
10 *station. Red dashed lines show the WMO ILC target ranges for each species.*

11

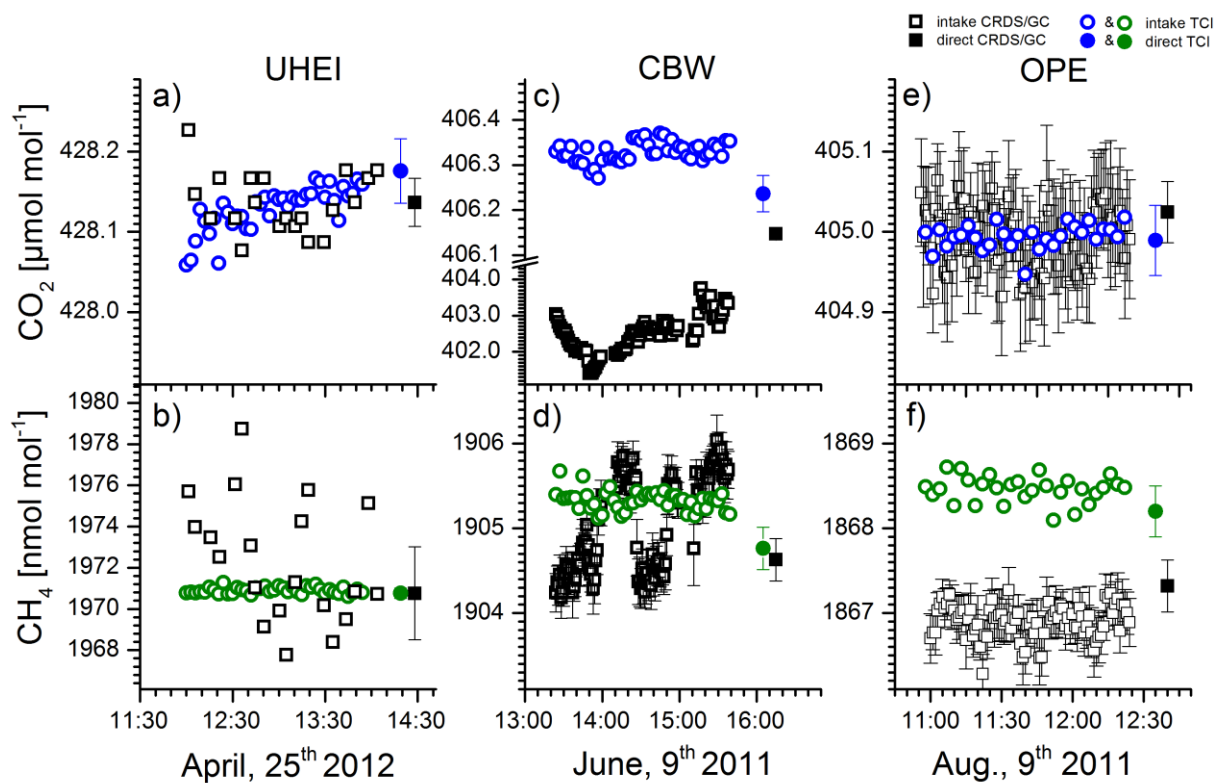


Figure 2: Sample intake system tests at the three stations: Open symbols represent individual measurements via the entire sample intake system, including intake line, pumps and drying systems. Filled symbols denote averaged measurements of the same cylinder without the intake system. Black symbols represent station instrument, coloured symbols measurements with the TCI.

