Atmospheric
Measurement
Techniques

# Assessment of recent advances in measurement techniques for atmospheric carbon dioxide and methane observations

Christoph Zellweger[1], Lukas Emmenegger[1], Mohd Firdaus[2], Juha Hatakka[3], Martin Heimann[4,9], Elena Kozlova[5], T. Gerard Spain[6], Martin Steinbacher[1], Marcel V. van der Schoot[7], and Brigitte Buchmann[8]

[1]Empa, Swiss Federal Laboratories for Materials Science and Technology, Laboratory for Air Pollution/Environmental Technology, 8600 Dübendorf, Switzerland
[2]Atmospheric Science and Cloud Seeding Division, Malaysian Meteorological Department, Ministry of Science, Technology and Innovation, Kuala Lumpur, Malaysia
[3]Finnish Meteorological Institute, Helsinki, Finland
[4]Max Planck Institute for Biogeochemistry, Jena, Germany
[5]College of Life and Environmental Sciences, University of Exeter, Exeter, UK
[6]National University of Ireland, Galway, Ireland
[7]Oceans and Atmosphere, Commonwealth Scientific and Industrial Research Organisation, Aspendale, Victoria, Australia
[8]Empa, Swiss Federal Laboratories for Materials Science and Technology, Department Mobility, Energy and Environment, 8600 Dübendorf, Switzerland
[9]Division of Atmospheric Sciences, Department of Physics, University of Helsinki, Helsinki, Finland

*Correspondence to:* Christoph Zellweger (christoph.zellweger@empa.ch)

**Abstract.** Until recently, atmospheric carbon dioxide ($CO_2$) and methane ($CH_4$) measurements were made almost exclusively using nondispersive infrared (NDIR) absorption and gas chromatography with flame ionisation detection (GC/FID) techniques, respectively. Recently, commercially available instruments based on spectroscopic techniques such as cavity ring-down spectroscopy (CRDS), off-axis integrated cavity output spectroscopy (OA-ICOS) and Fourier transform infrared (FTIR) spectroscopy have become more widely available and affordable. This resulted in a widespread use of these techniques at many measurement stations. This paper is focused on the comparison between a CRDS "travelling instrument" that has been used during performance audits within the Global Atmosphere Watch (GAW) programme of the World Meteorological Organization (WMO) with instruments incorporating other, more traditional techniques for measuring $CO_2$ and $CH_4$ (NDIR and GC/FID). We demonstrate that CRDS instruments and likely other spectroscopic techniques are suitable for WMO/GAW stations and allow a smooth continuation of historic $CO_2$ and $CH_4$ time series. Moreover, the analysis of the audit results indicates that the spectroscopic techniques have a number of advantages over the traditional methods which will lead to the improved accuracy of atmospheric $CO_2$ and $CH_4$ measurements.

## 1 Introduction

Long-term observations of atmospheric greenhouse gases (GHGs) such as carbon dioxide ($CO_2$) and methane ($CH_4$) are crucial for the understanding of regional and global GHG budgets and their evolution. This requires data sets traceable to a common reference. Central Calibration Laboratories (CCLs) operate within the Global Atmosphere Watch (GAW) programme of the World Meteorological Organization (WMO) to provide measurement standards on the international calibration scales. The WMO/GAW programme strives to achieve ambitious compatibility goals that enable scientific interpretation of continental- or global-scale atmospheric observations measured by different laboratories or in situ stations. The compatibility goals apply to the gas mole

fraction ranges observed in the unpolluted troposphere. Currently, these goals stand at $\pm 0.1$ ppm for $CO_2$ ($\pm 0.05$ ppm for the Southern Hemisphere) and $\pm 2$ ppb for $CH_4$, whilst the extended goals of $\pm 0.2$ ppm and $\pm 5$ ppb, respectively, apply to measurements in more polluted environments (WMO, 2014). Additionally, the traceability of the measurements at GAW stations to the international calibration scales is evaluated by regular system and performance audits by the designated World Calibration Centres (WCCs).

Continuous measurements of $CO_2$ have been available since the 1950s (Harris, 2010; Keeling, 1960), and global methane coverage through direct measurements became available in the late 1970s (Khalil and Rasmussen, 1983; Kirschke et al., 2013). Until only recently, atmospheric $CO_2$ measurements have been made almost exclusively by nondispersive infrared (NDIR) absorption technique (Komhyr et al., 1989) whilst $CH_4$ has been measured by gas chromatography equipped with flame ionisation detectors (GC/FID) (Dlugokencky et al., 1995). Over the past few years, spectroscopic techniques such as direct absorption spectroscopy (McManus et al., 2015), cavity ring-down spectroscopy (CRDS) (Chen et al., 2010; Crosson, 2008), cavity enhanced off-axis integrated cavity output spectroscopy (OA-ICOS) (O'Shea et al., 2013) and Fourier transform infrared (FTIR) spectroscopy (Griffith et al., 2012) have become commercially available for the measurements of atmospheric $CO_2$ and $CH_4$. Currently, many traditional NDIR $CO_2$ and GC/FID $CH_4$ systems are being replaced by modern spectroscopic instruments (Brailsford et al., 2012). These new techniques have some clear advantages concerning sensitivity, precision, linearity, time response and the measurement setup. They further require less frequent calibration. However, there exist only a few published studies (Flores et al., 2015; Rella et al., 2013; Schibig et al., 2015; Vardag et al., 2014) comparing these modern measurement techniques with $CO_2$ NDIR or $CH_4$ GC/FID, and crucial information is still lacking to demonstrate that the former can guarantee a smooth continuation of historic and ongoing time series.

Our paper presents a set of comparison experiments of NDIR $CO_2$ and GC/FID $CH_4$ measurements with a CRDS travelling instrument that were made as part of the system and performance audits of the World Calibration Centre for Surface Ozone, Carbon Monoxide, Methane and Carbon Dioxide (WCC-Empa) (Buchmann et al., 2009). In addition, we present $CO_2$ and $CH_4$ comparisons between an OA-ICOS and the travelling instrument. The concept of using the travelling instrument for a comprehensive assessment of atmospheric measurements is encouraged by the recommendations of the WMO/IAEA Meetings on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (WMO, 2012, 2014) and has shown to a highly valuable tool for quality control (Hammer et al., 2013; Zellweger et al., 2013). Our comparison experiments presented here were conducted at four stations within the GAW network which cover different climatological conditions ranging from tropical to subarctic conditions characterised by varied atmospheric water vapour contents. We present and discuss the influence of water vapour on the quality of the atmospheric measurements as most of our measurement campaigns were conducted without drying of the ambient air samples for the CRDS travelling instrument. Furthermore, we investigate the impact of data coverage on hourly averaged data, which represents the standard aggregation period for data submission to most data repositories. Then, we examine the CRDS data with regard to the repeatability of calibration cylinder measurements and discuss calibration strategies. Finally, we analyse the data collected during $CH_4$ and $CO_2$ performance audits over the past few years from the perspective of the measurement techniques used to obtain them.

## 2 Experiment

The quality assurance strategy of the GAW programme comprises system and performance audits (hereafter only called audit) carried out by WCCs. WCC-Empa is the designated WCC for $CH_4$ (since 2000) and $CO_2$ (since 2010) audits. The performance audits conducted by WCC-Empa are made using two different approaches. The first method, which is described in Sect. 2.1 below, is based on the comparison of travelling standards (calibrated standard gases). This method has been an integral part of all performance audits made by WCC-Empa since we started this activity in 1995. In addition to the comparisons of travelling standards, a second approach by parallel measurements using a travelling instrument was implemented more recently. The latter approach, which is described in more detail in Sect. 2.2, was introduced after it was recognised that standard comparisons alone often lack important sources of potential biases, for example effects in the air inlet system.

### 2.1 Performance audit using travelling standards

The concept of the audit procedure using travelling standards has been described in detail elsewhere (Buchmann et al., 2009; Klausen et al., 2003). In brief, an audit involves the comparison of travelling standards (i.e. compressed gas in high pressure cylinders) on the analytical system of the audited station (WMO, 2011b). The travelling standards are calibrated against primary laboratory reference standards traceable to the CCL before and after the audit. The audited station's personnel analyse the travelling standards and report the mole fractions, which are compared to the values assigned by the WCC. The result is analysed by a linear regression between the reference (WCC) and the station values. For the calibration of the travelling standards at WCC-Empa, a GC/FID (Varian 3800) system was used from 2000 to 2009 for $CH_4$; since 2009 a CRDS (Picarro Inc., G1301 $CO_2$/$CH_4$/$H_2O$ analyser) has been used for both $CH_4$ and $CO_2$ calibrations. Several standards of the CCL

**Table 1.** $CO_2$ performance audits using travelling standards from 2010 to 2015.

| Station | GAW ID | Year | Instrument | Method | Intercept (ppm) | Slope (–) | Bias at 405 ppm $CO_2$ (ppm) |
|---|---|---|---|---|---|---|---|
| Lauder | LAU | 2010 | FTIR | FTIR | −2.48 | 1.00660 | 0.19 |
| Cape Point | CPT | 2011 | Hartmann & Braun URAS 4 | NDIR | 4.65 | 0.98813 | −0.16 |
| Zugspitze | ZSF | 2011 | HP6890 | GC/FID | 2.83 | 0.99286 | −0.06 |
| Hohenpeissenberg | HPB | 2011 | Picarro G1301 | CRDS | −0.09 | 0.99996 | −0.11 |
| Bukit Koto Tabang | BKT | 2011 | Picarro G1301 | CRDS | 2.81 | 0.99285 | −0.09 |
| Pallas | PAL | 2012 | Picarro G2401 | CRDS | 0.85 | 0.99781 | −0.04 |
| Pallas | PAL | 2012 | LI-COR LI-7000 | NDIR | 0.62 | 0.99863 | 0.07 |
| Zeppelin Mountain | ZEP | 2012 | Picarro G2401 | CRDS | −0.25 | 1.00120 | 0.24 |
| Zeppelin Mountain | ZEP | 2012 | LI-COR LI-7000 | NDIR | 3.59 | 0.99000 | −0.46 |
| Cabo Verde | CVO | 2012 | LGR GGA-24EP | OA-ICOS | 1.28 | 0.99690 | 0.02 |
| Cabo Verde | CVO | 2012 | Siemens Ultramat 6F | NDIR | 0.17 | 0.99970 | 0.05 |
| Mace Head | MHD | 2013 | Picarro G1301 | CRDS | 0.90 | 0.99785 | 0.03 |
| Mace Head | MHD | 2013 | Picarro G2301 | CRDS | 1.16 | 0.99725 | 0.05 |
| Izaña | IZO | 2013 | LICOR LI-7000 | NDIR | 1.90 | 0.99521 | −0.04 |
| Izaña | IZO | 2013 | LICOR LI-6252 | NDIR | −4.02 | 1.01038 | 0.18 |
| Danum Valley | DMV | 2013 | LoFlo Mark II | NDIR | 1.64 | 0.99588 | −0.03 |
| Bukit Koto Tabang | BKT | 2014 | Picarro G1301 | CRDS | 0.91 | 0.99742 | −0.13 |
| Anmyeon-do | AMY | 2014 | Picarro G2301 | CRDS | −0.18 | 1.00079 | 0.14 |
| Jungfraujoch | JFJ | 2015 | Picarro G2401 | CRDS | 0.10 | 0.99975 | 0.00 |
| Jungfraujoch | JFJ | 2015 | SICK MAIHAK S710 | NDIR | −3.62 | 1.00907 | 0.06 |

(NOAA/ESRL, National Oceanic and Atmospheric Administration/Earth System Research Laboratory) are used as reference standards at WCC-Empa ensuring traceability to the CCL.

For the current study we analysed performance audit results for methane (2005–2014) and carbon dioxide (2010–2015). Details of the comparisons including instruments and analytical techniques are given in Table 1 for $CO_2$ and Table 2 for $CH_4$. In order to assess the performance of the individual comparisons in a standardised way, the bias in the centre of the mole fraction range (405 ppm for $CO_2$, 1900 ppb for $CH_4$) of the unpolluted troposphere (WMO, 2014) (360–450 ppm for $CO_2$, 1700–2100 ppb for $CH_4$) was calculated for these comparisons based on the linear regression analysis. This allows displaying the result of a performance audit using travelling standards as a single dot in a bias vs. slope plot, as illustrated in Fig. 1 for the example of $CO_2$ audits. The green dashed line in the left panel of Fig. 1 shows a case with no bias at 405 ppm $CO_2$ but with the corresponding minimal slope that is possible for the data still meeting the data quality objective (DQO) of 0.1 ppm in the range of 360–450 ppm $CO_2$. This case translates to a single point in the bias vs. slope plot, as shown by the green dot in the left panel of Fig. 1. For illustrative purpose, two additional cases are shown: the maximum allowed bias with the corresponding slope that still meets the extended DQO of 0.2 ppm (orange dashed line/dot) and a case with a slope/bias combination that does not meet the DQOs (red dashed line/dot) over the entire relevant mole fraction range.
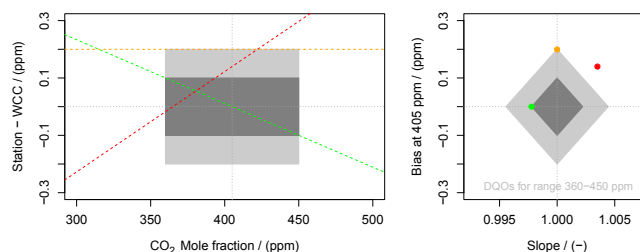


**Figure 1.** Left: deviation vs. reference value plot for $CO_2$ (illustrative) for three different cases (green, orange, red; details see text) for the mole fraction of 360–450 ppm $CO_2$. Right: illustrative bias vs. slope plot for the cases shown in the left panel (details see text). The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals.

