



Comprehensive laboratory and field testing of cavity ring-down spectroscopy analyzers measuring H₂O, CO₂, CH₄ and CO

C. Yver Kwok¹, O. Laurent¹, A. Guemri¹, C. Philippon¹, B. Wastine^{1,a}, C. W. Rella², C. Vuillemin¹, F. Truong¹, M. Delmotte¹, V. Kazan¹, M. Darding¹, B. Lebègue¹, C. Kaiser^{1,b}, I. Xueref-Rémy¹, and M. Ramonet¹

¹Laboratoire des Sciences du Climat et l'Environnement (LSCE/IPSL), CNRS-CEA-UVSQ, Centre d'Etudes Orme des Merisiers, Gif-sur-Yvette, France

²Picarro Inc., 3105 Patrick Henry Drive, Santa Clara, CA, USA

^anow at: Air Normand, 3 Place de la Pomme d'Or, 76000 Rouen, France

^bnow at: Laboratoire National de métrologie et d'essais, (LNE), 1 rue Gaston Boissier, 75724 Paris CEDEX 15, France

Correspondence to: C. Yver Kwok (camille.yver@lsce.ipsl.fr)

Received: 12 March 2015 – Published in Atmos. Meas. Tech. Discuss.: 27 April 2015

Revised: 11 August 2015 – Accepted: 31 August 2015 – Published: 23 September 2015

Abstract. To develop an accurate measurement network of greenhouse gases, instruments in the field need to be stable and precise and thus require infrequent calibrations and a low consumption of consumables. For about 10 years, cavity ring-down spectroscopy (CRDS) analyzers have been available that meet these stringent requirements for precision and stability. Here, we present the results of tests of CRDS instruments in the laboratory (47 instruments) and in the field (15 instruments). The precision and stability of the measurements are studied. We demonstrate that, thanks to rigorous testing, newer models generally perform better than older models, especially in terms of reproducibility between instruments. In the field, we see the importance of individual diagnostics during the installation phase, and we show the value of calibration and target gases that assess the quality of the data. Finally, we formulate recommendations for use of these analyzers in the field.

1 Introduction

The Integrated Carbon Observation System (ICOS) is a European research infrastructure project that is currently reaching its operational phase after a 5-year preparatory phase (<http://www.icos-infrastructure.eu/>). Its goal is to provide high-quality observations needed to understand the long-term trend and spatial distribution of greenhouse gas emis-

sions. For this purpose, ICOS is setting up monitoring networks of greenhouse gases over Europe in the atmosphere, ecosystems and at the surface of the oceans. In addition to the networks, central facilities have been designed during the preparatory phase to coordinate and standardize field operations and measurement protocols. Different methods will be developed to infer continental and regional carbon budgets from these measurements. The atmospheric approach using inverse modeling (Carouge et al., 2010a, b; Bousquet et al., 2011, 2013) relies on our ability to characterize very precisely the regional gradients of greenhouse gases over Europe. The efficient mixing of air masses in the troposphere has the advantage of integrating the signals from highly variable surface sources and sinks, but it may also have the inconvenient effect of smoothing very quickly the regional gradients. Previous studies have shown that the monthly mean gradients over Europe are generally lower than 10 ppm for CO₂ (Ramonet et al., 2009; Xueref-Remy et al., 2011a, b), and 100 ppb for CH₄ (from ICOS data). Moreover, the interannual variability from CO₂ is even lower, just a few parts per million (Ramonet et al., 2009). This typical greenhouse gas variability over Europe provides the most important data quality constraint of the ICOS-like observations. However, other parameters such as robustness, stability, and low-maintenance requirements also influence the choice of instruments to be used in field stations like the sites in the ICOS infrastructure. Gas chromatography (GC) systems (Yver et al., 2009; Schmidt et al., 2014) or monitors based on

nondispersive infrared (NDIR) sensors (Xueref-Remy et al., 2011a; Andrews et al., 2014; Schmidt et al., 2014) have been widely used over the last decades for high-precision measurements of CO₂ as well as CH₄ and N₂O for GC. Both technologies require hourly to daily calibrations to produce high-precision measurements and for GC a relatively high level of expertise to produce high-precision measurements. This usually leads to high maintenance and difficult installation of these instruments in remote locations. For about 10 years, cavity ring-down spectroscopy (CRDS) analyzers have been developed and commercialized by a few companies (Crosson, 2008; Peltola et al., 2014). This new generation of sensors for greenhouse gases can be easily deployed in the field and requires lower maintenance and consumables compared to GC and NDIR technologies.