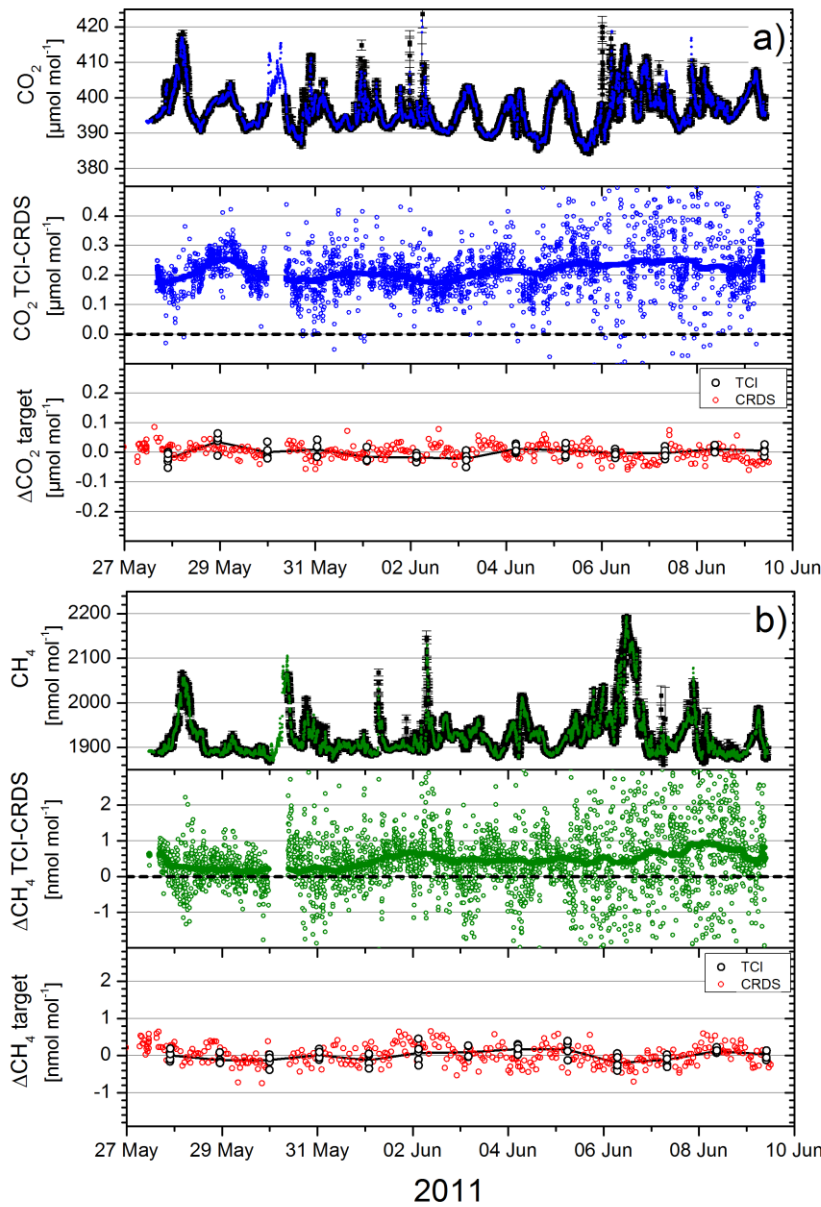


Figure 3: Upper panels in a & b: Comparison of CRDS (black) and TCI (blue/green) of CO_2 or CH_4 measurements performed at the 200m level at CBW. Middle panels: individual 3 minute mean CO_2 or CH_4 differences between the TCI and the smoothed CRDS measurements (open symbols); Averaged differences (4h running median filter) as filled symbols. Lower panels: Individual one minute CRDS target measurements in red and individual three minute TCI target measurements in black. For both instruments the target/surveillance gas deviations are given in a way to allow for direct trend comparison to the ambient air differences.