## 2.2 Performance audit by parallel measurements with a travelling instrument

Quality assessments based on the comparison of travelling standards alone have their limitations, since they do not cover all parts of the analytical system that may bias a measurement, such as inlet and drying systems (WMO, 2011a). Furthermore, comparisons of travelling standards during on-site audits as well as round robin experiments are only snapshots and are potentially biased, e.g. by coincidental instrument malfunction (results worse compared to normal operation) or by extraordinary care taken during analysis (results better compared to normal operation). Therefore, it has been rec-

**Table 2.** $CH_4$ performance audits using travelling standards from 2005 to 2014.

| Station/laboratory | GAW ID | Year | Instrument | Method | Intercept (ppb) | Slope (–) | Bias at 1900 ppb $CH_4$ (ppb) |
|---|---|---|---|---|---|---|---|
| Ryori | RYO | 2005 | Horiba GA-360 | NDIR | −29.6 | 1.0157 | 0.29 |
| Japan Meteorological Agency | NA | 2005 | Shimadzu 14BPF | GC/FID | 2.4 | 0.9995 | 1.49 |
| Zugspitze | ZSF | 2006 | HP 6890 | GC/FID | 9.1 | 0.9933 | −3.66 |
| Jungfraujoch | JFJ | 2006 | Agilent 6890 | GC/FID | −9.6 | 1.0062 | 2.20 |
| Cape Point | CPT | 2006 | Varian CP-3800 | GC/FID | −47.0 | 1.0259 | 2.25 |
| Pallas | PAL | 2007 | Agilent 6890N | GC/FID | 15.3 | 0.9921 | 0.32 |
| Barrow | BRW | 2008 | HP 6890 | GC/FID | 38.2 | 0.9793 | −1.18 |
| Izaña | IZO | 2009 | DANI-3800 | GC/FID | 9.1 | 0.9950 | −0.31 |
| Mt. Waliguan | WLG | 2009 | HP 5890 | GC/FID | 3.0 | 0.9976 | −1.53 |
| Mt. Waliguan | WLG | 2009 | Agilent 6890 | GC/FID | 0.1 | 1.0001 | 0.31 |
| Mt. Waliguan | WLG | 2009 | Picarro G1301 | CRDS | 3.7 | 0.9977 | −0.77 |
| GAW calibration lab Beijing | NA | 2009 | Agilent 6890N | GC/FID | 28.5 | 0.9843 | −1.34 |
| GAW calibration lab Beijing | NA | 2009 | Agilent 6890N | GC/FID | 13.8 | 0.9936 | 1.61 |
| GAW calibration lab Beijing | NA | 2009 | Picarro G1301 | CRDS | 0.8 | 1.0008 | 2.27 |
| Mace Head | MHD | 2009 | CARLE 100A | GC/FID | 0.7 | 0.9998 | 0.24 |
| Lauder | LAU | 2010 | FTIR | FTIR | −10.2 | 1.0060 | 1.20 |
| Cape Point | CPT | 2011 | Varian CP-3800 | GC/FID | −34.9 | 1.0202 | 3.46 |
| Zugspitze | ZSF | 2011 | HP6890 | GC/FID | 9.2 | 0.9947 | −0.92 |
| Hohenpeissenberg | HPB | 2011 | Picarro G1301 | CRDS | −0.2 | 1.0003 | 0.33 |
| Bukit Koto Tabang | BKT | 2011 | Picarro G1301 | CRDS | −0.6 | 1.0000 | −0.57 |
| Pallas | PAL | 2012 | Picarro G2401 | CRDS | 12.4 | 0.9929 | −1.05 |
| Zeppelin Mountain | ZEP | 2012 | Picarro G2401 | CRDS | 11.3 | 0.9939 | −0.25 |
| Mt. Cimone | CMN | 2012 | Agilent 6890N | GC/FID | 45.4 | 0.9764 | 0.64 |
| Cabo Verde | CVO | 2012 | LGR GGA-24EP | OA-ICOS | 15.0 | 0.9917 | −0.86 |
| Mace Head | MHD | 2013 | CARLE 100A | GC/FID | 4.0 | 0.9977 | −0.33 |
| Mace Head | MHD | 2013 | Picarro G1301 | CRDS | 8.2 | 0.9954 | −0.48 |
| Mace Head | MHD | 2013 | Picarro G2301 | CRDS | 11.9 | 0.9937 | −0.03 |
| Izaña | IZO | 2013 | DANI 3800 | GC/FID | −11.0 | 1.0064 | 1.12 |
| Izaña | IZO | 2013 | Varian 3800 | GC/FID | −8.0 | 1.0043 | 0.17 |
| Bukit Koto Tabang | BKT | 2014 | Picarro G1301 | CRDS | 0.0 | 0.9999 | −0.23 |
| Anmyeon-do | AMY | 2014 | Picarro G2301 | CRDS | 8.0 | 0.9955 | −0.63 |
| Jungfraujoch | JFJ | 2015 | Picarro G2401 | CRDS | 1.9 | 0.9992 | 0.36 |

ommended that the quality control procedure during on-site audits should include parallel measurements with a travelling instrument whenever feasible (WMO, 2011a, 2012, 2014).

The concept of the practical realisation of the on-site data comparison is illustrated in Fig. 2. The core part is a CRDS analyser (Picarro Inc., G2401) as travelling instrument (a). An independent inlet system (b) is used during the parallel measurement with the travelling instrument, and, if feasible, the travelling instrument samples from the station and the independent inlet system sequentially (c). Furthermore, the travelling instrument is independently calibrated using its own set of standards (d), which normally comprises a subset of the travelling standards used for the audit. For further confirmation of the compatibility of the two systems, the travelling standards (e) are measured on both the station analyser and the travelling instrument. Dry air mole fractions are compared for both the travelling standards and the travelling instrument comparisons.

For this study we used the data from two Picarro G2401 $CO/CH_4/CO_2/H_2O$ CRDS instruments (Picarro Inc., USA). The instruments were calibrated every 30–40 h using dry compressed air as working standard. In most cases the sample air was not dried prior to analysis with the travelling instrument and a humidity correction using the Empa method (Rella et al., 2013) was applied to all data. Experimental details of the Empa method are described in Zellweger et al. (2012). Briefly, a small amount of water (approximately 0.8 mL) was directly injected into a constant flow (approximately 500 mL min$^{-1}$) of a working standard which was delivered to the instrument. The resulting water vapour influence was then fitted by a quadratic function. The correction function was determined several times for each instrument, and a single function that was initially obtained was used to apply the correction. Yver Kwok et al. (2015) have recently published a comprehensive assessment of the performance of the Picarro G2401 analyser, and our setup of the travelling in-
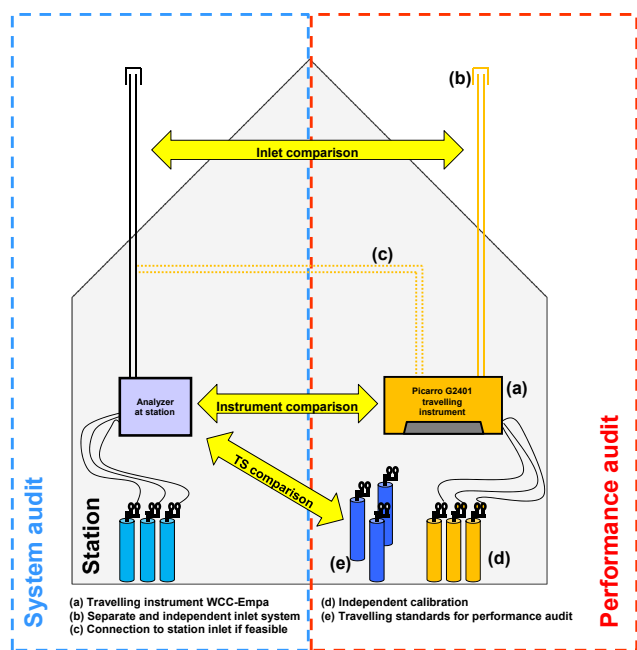
**Figure 2.** Schematic of the comparison procedure for the ambient air measurements during audits by WCC-Empa.

strument was done along the lines of their recommendations. However, in contrast to their approach, we were running calibrations more frequently but only as one cycle. Furthermore, the background signal of both travelling instruments was initially adjusted using so-called zero air ($CO_2$ and $CH_4$-free natural air). The following calibration strategy was used:

- A working standard (calibrated against certified CCL laboratory standards before and after each campaign) with mole fractions close to ambient air was analysed every 30–40 h.

- A LOESS fit was applied to these data.

- The ratio of the assigned working standard value to the LOESS fit was used to apply a drift correction to all data.

- To verify the calibration, two additional cylinders were measured as target standards. The same calibration and water vapour corrections as for ambient air were applied.

The following calibration scales were used: the $CO_2$ scale WMO-X2007 (Zhao and Tans, 2006) and the $CH_4$ scale WMO-X2004 (Dlugokencky et al., 2005).

An example of the working and target standard measurements is shown in Fig. 3. It can be seen that the variation of the target gas measurements did not exceed the range of ± half of the WMO compatibility goals of 0.1 ppm for $CO_2$ and 2 ppb for $CH_4$, which is usually required for

intra-laboratory repeatability (WMO, 2014). Furthermore, the maximum drift between two consecutive working standard measurements was always smaller than half the compatibility goal, indicating that calibrations were made with sufficient frequency. Similar stability was achieved during all measurement campaigns.

The optimal calibration frequency was further investigated with laboratory experiments. For this purpose, a gas standard (dry natural air) was continuously measured using one of the Picarro G2401 WCC-Empa travelling instruments. To prolong the length of the measurement period, the sample flow was reduced to $30 \, \text{mL} \, \text{min}^{-1}$ by a needle valve at the inlet port, which still allowed the stabilisation of the cavity pressure to 140 torr (186.6 hPa). The standard gas was calibrated against NOAA/ESRL standards before and after the experiment to ensure that no drift occurred over the observation period, which may happen when a standard gas is losing pressure (Leuenberger et al., 2015). The initial pressure of the standard was 12.96 MPa and dropped to 6.21 MPa at the end of the experiment. Figure 4 summarises the results of these measurements. The upper left panel shows the $CO_2$ variation of the standard gas (5 s raw data) during the experiment duration of 455 h (approx. 19 days). A slight upward drift was observed, which we consider as instrumental drift since the (secondary) standard has been proven to be stable with respect to the NOAA standards over the course of the experiment. The variations for methane are shown in the upper right panel; again, instrument drift was observed but, in contrast to $CO_2$, no monotonic trend was detected. The lower panels of Fig. 4 show the Allan deviation for $CO_2$ and $CH_4$ (Werle et al., 1993) using the data above, which allows an estimate of the optimal calibration intervals. Based on this experiment, the optimal averaging time is approximately 20 min for both $CO_2$ and $CH_4$. This is shorter than the $CO_2$ minimum at 58 min that was found by Flowers et al. (2012) but compares well with the results of Yver Kwok et al. (2015). However, the Allan deviation only slightly increases up to $1 \times 10^5$ s (27.8 h); therefore, a calibration interval of 30 h is regarded as an optimal compromise with regard to stability, data coverage and the consumption of calibration gas.