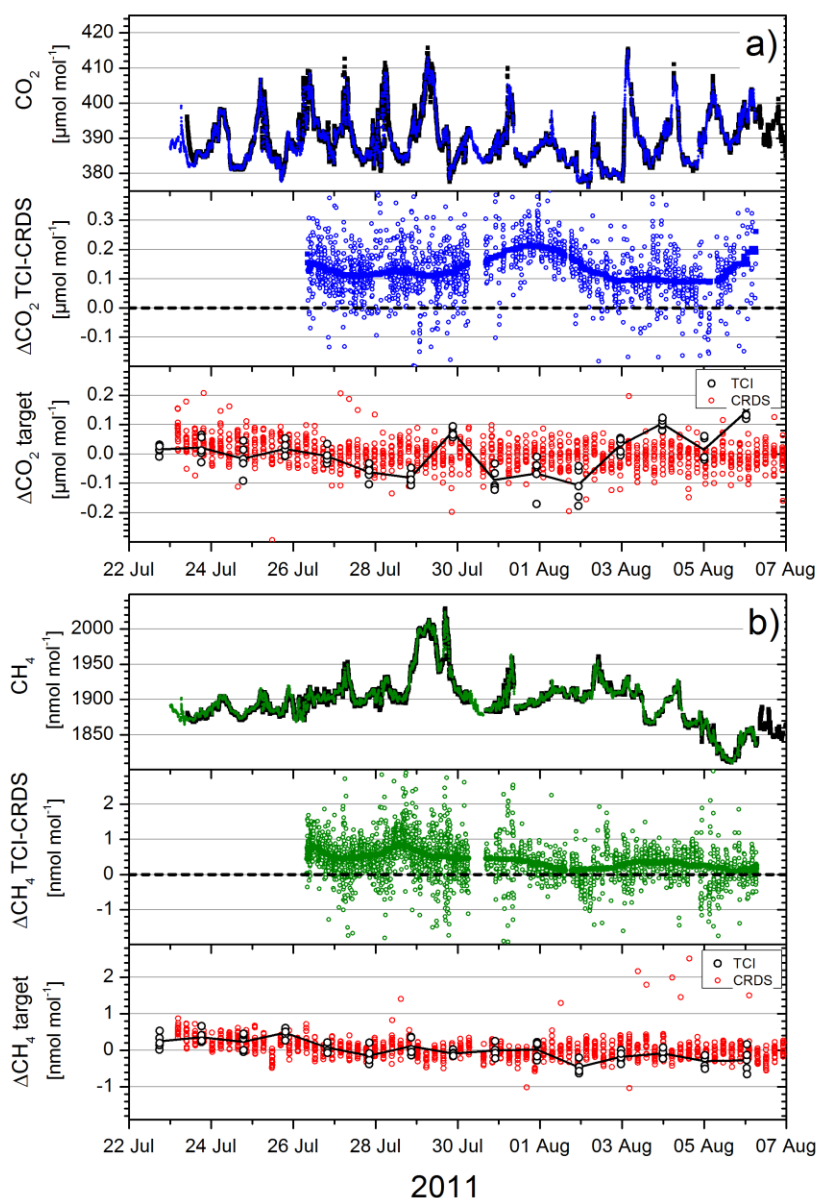


Figure 4: Upper panels in a & b: Comparison of CRDS (black) and TCI (blue/green) of CO_2 or CH_4 measurements performed at the 120m level at OPE. Middle panels: individual 3-minute mean CO_2 or CH_4 differences between the TCI and the smoothed CRDS measurements (open symbols); Averaged differences (4h running median filter) as filled symbols. Note, the first two days had to be excluded due to known problems of the CRDS intake system. Lower panels: Individual one minute CRDS CO_2 or CH_4 target measurements in red and individual three minute TCI target measurements in black. For both instruments the target/surveillance gas deviations are given in a way to allow for direct trend comparison to the ambient air differences.

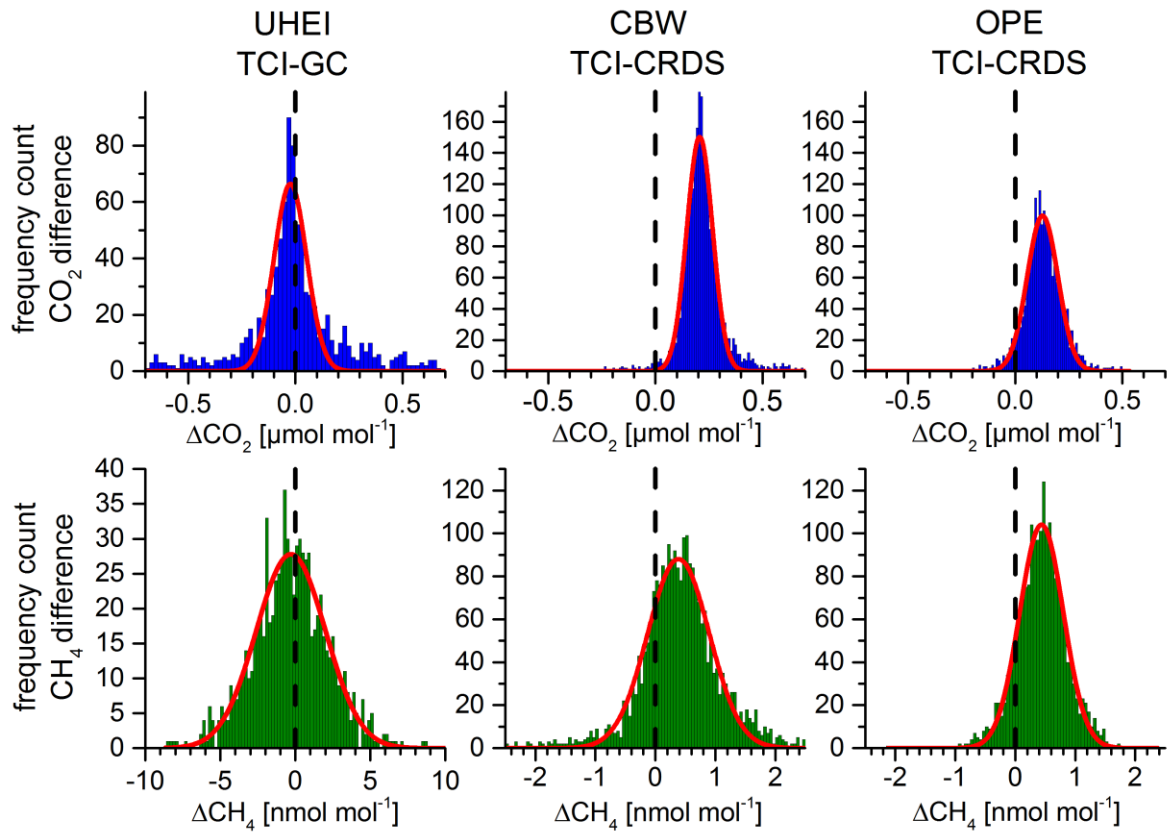


Figure 5: Histograms of the individual CO_2 (upper panel) and CH_4 (lower panel) differences for the different stations. For each histogram a Gaussian distribution is fitted and given in red. Note the different scale for CH_4 in Heidelberg.