On-site comparison experiments with the travelling instrument were made at the following GAW stations:

i. Danum Valley, Malaysia (DMV), operated by the Malaysian Meteorological Department, Kuala Lumpur, Malaysia, with support from Commonwealth Scientific and Industrial Research Organisation, Oceans & Atmosphere, Aspendale, Victoria, Australia;

ii. Pallas, Finland (PAL), operated by the Finnish Meteorological Institute, Helsinki, Finland;

iii. Cape Verde Atmospheric Observatory (CVO), operated by the National Institute of Meteorology and Geophysics, Cabo Verde, with support from the Max Planck
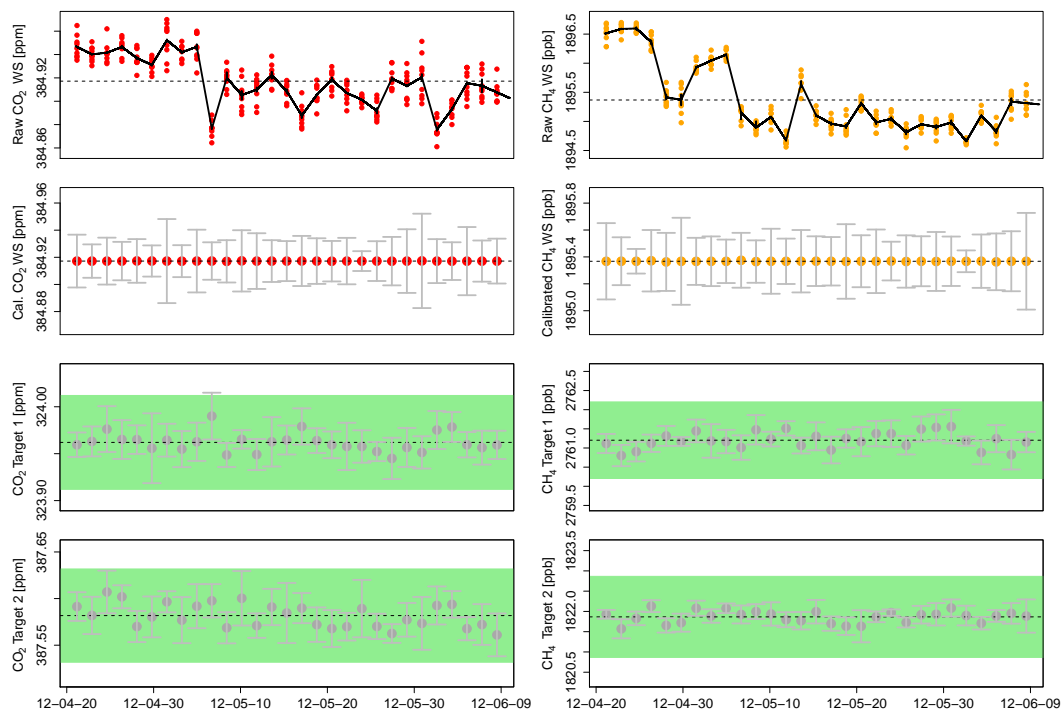
**Figure 3.** Left-hand side panels: $CO_2$ working and target cylinder measurements of the travelling instrument at PAL. Top panel: raw 1 min readings of the WS (red points) with LOESS fit (solid black line) and the mean reading (dotted grey line). Second panel from the top: average WS readings after calibration. Lower two panels: target cylinder measurements. The green area represents the average reading ± half of the WMO compatibility goal. Right-hand side panels: Same results for $CH_4$.
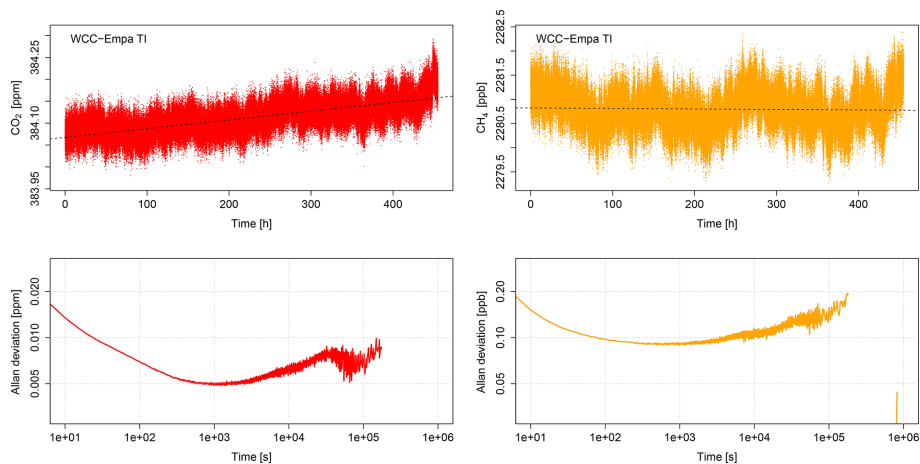


**Figure 4.** Upper left panel: $CO_2$ working standard measured over a period of 455 h with the Picarro G2401 (S/N 1497-CFKADS2098) travelling instrument (5 s averages). The black dashed line is the linear regression through all data. Lower left panel: Allan deviation plots based on the data above. Right: same for $CH_4$.

Institute for Biogeochemistry, Jena, Germany, and the University of York, United Kingdom;

iv. Mace Head, Ireland (MHD), operated by the National University of Ireland, Galway, Ireland.

All comparison experiments with the WCC-Empa travelling instrument were made using a separate inlet system, i.e. a separate air sampling tubing line (1/4″ OD Synflex 1300 or 1/2″ OD at DMV), leading to the same air intake location as the station. This WCC-Empa inlet line was flushed by an additional pump at a flow rate of approximately 2 L min$^{-1}$. At PAL, the travelling instrument switched occasionally to the single station inlet. CVO has two separate inlets: reactive gases are sampled 8 m above the ground on top of

**Table 3.** Overview of $CO_2$ and $CH_4$ comparison experiments presented in this study.

| Location | Coordinates | Start | End | Station instrument | Travelling instrument | Compound |
|---|---|---|---|---|---|---|
| PAL | 67.973° N, 24.116° E | 31 Apr 2012 | 9 Jun 2012 | LI-COR LI-7000 | CFKADS2001 humid meas. | $CO_2$ |
| DMV | 4.981° N, 117.844° E | 6 Dec 2013 | 25 Feb 2014 | LoFlo Mark II | CFKADS2098 humid meas. | $CO_2$ |
| CVO | 16.864° N, 24.868° W | 12 Dec 2012 | 4 Feb 2013 | LGR GGA-24EP | CFKADS2001 dry meas. | $CO_2/CH_4$ |
| MHD | 53.325° N, 9.900° W | 24 Jul 2013 | 27 Aug 2013 | CARLE 100A GC/FID | CFKADS2098 humid meas. | $CH_4$ |

the measurement container, whereas the GHG inlet is located on top of a tower 30 m above the ground. Comparisons with the WCC travelling instrument at CVO were performed by switching occasionally to the inlet for reactive gases. It should be noted that only comparisons were selected where neither the WCC-Empa travelling instrument nor the station analyser had instrumental problems. Furthermore, comparisons between the travelling instrument and other CRDS analysers are not shown in this paper, since the scope of the current work focuses on the comparison of CRDS with those techniques that have been widely used in the past. Results of the WCC-Empa travelling instrument and another CRDS instrument at PAL were published by Rella et al. (2013). An overview of the comparisons, including duration and instruments, is presented in Table 3. More information on the stations is available from the GAW Station Information System (GAWSIS, 2016).

## 3 Results

The $CO_2$ and $CH_4$ ambient air comparison experiments selected for this study were carried out at four GAW stations (cf. Table 3). The selected sites span a range of climatologies from tropical to subarctic conditions. Furthermore, $CO_2$ and $CH_4$ variability were distinctly different, ranging from remote baseline conditions at CVO with almost no temporal variation to highly variable conditions due to atmosphere–biosphere exchange processes at DMV. Figure 5 shows the diurnal variations for $CO_2$ and $CH_4$ as well as the frequency distribution of $CO_2$, $CH_4$ and $H_2O$ measured by the travelling instrument during the 1–2-month-long campaigns. Almost no or little diurnal variation was observed at the remote stations CVO and PAL, whereas the measurements at MHD and DMV showed significantly more variability because of sporadic signals from $CO_2$ and $CH_4$ source regions and due to vegetation uptake and respiration. The selected campaigns also cover different atmospheric water vapour contents, ranging from dry subarctic conditions at PAL ($H_2O < 1\%$) to temperate (MHD, $H_2O$ 1–2 %) and

tropical (DMV, $H_2O > 2\%$) conditions. At CVO, the air sampled with the travelling instrument was dried with a Nafion dryer, because carbon monoxide (CO) was also studied during this particular campaign (not shown here), and the water vapour correction for CO at that time and for this specific instrument prevented sufficiently precise humid CO measurements to be made by the travelling instrument (Zellweger et al., 2012).

### 3.1 Carbon dioxide ambient air comparisons

Figures 6–8 show $CO_2$ comparisons made at the PAL, DMV and CVO GAW stations. In the upper panels, the comparison with the highest available common time resolution (1 min) is shown, together with the deviation of the station instrument compared to the travelling instrument as function of time, and a histogram of the observed bias. The middle panels show the same but for hourly aggregated values where all available data were considered. The lower panels also show hourly aggregates, but mean values were calculated using only high-resolution data with concurrent data availability of the travelling and station instruments.

The temporal variation was well captured at all stations even at the highest time resolution of 1 min. To account for different residence times in the inlet system, data of the travelling instrument were slightly shifted (up to 53 s) to obtain the best possible agreement between the time series. The mean $CO_2$ bias based on 1 min data was $0.08 \pm 0.06$ ppm at PAL, $0.01 \pm 0.67$ ppm at DMV and $0.06 \pm 0.08$ ppm at CVO. These deviations are in good agreement with the results obtained when WCC-Empa travelling standards were measured on the station analysers as summarised in Fig. 9, where the bias is shown for individual travelling standards (black dots), including a linear regression analysis with 95 % confidence intervals. The resulting deviation is very close to the bias observed during the ambient air comparison, which is also shown as small red points (hourly data) in Fig. 9. Normally, the performance audit covers a wider mole fraction range compared to the ambient variability of a station. This gives
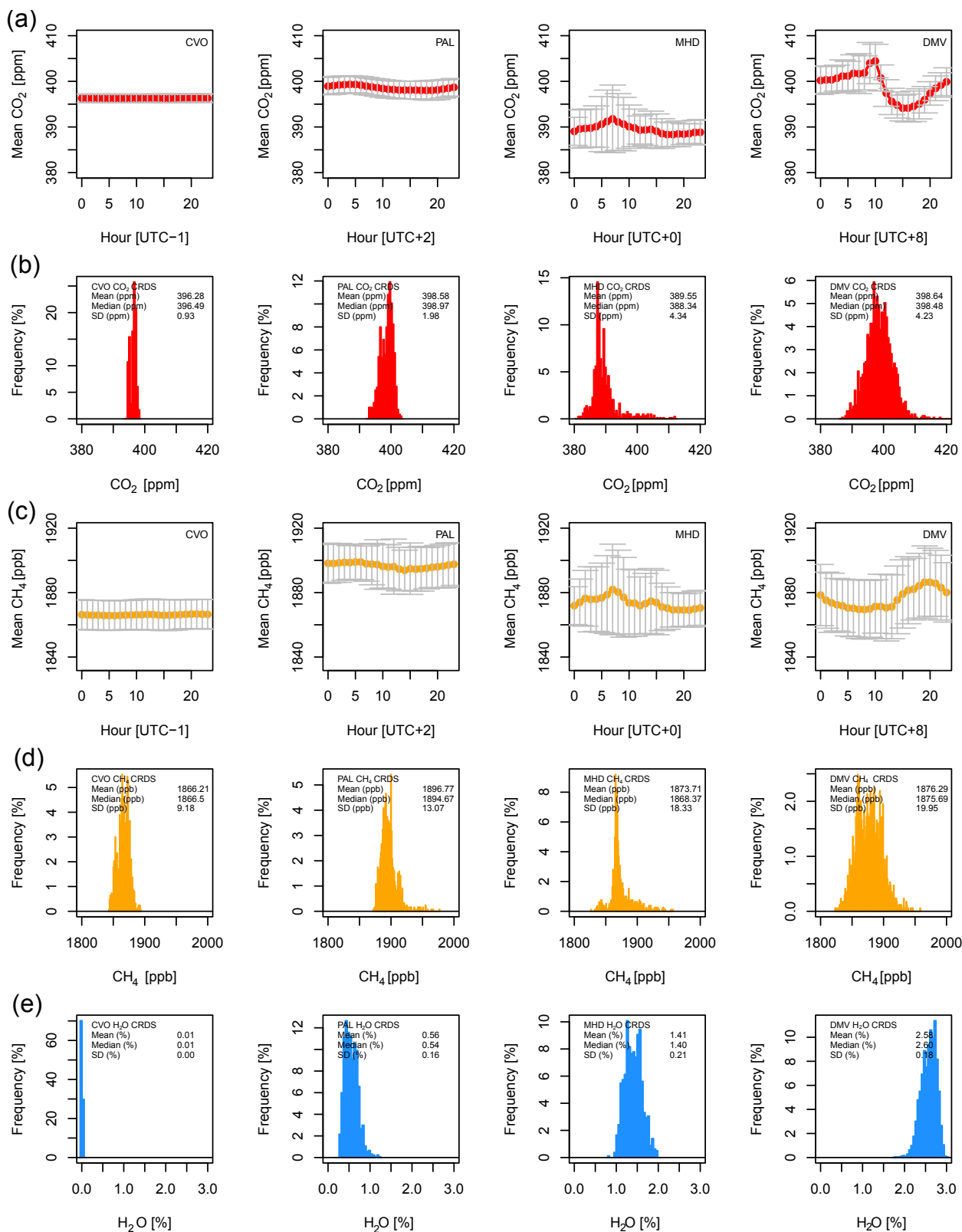
**Figure 5. (a)** Mean $CO_2$ diurnal variation measured during the comparison campaigns at CVO, PAL, MHD and DMV with the travelling instrument. The error bars are the standard deviation of each hourly value. **(b)** Frequency distribution of hourly $CO_2$ mole fractions (bin size 0.5 ppm). **(c)** Same as **(a)** for $CH_4$. **(d)** Same as **(b)** for $CH_4$, bin size 1 ppb. **(e)** Frequency distribution of the hourly $H_2O$ content of the atmosphere (bin size 0.05 %), except for CVO, where the $H_2O$ content after the Nafion dryer is shown.
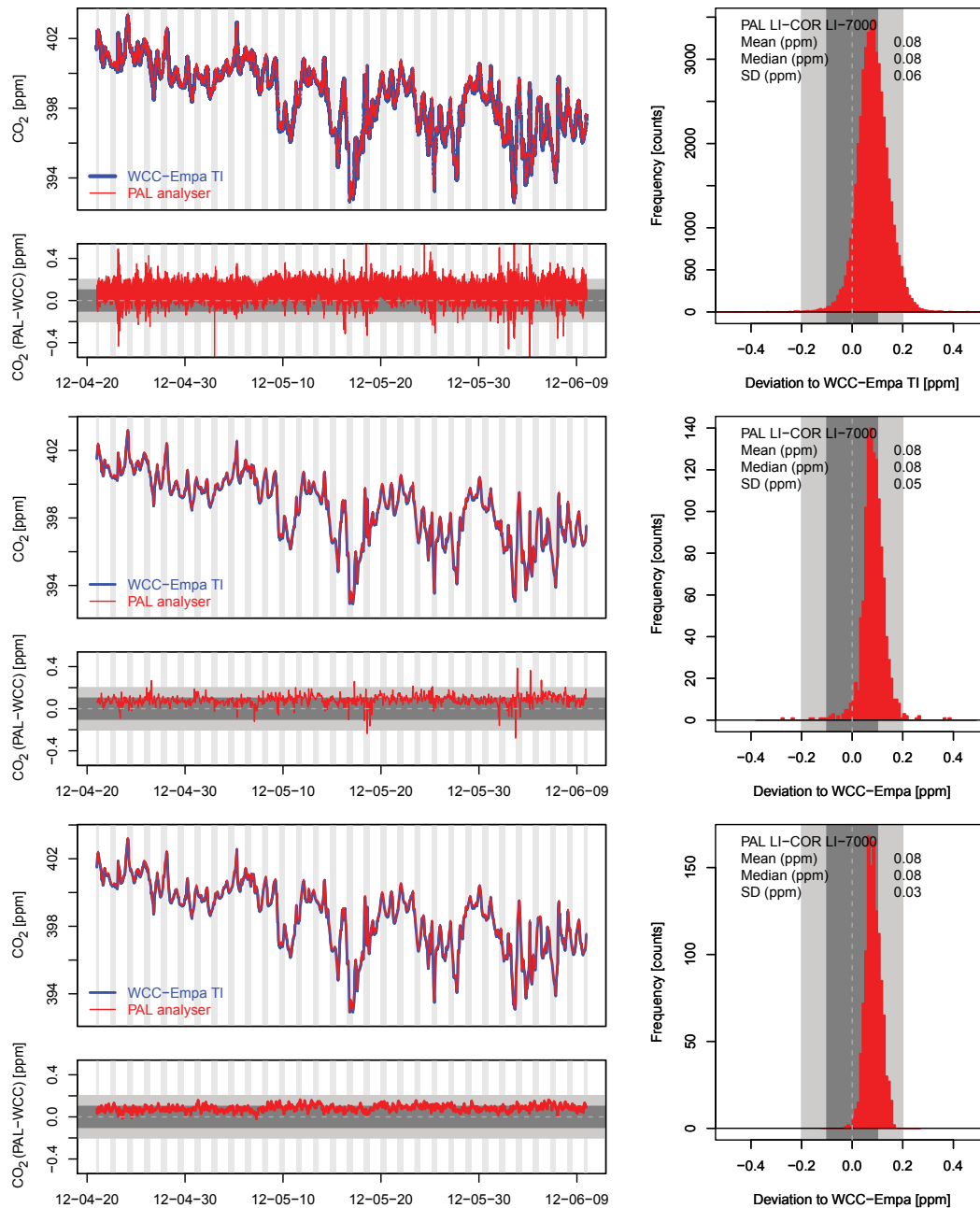
**Figure 6.** $CO_2$ comparison at PAL between the WCC-Empa travelling instrument and the PAL LI-COR LI-7000 instrument. Left: $CO_2$ time series and $CO_2$ bias vs. time. Right: deviation histogram. Upper set: 1 min data; middle set: 1 h data, calculated from all available 1 min values; lower set: 1 h data, calculated from 1 min values with concurrent PAL and WCC-Empa data. The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals; vertical grey bars (left diagrams) illustrate when different inlets were used (see text for details).

valuable information about either the compensation of the instrument nonlinearity, the consistency of the used standards or a combination of both, which would not be available from the ambient air comparison alone. The above results indicate that the nonlinearity of the analysers was well corrected at PAL and DMV. In contrast, the travelling standard comparison shows a larger bias at CVO. Most likely this is due to the fact that the travelling standards are significantly out of the CVO calibration range, which is narrow in response to the small variability of the $CO_2$ mole fraction at this station.

The mean agreement and standard deviation remained almost unchanged after aggregation to hourly values at all sta-
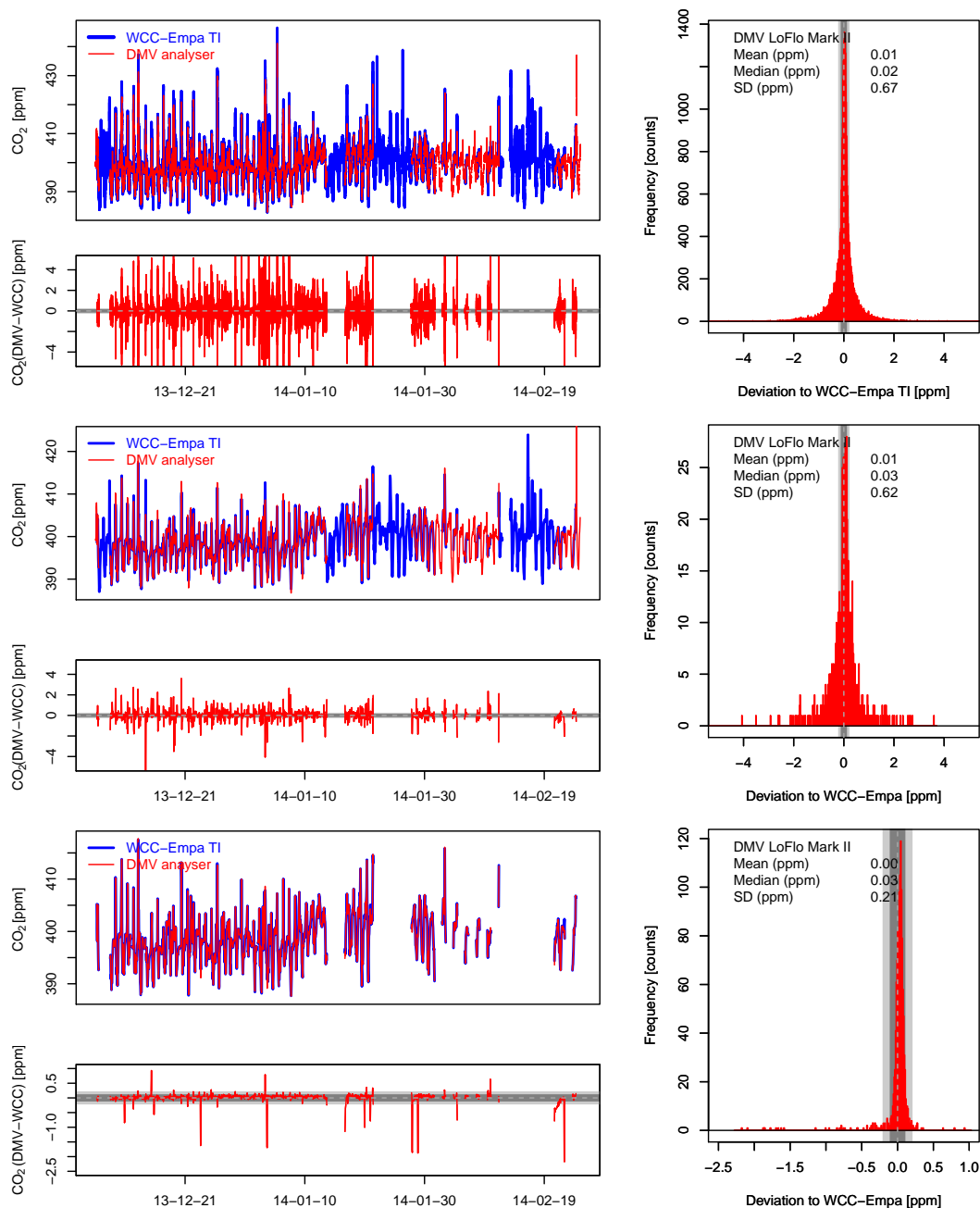
**Figure 7.** $CO_2$ comparison at DMV between the WCC-Empa travelling instrument and the DMV LoFlo Mark II instruments. Left: $CO_2$ time series and $CO_2$ bias vs. time. Right: deviation histogram. Upper set: 1 min data; middle set: 1 h data, calculated from all available 1 min values; lower set: 1 h data, calculated from 1 min values with both DMV and WCC-Empa data coverage. The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals.

tions, indicating that the procedures for time synchronisation and sample residence time correction were appropriate. The data availability was different for the travelling instrument and the station instruments mainly due to different requirements concerning calibration frequency of the station analysers. The CRDS travelling instrument was normally calibrated every 30 h for 45 min, which results in a very high

data availability and a uniform data coverage. In contrast, the NDIR $CO_2$ instruments at PAL and DMV require more frequent calibrations, which have been implemented using different approaches. The data availability was more or less homogeneous over time at PAL, whereas at DMV the hourly average $CO_2$ mole fraction value is calculated using only the final 44 min of each hour due to the automated hourly zero
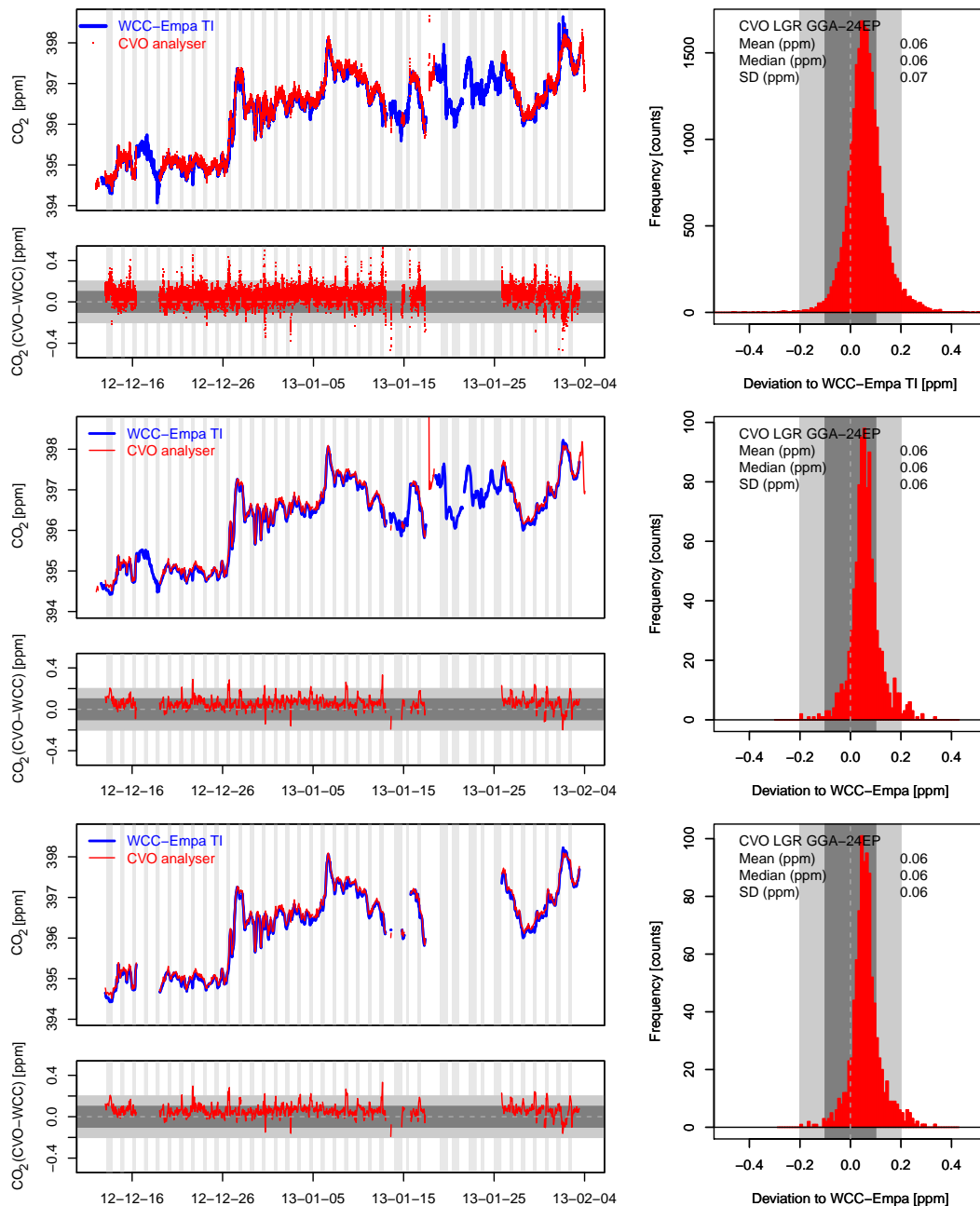
**Figure 8.** $CO_2$ comparison at CVO between the WCC-Empa travelling instrument and the CVO LGR GGA-24EP instruments. Left: $CO_2$ time series and $CO_2$ bias vs. time. Right: deviation histogram. Upper set: 1 min data; middle set: 1 h data, calculated from all available 1 min values; lower set: 1 h data, calculated from 1 min values with concurrent CVO and WCC-Empa data from both inlets. The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals; vertical grey bars (left diagrams) illustrate when different inlets were used (see text for details).