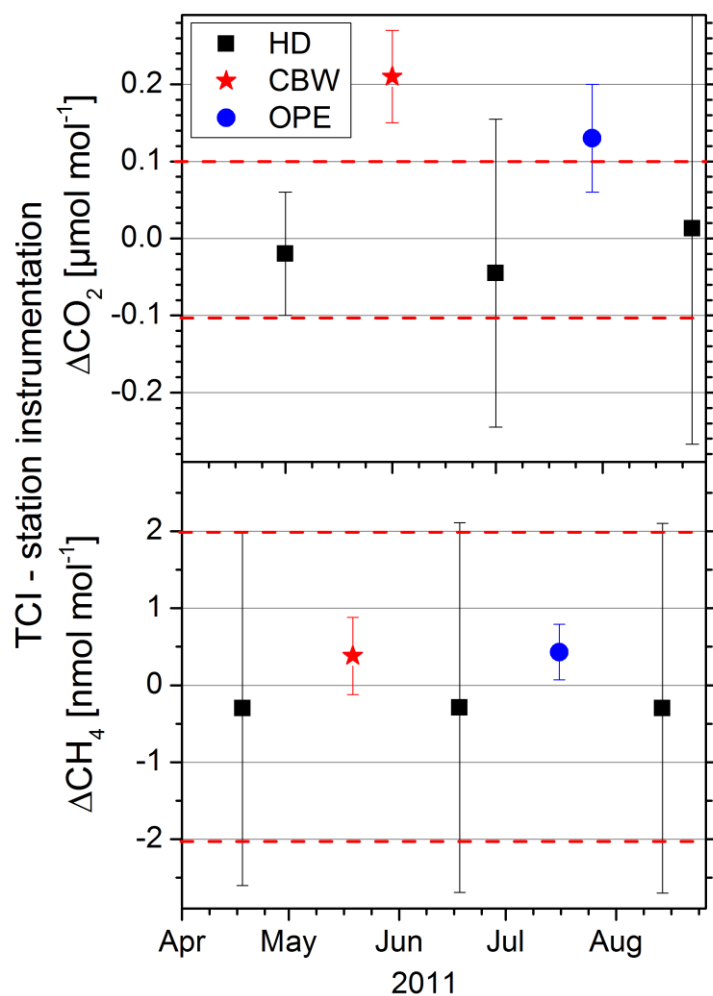
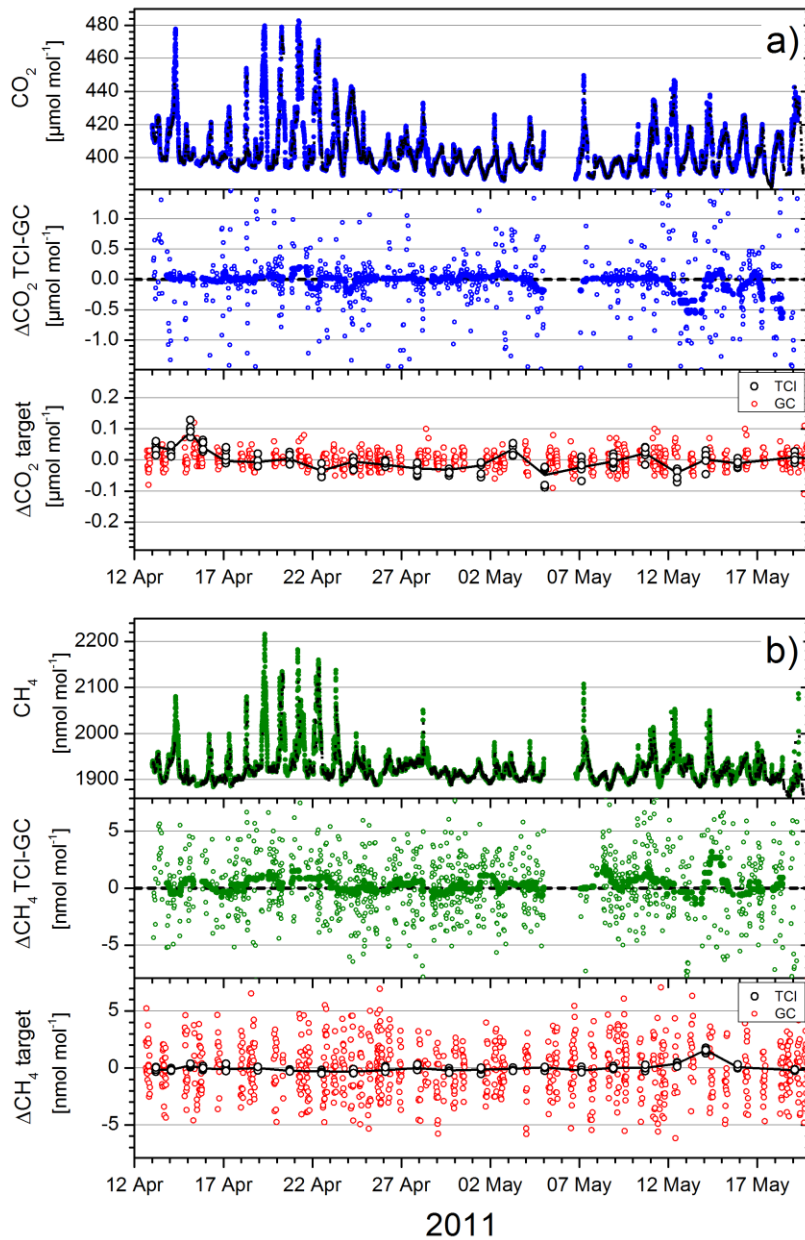


Figure 6: Summary of the differences of atmospheric measurements and their 1σ standard deviations (Gauss fit) between the TCI and the respective stations for CO₂ (upper panel) and CH₄ (lower panel).



1

2 *Figure A1: Upper panels of a) and b): Comparison of GC (black) and TCI (blue/green) of*
 3 *CO₂ or CH₄ measurements performed in Heidelberg. Middle panels: individual CO₂ or CH₄*
 4 *differences between the buffered GC and the smoothed TCI measurements (open symbols);*
 5 *Averaged differences (15h running median filter) are shown as filled symbols. The smoothing*
 6 *window is larger, compared to the ICOS Demonstration Sites, to account for the lower*
 7 *comparison frequency of the individual measurements. Lower panels: Individual CO₂ or CH₄*
 8 *GC target injections in red and individual TCI target measurements in black.*

9