drift correction mode employed at the start of every hour to monitor short-term detector drifts. For OA-ICOS instrument at CVO, a scheme alternating between a working standard and ambient air measurement in an interval of 1 min was in place resulting in a homogenous data availability of every second minute. This has been done to account for short-term

drifts. The data availability of the station analysers and the travelling instrument as a function of the minute of the hour as well as the minute of the day is shown in Fig. 10. Data coverage becomes extremely important with rapid atmospheric changes, e.g. during flux measurements (Peltola et al., 2014), and at sites influenced by local biospheric processes.
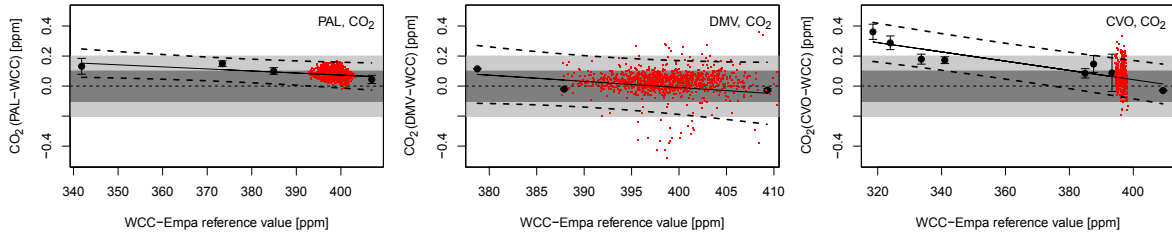
**Figure 9.** Results of the $CO_2$ travelling standard comparisons (performance audits) at PAL (left), DMV (middle) and CVO (right). The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals. The red points correspond to the observed differences based on hourly data during the ambient air comparison. Solid and dashed lines represent the fit and the 95 % confidence intervals for the linear regression through the travelling standard comparison.
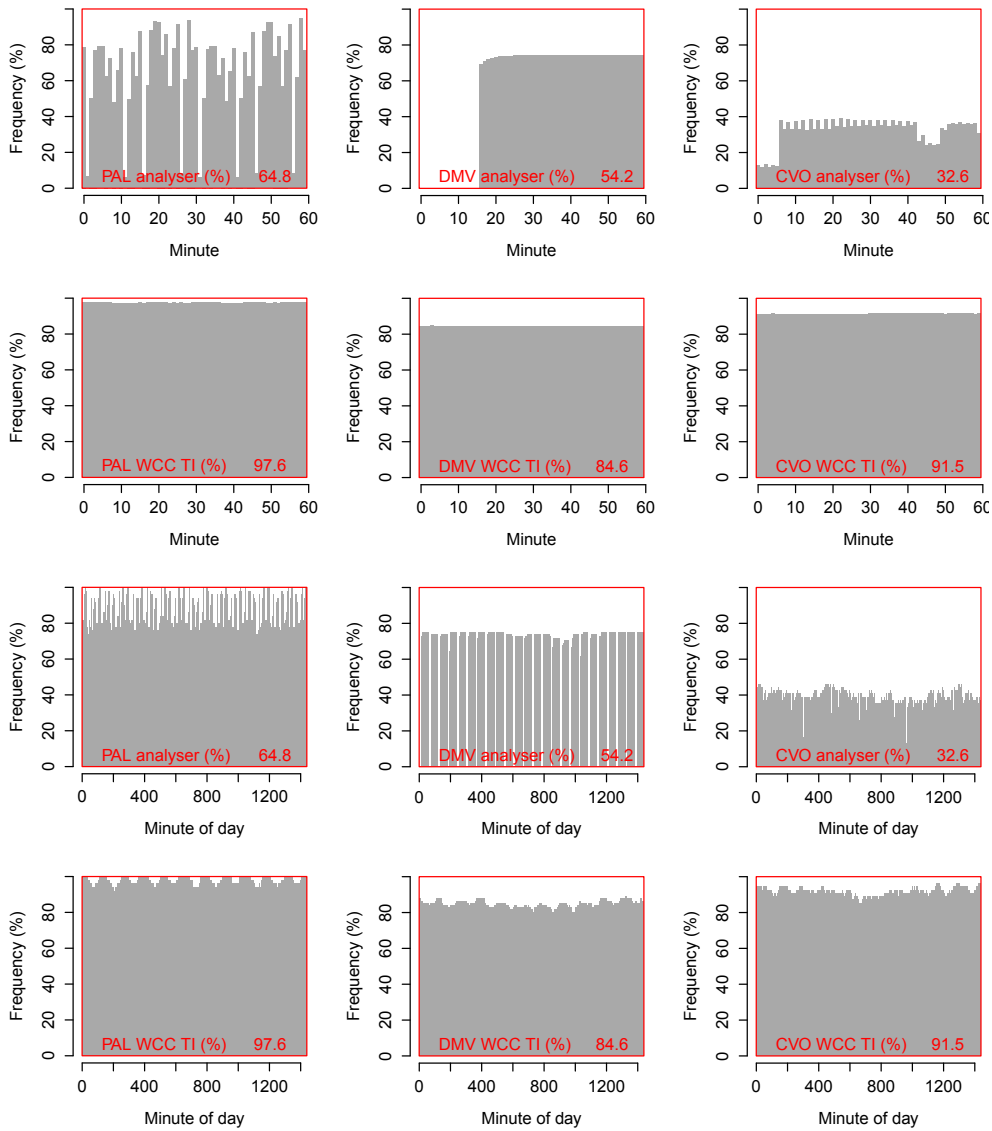


**Figure 10.** First row: mean $CO_2$ data availability of the PAL, DMV and CVO station instruments as a function of the minute of the hour. Second row: same as above, for the WCC-Empa travelling instrument. Third row: mean $CO_2$ data availability for the station instruments as a function of the minute of the day. Fourth row: same as above, for the WCC-Empa travelling instrument.

The influence of the data coverage was assessed by the calculation of hourly averages with concurrent data availability for the highest time resolution, which is shown in the lower panels of Figs. 6–8. As expected, no influence was found for the CVO data series, which is due to the combination of very small ambient air variability and homogeneous data coverage of the CVO analyser. Some effect was observed at PAL, where the standard deviation of the mean bias changes from 0.05 to 0.03 ppm if only concurrent data points are considered for averaging. It should be noted that in this case all occurrences with a bias > 0.2 ppm disappear. These results can be compared to the simultaneous comparison exercise of the travelling instrument with an on-site CRDS instruments (Rella et al., 2013), where also a mean standard deviation of the difference between the two instruments of 0.03 ppm was observed. As expected, the effect of data coverage on hourly mean values was the most pronounced at DMV due to the largest ambient variability and the omission of the first 16 min of data from each hourly calculation. The standard deviation of the mean bias decreased significantly from 0.62 to 0.21 ppm when only concurrent data were considered for the aggregation of hourly means. In order to test the influence of the temporal coverage, we also calculated the difference between hourly averages of the travelling instrument using all available data and only data of the travelling instrument with concurrent data of the DMV instrument. The resulting distribution of the differences looks very similar to the bias observed between the travelling instrument and the DMV analyser, with a standard deviation of 0.54 ppm. Therefore, the differences that we observe between the DMV analyser and the travelling instrument originate almost entirely from different temporal coverage. The bias induced by the data coverage can well exceed the extended WMO/GAW compatibility goals for hourly values in case of rather large variability of the ambient air $CO_2$ mole fraction. As reported here, homogeneous data coverage over the period of consideration can be more important than the absolute data availability. As an example, measurements made every second minute (CVO) better characterise an hourly average compared to a setup in which the first 16 min of each hour is not measured (DMV). Consequently, in the case of reduced data coverage due to the applied calibration scheme or analytical technique the resulting uncertainties cannot be neglected but should be evaluated and reported along with the uncertainty budget of the analysis.

### 3.2 Methane ambient air comparisons

Figures 11 and 12 show $CH_4$ comparison experiments made at the CVO and MHD GAW stations in the same format as Figs. 6–8 for $CO_2$. At CVO, the $CH_4$ data coverage is the same as for $CO_2$, since it is measured with the same OA-ICOS instrument (Los Gatos Research, LGR-GGA-24EP), whereas the GC/FID system (Carle 100A) at MHD analyses only two ambient air samples per hour, resulting in a maxi-

mal data availability of 3.3 % if single injections are being considered to be representative for 1 min. The actual data availability at MHD, however, was 2.2 % due to further instrument downtimes. The observed $CH_4$ bias between CVO and WCC-Empa is $-0.61 \pm 0.49$ ppb based on 1 min data, well within the WMO/GAW compatibility goal of $\pm 2$ ppb. Further averaging to hourly values slightly reduces the standard deviation of the bias to 0.34 ppb, whereas similarly to $CO_2$ no significant improvement is observed if only concurrent high-resolution data are considered for the hourly aggregate. The scatter of the observed bias is significantly larger for the MHD comparison, averaging to $-0.57 \pm 3.79$ ppb based on comparison with single injections of the MHD GC/FID with concurrent travelling instrument 1 min data. This is slightly larger compared to the difference that was observed during a 2-month comparison campaign by Vardag et al. (2014), when a mean bias to a FTIR of $-0.04 \pm 3.38$ ppb was observed. The standard deviation, however, is comparable, and a significant part of it can be attributed to the repeatability of the GC/FID system at MHD. The observed relative standard deviation under repeatability conditions (multiple injections of travelling standard) was 0.12 % during the audit, which results in a scatter of 2.25 ppb for the travelling instrument–GC/FID comparison. The aggregation of hourly values nearly doubles the scatter to $\pm 4.20$ ppb. This can be expected since two single injections per hour lead to poor statistics and do not sufficiently reflect the observed variability of ambient $CH_4$ at MHD.

As for $CO_2$, the results of the comparison of the travelling standards during audit were confirmed by $CH_4$ ambient air comparisons both at CVO and MHD. The deviations were in good agreement with the results of the travelling standard comparison. This is shown in Fig. 13, where the bias is plotted for individual travelling standard (black dots), including a linear regression analysis with 95 % confidence intervals. Additionally, the bias observed during the ambient air comparison (orange points in Fig. 13) agrees well with the bias determined by the performance audit. In the case of CVO, the observed scatter during the ambient air measurement was comparable with the 95 % confidence bands of the linear interpolation of the audit results, whereas the ambient air scatter at MHD was larger. Again, this is expected due to the different data coverage of the different techniques.

### 3.3 Influence of the inlet system

During the comparison experiments at PAL and CVO the travelling instrument switched occasionally from the WCC inlet to the station inlet. This was used to further investigate the influence of the inlet system. At PAL, no difference was observed between the station and the WCC-Empa inlet systems for $CO_2$. The vertical grey shaded areas in Fig. 6 denote the periods when the travelling instrument sampled air from the PAL inlet, whereas the dedicated WCC-Empa inlet was used the rest of the time. The difference between the two
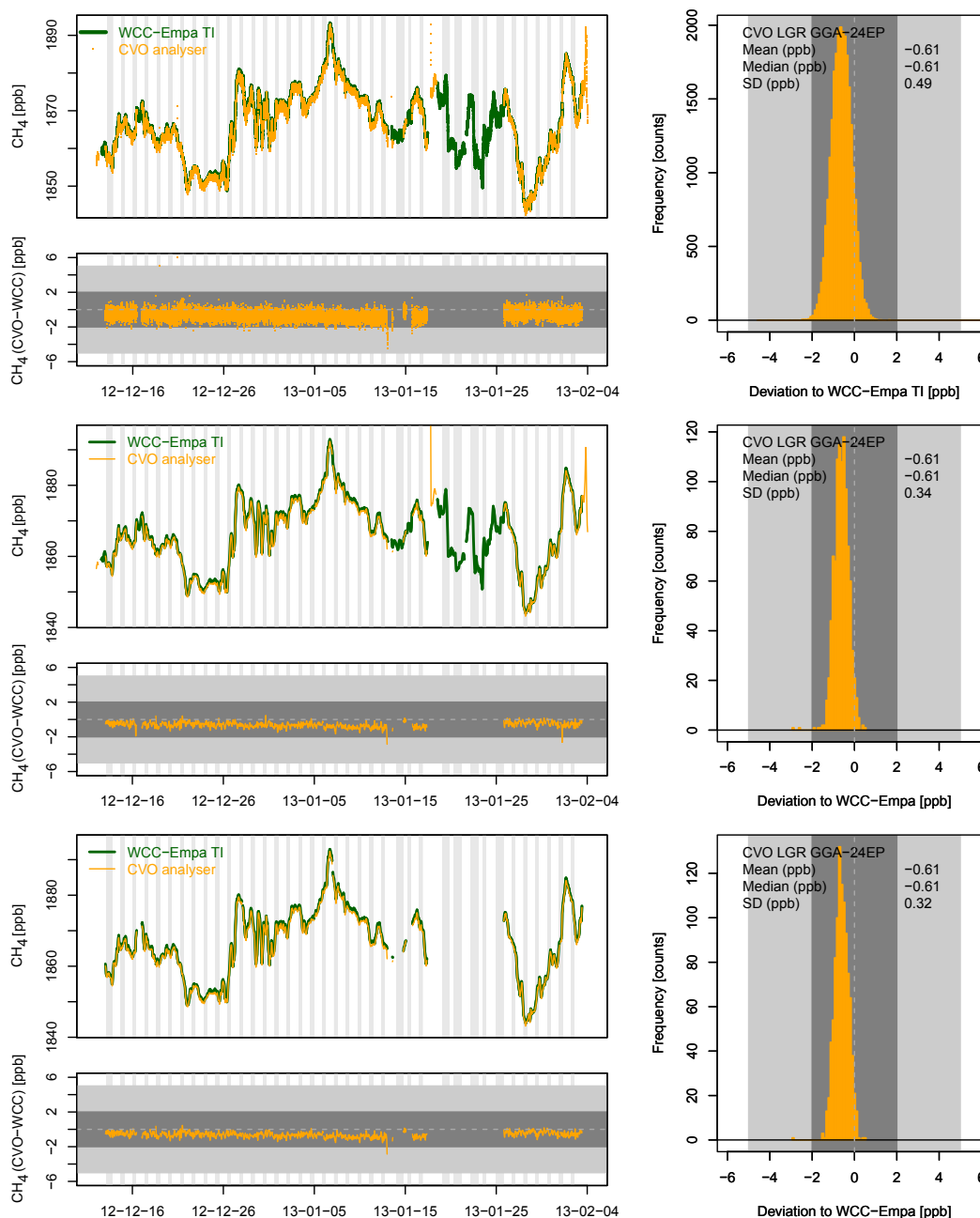
**Figure 11.** $CH_4$ comparison at CVO between the WCC-Empa travelling instrument and the CVO LGR GGA-24EP instruments. Left: $CH_4$ time series and $CH_4$ bias vs. time. Right: deviation histogram. Upper set: 1 min data; middle set: 1 h data, calculated from all available 1 min values; lower set: 1 h data, calculated from 1 min values with both CVO and WCC-Empa data coverage. The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals; vertical grey bars illustrate when different inlets were used (see text for details).

inlets was not significant ($0.01 \pm 0.06$ ppm), indicating that the PAL inlet system is fully appropriate. This was also the case for the CVO $CH_4$ measurements, where the mean bias of the CVO instrument was $-0.71 \pm 0.50$ ppb measured at the CVO inlet compared to $-0.59 \pm 0.49$ ppb at the WCC-Empa inlet. However, a small $CO_2$ bias was observed at CVO, with

a mean deviation of $0.05 \pm 0.06$ ppm measured at the CVO GHG inlet compared to $0.13 \pm 0.09$ ppm at the reactive gases inlet. This is not unexpected, since the reactive gases inlet is located 24 m below the CVO GHG inlet and thus may be more influenced by the local and regional $CO_2$ emissions.
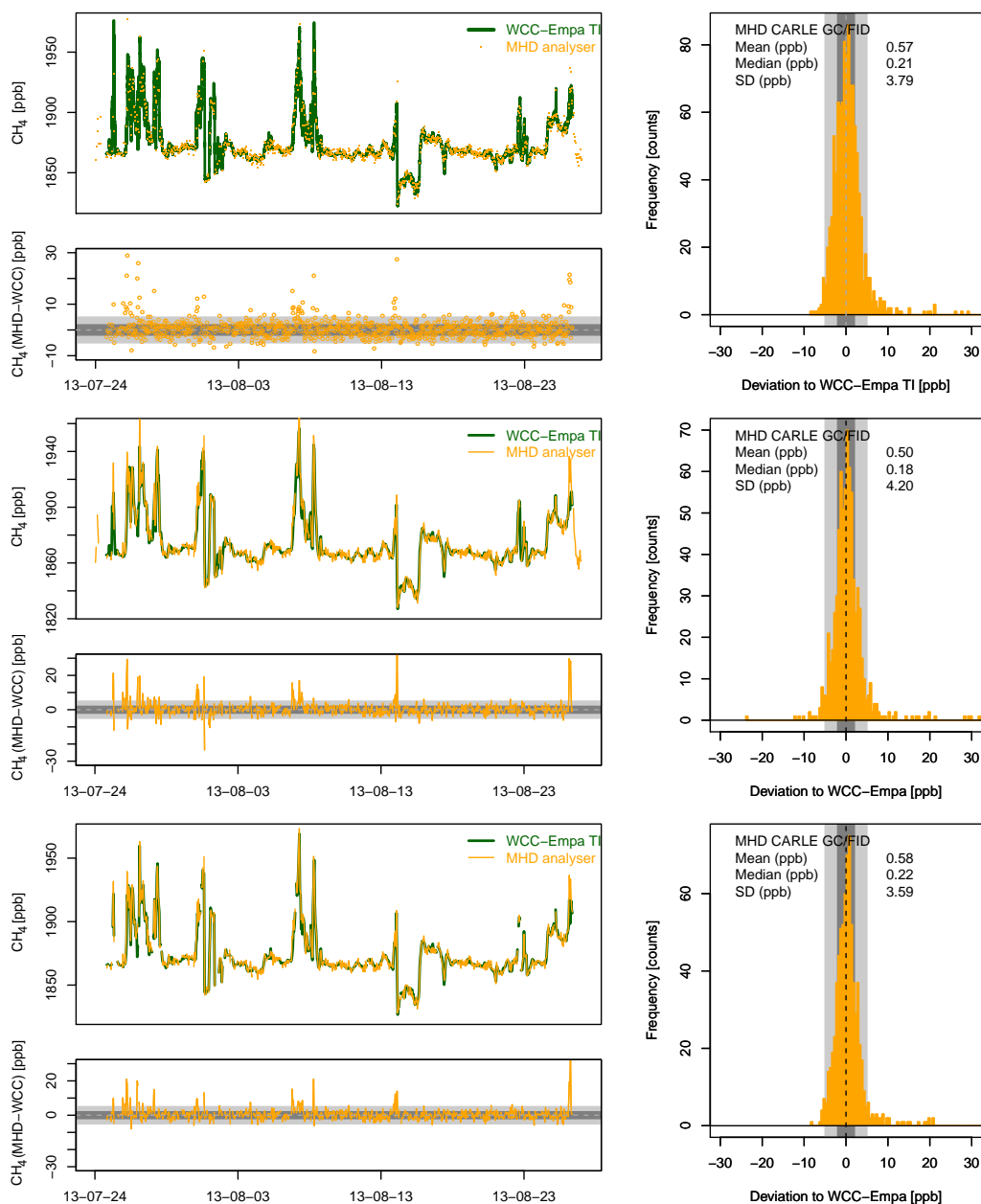
**Figure 12.** CH₄ comparison at MHD between the WCC-Empa travelling instrument and the MHD CARLE GC/FID instruments. Left: CH₄ time series and CH₄ bias vs time. Right: deviation histogram. Upper set: 1 min/single injection data; middle set: 1 h data, calculated from all available 1 min and single injection values; lower set: 1 h data, calculated from 1 min and single injection values with both MHD and WCC-Empa data coverage. The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals.

Again, this result indicates that the inlet system is fully appropriate at CVO as well.

In other cases (Zellweger et al., 2013), which are not shown here due to the scope of the present study, significant deviations between inlet systems can be observed. Those were mainly caused by leaks or inefficient drying procedures. The use of independent inlet systems during ambient air comparisons can clearly provide valuable information on the overall performance of the measurement setup, which cannot be obtained by the comparisons of standard gases alone. In the cases presented here, the inlet designs were appropriate and did not contribute to a potential bias.
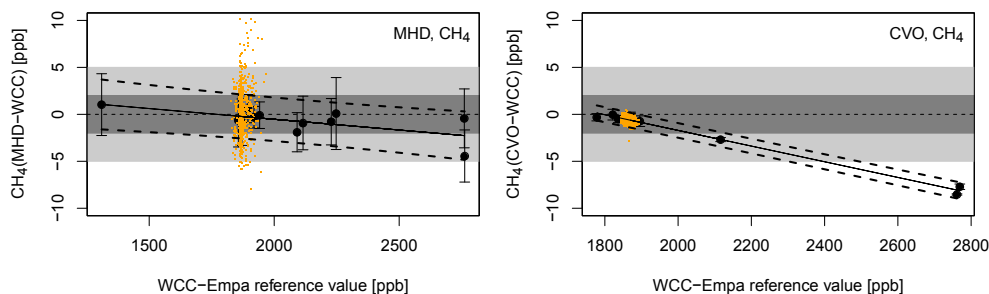
**Figure 13.** Results of the $CH_4$ travelling standard comparisons (performance audits) at MHD (left) and CVO (right). The grey areas correspond to the WMO/GAW compatibility and extended compatibility goals. The orange points correspond to the observed differences based on hourly data during the ambient air comparison.

## 3.4 Water vapour correction

Traditional measurement techniques for GHGs require drying of the sample gas to achieve the WMO/GAW compatibility goals. A recent study by Rella et al. (2013) shows that measurements of humid air can be made using CRDS instruments when an appropriate correction for water vapour interference is applied. They concluded that it is possible to make $CO_2$ and $CH_4$ measurements within the GAW compatibility goals (for the Northern Hemisphere) for water vapour levels up to at least 2 % by determining the water vapour correction function once on a per-instrument basis. During our study, all WCC-Empa measurements, except at CVO, were made without drying the air sample.

The water vapour correction functions were experimentally determined several times per instrument, which provides information about the stability and reproducibility of the correction function. The correction function is a second-order polynomial, as shown by following equations:

$$CO_2(dry) = CO_2(wet)/\left(1 + a \cdot H_2O + b \cdot H_2O^2\right), \qquad (1)$$

$$CH_4(dry) = CH_4(wet)/\left(1 + c \cdot H_2O + d \cdot H_2O^2\right). \qquad (2)$$

$CO_2$(wet), $CH_4$(wet), and $H_2O$ are the (humid) mole fractions in ppm ($CO_2$, $CH_4$) or % ($H_2O$) reported by the analyser.

Figure 14 shows the differences between the first and the subsequent measurements for the two instruments at a nominal mole fraction of 400 ppm ($CO_2$) and 1900 ppb ($CH_4$) with the WMO/GAW compatibility targets. The CRDS instruments used for this study have built-in water vapour correction functions and report also dry mole fractions. However, to achieve the best possible correction equation, it is advisable that instrument specific correction functions are determined for each instrument using the method described by Rella et al. (2013). The robustness of the correction can be further improved by pooling a number of correction functions obtained by several experiments or by selecting a correction function that is representative for a larger set of experiments.
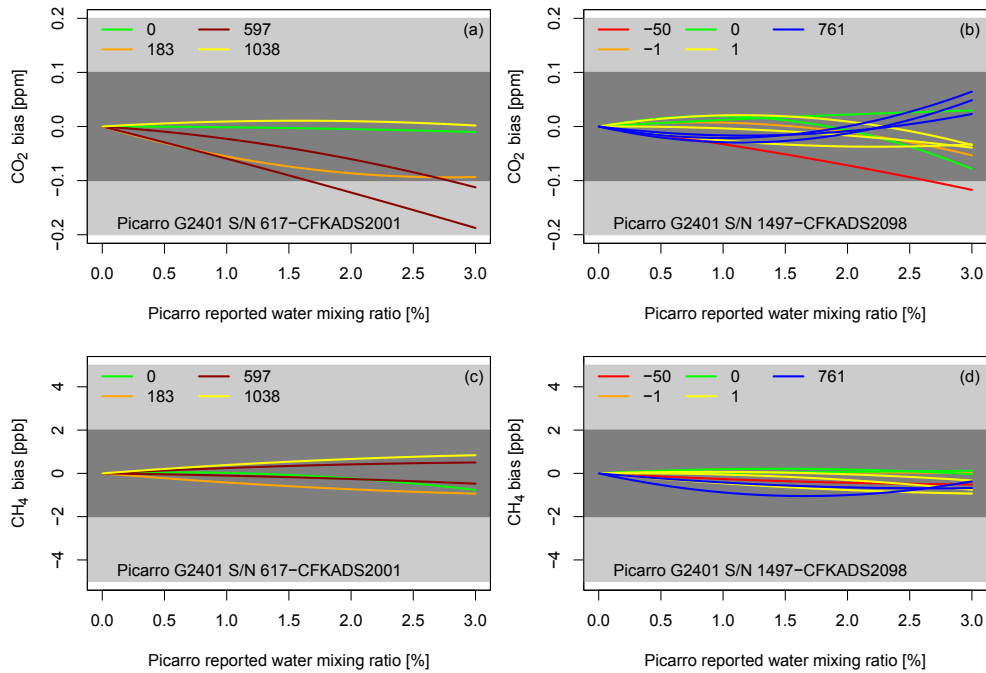
Figure 14 shows that the $CO_2$ correction functions result in dry mole fractions that are within the WMO/GAW compatibility goals for water vapour up to 2 % when functions obtained at the same day are pooled (CFKADS#2001) and up to >2.5 % for the newer analyser (CFKADS#2098), which is in line with the results published by Rella et al. (2013). The differences between determinations of the water correction function at same day and after longer time periods were similar, which indicates that the short-term and the long-term sensitivity changes are in the same order or that the repeatability of the droplet test using the Empa method (Rella et al., 2013; Zellweger et al., 2012) is the limiting factor. Also, it was confirmed that the $CH_4$ correction functions result in dry mole fractions that are within the WMO/GAW compatibility goals for water vapour up to 3 %. In our case, only one correction function that was initially retrieved was used to correct the data. These initial coefficients, as determined using the Empa method described by Rella et al. (2013), are summarised in Table 4.

The difference between dry station instrument measurements and humid travelling instrument measurements as a function of water vapour is illustrated in Fig. 15 for all comparisons. No dependency was found between the observed $CO_2$ bias and $H_2O$ measured by the travelling instrument at PAL. This result is consistent with a comparison using the data of the travelling instrument and another CRDS instrument with dry sample air that has been published by Rella et al. (2013). It is noteworthy that the same result was obtained at DMV despite the much higher $H_2O$ content of the ambient air. A small remaining dependency cannot be excluded based on the current study due to the high variability of the observed bias; however, the contribution to the overall uncertainty would be small compared to other sources of uncertainty. Since the water vapour interference is the limiting factor for the $CO_2$ measurements due to the uncertainty caused by the correction, it is important to note that the results of PAL and particularly DMV confirm the applicability of $CO_2$ correction functions.

**Table 4.** Water vapour correction coefficients for the WCC-Empa travelling CRDS instruments; coefficients $a$ and $b$ refer to the correction of $CO_2$, while $c$ and $d$ refer to $CH_4$ (see Eqs. 1 and 2).

| Analyser | Date | $a$ | $b$ | $c$ | $d$ |
|---|---|---|---|---|---|
| CFKADS2001 | 30 Dec 2012 | −0.015331 | 0.000062 | −0.012452 | 0.000065 |
| CFKADS2098 | 27 Jun 2013 | −0.015625 | 0.000095 | −0.013050 | 0.000175 |



**Figure 14. (a)** Bias at 400 ppm $CO_2$ compared to initial correction function of the Picarro G2401 S/N 617-CFKADS2001. The different lines correspond to the individual determinations of the correction functions. The number in the legend refers to the number of days from the determination of the reference correction function. **(b)** Same as **(a)**, for Picarro G2401 S/N 1497-CFKADS2098. **(c and d)** Same as **(a and b)**, for $CH_4$ at 1900 ppb. The grey areas correspond to the WMO/GAW compatibility (dark grey) and extended compatibility (light grey) goals.
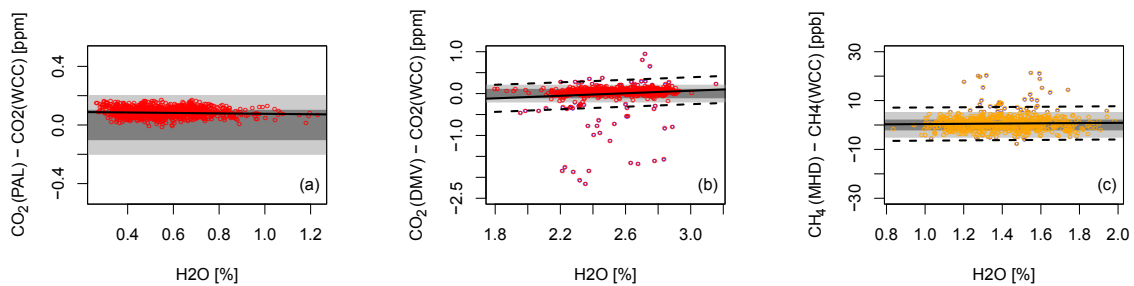


**Figure 15. (a)** $CO_2$ bias of PAL vs. $H_2O$ measured by the travelling instrument based on hourly matched data. The solid black lines denote the linear regressions, while the dashed lines are the 95 % confidence bands. **(b)** Same as **(a)**, for DMV. **(c)** Same as **(a)**, for $CH_4$ at MHD. The grey areas correspond to the WMO/GAW compatibility and extended compatibility goals.
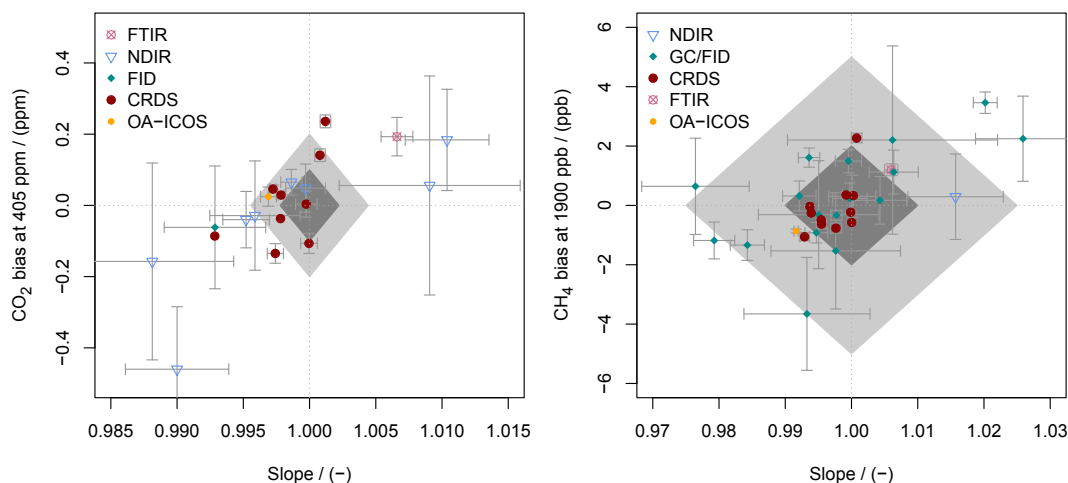
**Figure 16.** Left: $CO_2$ bias at 405 ppm vs. the slope of the performance audit for individual travelling standard comparisons involving different measurement techniques of the audited station analysers. Right: same for $CH_4$, with bias at 1900 ppb. The dark grey and light grey areas correspond to the WMO/GAW compatibility and extended compatibility goals for the range from 1700 to 2100 ppb $CH_4$ and from 360 to 450 ppm $CO_2$.

## 3.5 Evaluation of audit results

With our study, we aimed to evaluate the performance of modern spectroscopic analysers (CRDS, OA-ICOS, and FTIR) in comparison to the traditional techniques. WCC-Empa conducted 32 station audits with a travelling standard for $CH_4$ (2005–2014) and 20 for $CO_2$ (2010–2015). Details of the comparisons including instruments and analytical techniques are given in Table 1 for $CO_2$ and Table 2 for $CH_4$. Each comparison was evaluated by linear regression analysis as shown in Fig. 13. To judge whether the resulting slope/intercept combinations meet the WMO/GAW compatibility and extended compatibility goals, the bias in the centre of the mole fraction range (405 ppm for $CO_2$, 1900 ppb for $CH_4$) of the unpolluted troposphere (WMO, 2014) (360–450 ppm for $CO_2$, 1700–2100 ppb for $CH_4$) was plotted against the slope of the individual travelling standard comparisons. This is shown in Fig. 16 along with the allowed bias/slope combinations corresponding to the compatibility (dark grey area) and extended compatibility goals (light grey area) over the entire mole fraction range of the unpolluted troposphere. Only comparisons that were on the same calibration scale and without any known instrument malfunctions were considered. It can be clearly seen that large differences exist among the evaluated analytical techniques. Newer spectroscopic techniques such as CRDS and OA-ICOS show generally better performance with respect to accuracy and measurement uncertainty compared to NDIR ($CO_2$) and GC/FID ($CH_4$). Moreover, these techniques also provide better data coverage, which further reduces the uncertainty.

The results of the above analysis are further presented in Fig. 17, which summarises the percentage of comparisons
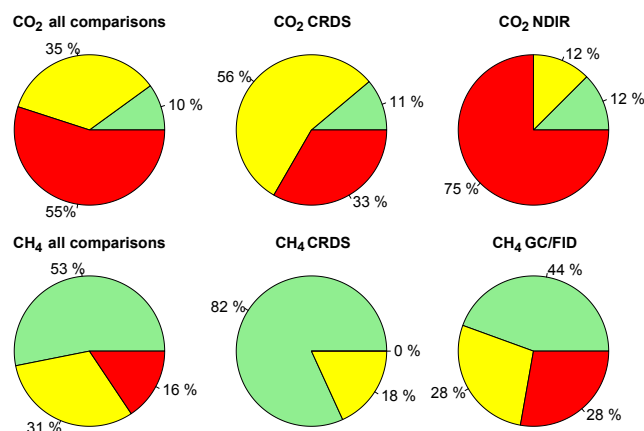


**Figure 17.** Upper panel: percentage of $CO_2$ performance audit results that were within the WMO/GAW compatibility goals in the range from 360 to 450 ppm (green), the extended compatibility goals (yellow), or outside the compatibility goals (red area). Results for all, but only CRDS and NDIR comparisons are shown. Lower panel: same for $CH_4$ in the range of 1700–2100 ppb; for all, but only CRDS and GC/FID comparisons.

that met the compatibility and extended compatibility goals. We show all comparisons and then separately only travelling standard–CRDS, travelling standard–NDIR ($CO_2$) and travelling standard–GC/FID ($CH_4$) comparisons.

It is obvious that reaching the compatibility goals for $CO_2$ remains a challenge; out of the 20 comparisons, only 2 (10 %) met the compatibility goal with 7 (35 %) meeting the extended goal. However, these results include the entire $CO_2$ mole fraction range relevant for the troposphere. Often, the calibration ranges at stations are intentionally limited to

the ambient mole fraction ranges typical for their location. Such ranges can be significantly smaller than those used for Figs. 16 and 17, e.g. at CVO. Therefore, slope/bias pairs that are outside the compatibility goals do not necessarily imply that the measurements at a station are biased. However, they provide useful information about the performance of the instrument as well as the calibration over the entire mole fraction range of the unpolluted troposphere.

For $CO_2$, the overall compatibility of the travelling standard–CRDS comparisons was significantly better compared to the travelling standard–NDIR comparisons. From the total nine travelling standard–CRDS comparisons, five (56 %) were within ±0.2 ppm and one within ±0.1 ppm (11 %). Out of the eight travelling standard–NDIR comparisons, one (12 %) reached the ±0.1 ppm and another one (12 %) the ±0.2 ppm limit. All other comparisons (75 %) were outside the compatibility goals. This is primarily due to the poorer repeatability and limitations of the NDIR technique in general (drift and nonlinearity of the detector) and, secondly, to the calibration linearity issues which become relevant when comparison results are expanded beyond the typical mole fraction range for a particular measurement site. This was also reflected by the error bars in Fig. 16, which in most cases were significantly larger for travelling standard–NDIR comparisons compared to travelling standard–CRDS $CO_2$ comparisons.

Meeting the WMO/GAW compatibility goals for $CH_4$ is significantly less challenging. Out of the 32 comparisons, 17 (53 %) were within the compatibility goals and 10 (31 %) were within the extended compatibility goals. A total number of five (16 %) comparisons did not meet the extended compatibility goals.

Figure 17 presents the results analysed as a function of the measurement techniques. All $CH_4$ travelling standard–CRDS comparisons were within the extended compatibility goals; nine (82 %) of the comparison met the ±2 ppb limit, whereas the ±5 ppb limit was met in the remaining two (18 %) cases. This is significantly more than for the GC/FID systems, of which only eight (44 %) met the ±2 ppb limit and five (28 %) the ±5 ppb limit. The remaining five (28 %) of the $CH_4$ travelling standard–CRDS comparisons were outside the compatibility goals. Focusing on the station relevant $CH_4$ mole fraction range should not be the reason of the worse result compared to CRDS, since the GC/FID method is known to be linear in the mole fraction range discussed here. More likely, worse instrument repeatability compared to CRDS plays an important role here, which is further illustrated by the significantly larger error bars of the bias/slope pairs in Fig. 16 for GC/FID systems.

Other techniques (OA-ICOS, FTIR) also indicate better repeatability compared to GC/FID ($CH_4$) and NDIR ($CO_2$); however, an insufficient number of comparison studies have been made to reliably show the superior performance of these techniques.

## 4  Conclusions

The results of the analysis of both travelling standards and side-by-side comparisons with a travelling instrument show that laser-based spectrometers such as the Picarro CRDS can be suitable for accurate $CO_2$ and $CH_4$ measurements. This is important with respect to the continuation of long-term time series of $CO_2$ and $CH_4$. Due to the higher temporal data coverage, repeatability and linearity, the accuracy of $CO_2$ and $CH_4$ measurements is expected to improve when traditional technologies such as NDIR analysers for $CO_2$ and GC/FID systems for $CH_4$ are replaced. Furthermore, CRDS-derived $CO_2$ and $CH_4$ measurements do not require the drying of the sampled air and water vapour corrections appear to be valid even under very humid conditions as encountered at tropical sites. The resulting remaining uncertainty is assumed to be small compared to other contributing factors to the overall uncertainty. In particular, incomplete data coverage unavoidable for quasi-continuous methods (e.g. GC measurements) or techniques requiring frequent calibrations (e.g. NDIR measurements) remains one of the significant contributing factors to the overall uncertainty. This becomes most important at locations with high temporal variability of the observed parameters, e.g. due to local and regional emission sources or atmosphere–biosphere exchange processes.

A thorough analysis of the $CO_2$ and $CH_4$ stability of CRDS instruments indicates that the optimal calibration frequency is approximately 30 h. This frequency is sufficient to compensate for the instrumental drift, and at the same time it allows to reduce the loss of ambient air measurements when calibrating an instrument. We believe that the modern measurement techniques such as CRDS will increase the number of GAW stations complying with the WMO/GAW compatibility goals for both $CO_2$ and $CH_4$. However, the fact remains that the compatibility goal of ±0.1 ppm for $CO_2$ can still be very challenging for many stations. Out of the measurement techniques employed in this study, the CRDS analyser has shown the best performance. Next to this specific type of laser spectrometer thoroughly assessed here, other techniques such as FTIR or OA-ICOS are becoming more widely available. To date, not enough comparison data are available from our audits to draw firm conclusions on the performance of these techniques, but initial results indicate that they also have the potential of being superior compared to traditional methods (NDIR, GC/FID). However, despite the advantages of the new techniques, care has to be taken with regard to calibration strategies, sample inlet setup and appropriate water vapour corrections.

Our analysis has demonstrated that, providing an adequate design of the measurement system, the performance assessment with either travelling standards or a travelling instrument leads to a similar conclusion. The two comparison methods supply complementary information: the approach which utilises travelling standards is better suited to characterise the performance of the instrument e.g. with regard

to linearity, whereas the side-by-side comparison has the advantage of incorporating the whole system including an inlet, a drying system and the instrument calibration over a longer period of time. Therefore, this two-pillar audit scheme with comparison of travelling standards and multi-week side-by-side comparisons of station instrumentation with travelling instruments proves to be a valid approach for data quality assessments at atmospheric measurement stations.

## 5   Data availability

Data from the performance audits at different GAW stations are available from the corresponding audit reports (http://www.empa.ch/web/s503/wcc-empa). Other data used in the paper is available upon request to the corresponding author as the raw data is not publicly archived.

## References

Brailsford, G. W., Stephens, B. B., Gomez, A. J., Riedel, K., Mikaloff Fletcher, S. E., Nichol, S. E., and Manning, M. R.: Long-term continuous atmospheric $CO_2$ measurements at Baring Head, New Zealand, Atmos. Meas. Tech., 5, 3109–3117, doi:10.5194/amt-5-3109-2012, 2012.

Buchmann, B., Klausen, J., and Zellweger, C.: Traceability of Long-Term Atmospheric Composition Observations across Global Monitoring Networks, Chimia, 63, 657–660, 2009.

Chen, H., Winderlich, J., Gerbig, C., Hoefer, A., Rella, C. W., Crosson, E. R., Van Pelt, A. D., Steinbach, J., Kolle, O., Beck, V., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Santoni, G. W., and Wofsy, S. C.: High-accuracy continuous airborne measurements of greenhouse gases ($CO_2$ and $CH_4$) using the cavity ring-down spectroscopy (CRDS) technique, Atmos. Meas. Tech., 3, 375–386, doi:10.5194/amt-3-375-2010, 2010.

Crosson, E. R.: A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor, Appl. Phys. B-Lasers O., 92, 403–408, 2008.

Dlugokencky, E. J., Steele, L. P., Lang, P. M., and Masarie, K. A.: Atmospheric methane at Mauna Loa and Barrow observatories: Presentation and analysis of in situ measurements, J. Geophys. Res., 100, 23103–23113, 1995.

Dlugokencky, E. J., Myers, R. C., Lang, P. M., Masarie, K. A., Crotwell, A. M., Thoning, K. W., Hall, B. D., Elkins, J. W., and Steele, L. P.: Conversion of NOAA atmospheric dry air $CH_4$ mole fractions to a gravimetrically prepared standard scale, J. Geophys. Res., 110, D18306, doi:10.1029/2005JD006035, 2005.

Flores, E., Rhoderick, G. C., Viallon, J., Moussay, P., Choteau, T., Gameson, L., Guenther, F. R., and Wielgosz, R. I.: Methane Standards Made in Whole and Synthetic Air Compared by Cavity Ring Down Spectroscopy and Gas Chromatography with Flame Ionization Detection for Atmospheric Monitoring Applications, Anal. Chem., 87, 3272–3279, 2015.

Flowers, B. A., Powers, H. H., Dubey, M. K., and McDowell, N. G.: Inter-comparison of two high-accuracy fast-response spectroscopic sensors of carbon dioxide: a case study, Atmos. Meas. Tech., 5, 991-997, 2012.

GAWSIS: available at: https://gawsis.meteoswiss.ch, last access: 20 June 2016.

Griffith, D. W. T., Deutscher, N. M., Caldow, C., Kettlewell, G., Riggenbach, M., and Hammer, S.: A Fourier transform infrared trace gas and isotope analyser for atmospheric applications, Atmos. Meas. Tech., 5, 2481–2498, doi:10.5194/amt-5-2481-2012, 2012.

Hammer, S., Konrad, G., Vermeulen, A. T., Laurent, O., Delmotte, M., Jordan, A., Hazan, L., Conil, S., and Levin, I.: Feasibility study of using a "travelling" $CO_2$ and $CH_4$ instrument to validate continuous in situ measurement stations, Atmos. Meas. Tech., 6, 1201–1216, doi:10.5194/amt-6-1201-2013, 2013.

Harris, D. C.: Charles David Keeling and the Story of Atmospheric CO2 Measurements, Anal. Chem., 82, 7865–7870, 2010.

Keeling, C. D.: The Concentration and Isotopic Abundances of Carbon Dioxide in the Atmosphere, Tellus, 12, 200-203, 1960.

Khalil, M. A. K. and Rasmussen, R. A.: Sources, Sinks, and Seasonal Cycles of Atmospheric Methane, J. Geophys. Res., 88, 5131–5144, 1983.

Kirschke, S., Bousquet, P., Ciais, P., et al.: Three decades of global methane sources and sinks, Nat. Geosci., 6, 813–823, 2013.

Klausen, J., Zellweger, C., Buchmann, B., and Hofer, P.: Uncertainty and bias of surface ozone measurements at selected Global Atmosphere Watch sites, J. Geophys. Res., 108, 4622, doi:10.1029/2003JD003710, 2003.

Komhyr, W. D., Harris, T. B., Waterman, L. S., Chin, J. F. S., and Thoning, K. W.: Atmospheric carbon dioxide at Mauna Loa Observatory: 1. NOAA global monitoring for climatic change measurements with a nondispersive infrared analyzer, 1974–1985, J. Geophys. Res., 94, 8533–8547, 1989.

Leuenberger, M. C., Schibig, M. F., and Nyfeler, P.: Gas adsorption and desorption effects on cylinders and their importance for long-term gas records, Atmos. Meas. Tech., 8, 5289–5299, doi:10.5194/amt-8-5289-2015, 2015.

McManus, J. B., Zahniser, M. S., Nelson, D. D., Shorter, J. H., Herndon, S. C., Jervis, D., Agnese, M., McGovern, R., Yacovitch, T. I., and Roscioli, J. R.: Recent progress in laser-based trace gas instruments: performance and noise analysis, Appl. Phys. B-Lasers O., 119, 203–218, 2015.

O'Shea, S. J., Bauguitte, S. J.-B., Gallagher, M. W., Lowry, D., and Percival, C. J.: Development of a cavity-enhanced absorption spectrometer for airborne measurements of $CH_4$ and $CO_2$, At-

mos. Meas. Tech., 6, 1095–1109, doi:10.5194/amt-6-1095-2013, 2013.

Peltola, O., Hensen, A., Helfter, C., Belelli Marchesini, L., Bosveld, F. C., van den Bulk, W. C. M., Elbers, J. A., Haapanala, S., Holst, J., Laurila, T., Lindroth, A., Nemitz, E., Röckmann, T., Vermeulen, A. T., and Mammarella, I.: Evaluating the performance of commonly used gas analysers for methane eddy co-variance flux measurements: the InGOS inter-comparison field experiment, Biogeosciences, 11, 3163–3186, doi:10.5194/bg-11-3163-2014, 2014.

Rella, C. W., Chen, H., Andrews, A. E., Filges, A., Gerbig, C., Hatakka, J., Karion, A., Miles, N. L., Richardson, S. J., Steinbacher, M., Sweeney, C., Wastine, B., and Zellweger, C.: High accuracy measurements of dry mole fractions of carbon dioxide and methane in humid air, Atmos. Meas. Tech., 6, 837–860, doi:10.5194/amt-6-837-2013, 2013.

Schibig, M. F., Steinbacher, M., Buchmann, B., van der Laan-Luijkx, I. T., van der Laan, S., Ranjan, S., and Leuenberger, M. C.: Comparison of continuous in situ $CO_2$ observations at Jungfraujoch using two different measurement techniques, Atmos. Meas. Tech., 8, 57–68, doi:10.5194/amt-8-57-2015, 2015.

Vardag, S. N., Hammer, S., O'Doherty, S., Spain, T. G., Wastine, B., Jordan, A., and Levin, I.: Comparisons of continuous atmospheric $CH_4$, $CO_2$ and $N_2O$ measurements – results from a travelling instrument campaign at Mace Head, Atmos. Chem. Phys., 14, 8403–8418, doi:10.5194/acp-14-8403-2014, 2014.

Werle, P., Mücke, R., and Slemr, F.: The Limits of Signal Averaging in Atmospheric Trace-Gas Monitoring by Tunable Diode-Laser Absorption Spectroscopy (TDLAS), Appl. Phys. B-Lasers O., 57, 131–139, 1993.

WMO: 15th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurements Techniques, Jena, Germany, 7–10 September 2009 (WMO TD No. 1553), World Meteorological Organization, Geneva, Switzerland, GAW Report No. 194, 318 pp., 2011a.

WMO: Addendum for the Period 2012–2015 to the WMO Global Atmosphere Watch (GAW) Strategic Plan 2008–2015, World Meteorological Organization, Geneva, Switzerland, GAW Report No. 197, 57 pp., 2011b.

WMO: 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT-2011), Wellington, New Zealand, 25–28 October 2011, World Meteorological Organization, Geneva, Switzerland, GAW Report No. 206, 67 pp., 2012.

WMO: 17th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2013), Beijing, China, 10–13 June 2013, World Meteorological Organization, Geneva, Switzerland, GAW Report No. 213, 168 pp., 2014.

World Calibration Centre (WCC-Empa): Data from the performance audits at different GAW stations, available at: http://www.empa.ch/web/s503/wcc-empa, last access: 23 September 2016.

Yver Kwok, C., Laurent, O., Guemri, A., Philippon, C., Wastine, B., Rella, C. W., Vuillemin, C., Truong, F., Delmotte, M., Kazan, V., Darding, M., Lebègue, B., Kaiser, C., Xueref-Rémy, I., and Ramonet, M.: Comprehensive laboratory and field testing of cavity ring-down spectroscopy analyzers measuring $H_2O$, $CO_2$, $CH_4$ and CO, Atmos. Meas. Tech., 8, 3867–3892, doi:10.5194/amt-8-3867-2015, 2015.

Zellweger, C., Steinbacher, M., and Buchmann, B.: Evaluation of new laser spectrometer techniques for in-situ carbon monoxide measurements, Atmos. Meas. Tech., 5, 2555–2567, doi:10.5194/amt-5-2555-2012, 2012.

Zellweger, C., Steinbacher, M., Emmenegger, L., and Buchmann, B.: Added value of side-by-side comparisons with a travelling analyzer during WCC-Empa audits, in: Proceedings of the 17th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2013), Beijing, China, 10–13 June 2013.

Zhao, C. L. and Tans, P. P.: Estimating uncertainty of the WMO mole fraction scale for carbon dioxide in air, J. Geophys. Res., 111, D08S09, doi:10.1029/2005JD006003, 2006.