In the framework of ICOS, the ICOS Atmospheric Thematic Centre (ATC) metrology laboratory (MLab hereafter) based at the Laboratoire des Sciences du Climat et de l'Environnement (LSCE) is responsible for testing every instrument within the ICOS atmospheric network. Here, we present the results of tests for CRDS instruments that measure CO₂, CH₄ and CO (Models ESP1000, G1301, G1302, G2301, G2302, G2401, Picarro, Inc., Santa Clara, CA, USA). We show the results for 47 instruments tested in the laboratory and from 15 field instruments measuring on sites instrumented for greenhouse gas for at least a year. These instruments have been dispatched to stations in various countries and environments with several stations in France within the SNO-RAMCES (National Observation Service – ICOS France <https://icos-atc.lsce.ipsl.fr/?q=stations>), stations in harsh environments as part of the CarboAfrica project (http://www.carboafrika.net/index_en.asp) or the ICOS-INWIRE project (<http://www.icos-inwire.lsce.ipsl.fr/>) in Côte d'Ivoire, French Guyana, Bolivia, and other ICOS stations in Europe. Other types of analyzers using different technologies or measuring different components such as N₂O or carbon isotopes have also been tested by the MLab but are not discussed in this study.

In the first section, we describe the analysis technique and the different models of instruments. Then, the protocols and the metrics used to assess the performance of the instruments are defined. In the last section, results for each species (CO₂, CH₄ and CO) are presented and discussed. Finally, in the conclusion, we sum up recommendations for use of these instruments in the field.

2 Instruments

All the results presented in this study come from tests performed at the manufacturer, at the MLab and in the field on CRDS analyzers manufactured by the company Picarro, Inc. between 2008 and 2014. They cover five different instrument models and up to four species (CO₂, CH₄, CO, H₂O).

The CRDS technique can be described as follows. A laser source is used to excite a measurement cell, which consists of a low-loss optical resonant cavity composed of at least two concave high-reflectivity mirrors. As the injected laser light propagates back and forth between the mirrors, a portion of the light is retransmitted through the mirror after each pass. A photosensitive detector located behind one of the mirrors monitors the time decay of the laser light. The decay (or ring-down) time depends on the cavity loss but also on the presence of any absorber species inside the cavity. Thus, higher concentrations of the target analyte molecule in the cavity correspond to shorter ring-down times. This technique and the details for each model are detailed in Crosson (2008), Chen et al. (2010, 2013), and Rella et al. (2013). For the instruments considered here, near-infrared telecom lasers are used. It is important to note that although the chosen CO₂, CH₄ and H₂O absorption lines are fairly well separated from other absorption features, the absorption line for CO lies between CO₂ and H₂O lines and has a relatively low strength compared to the others, as can be seen in Fig. 1. It is then challenging to measure CO at a high precision.

The three models, named ESP1000, G1301 and G2301 measure CO₂, CH₄ and H₂O using two lasers (one for CO₂, one for CH₄ and H₂O). The G1302 and G2302 instruments measure CO₂, CO and H₂O and the G2401 instruments measure all four species (using one more laser for CO). It has to be noted that three of the G1302 tested at the MLab were the first instruments of this model and two of them were tested only before they were upgraded for H₂O–CO₂–CO cross-talks (CKADS04 and CKADS07). The actual performances of these instruments should be better than has been demonstrated here. The instruments considered in this analysis are listed in Table 1 with their serial number, model, time of purchase, database identifier when existing, field sites if used here, measured species (except H₂O) as well as the main results of different repeatability tests.

3 Protocols and metrics used in the study

The metrics defined in the study follow the International Vocabulary of Metrology guidelines (VIM, <http://www.bipm.org/en/publications/guides/vim.html>) and the Global Atmosphere Watch guidelines (GAW, http://gaw.empa.ch/glossary/glossary.html#section_2). They are usually calculated under repeatability conditions of measurements where all conditions stay identical over a short period of time. The continuous measurement repeatability commonly called precision is a repeatability measure applied to continuous measurements. The long-term repeatability which was commonly called reproducibility in the atmospheric community is a repeatability measure over an extended period of time (Andrews et al., 2014; Schmidt et al., 2014). The hereafter called short-term repeatability is what is defined as the repeatability in the VIM and GAW glossary. In the present

Table 1. The 47 analyzers considered in this study. Their serial number, model, ICOS number when attributed and period of purchase are indicated in the first columns. If field results are presented here, the site as well as the number of months the instrument was used in the span of this study (+ indicates that the instrument is still running) and the percentage of valid (V) and invalid (I) data are detailed. For the MLab site, these are the reference instruments used for comparison. Comparable results (continuous measurement repeatability (CMR), short-term repeatability (STR) and long-term repeatability (LTR)) between factory, MLab and field are shown for all species.

Serial no.	Model	ICOS ID	Purchased date	Site	On the site for (months)	Species	Picarro CMR	ATC 5 min/1 min	Field	Picarro STR	ATC	Field LTR	
CFADS14	ESP1000	38	Feb 2008	BIS	37, V: 61 % I: 39 %	CO ₂ CH ₄	0.083 0.54	0.050 0.34	0.038 0.27	0.022 0.09	0.01 0.09	0.018 0.19	
CFADS44	ESP1000	25	Feb 2009			CO ₂ CH ₄	0.043 0.37	0.051 0.37		0.046 0.38		0.03 0.33	
CFADS02	ESP1000	23	Feb 2008	LTO	61, V: 72 % I: 28 %	CO ₂ CH ₄	0.095 0.40	0.042 0.39	0.104 0.52	0.022 0.20	0.029 0.14	0.066 0.23	
CFADS03	ESP1000	77	Feb 2008	MLab		CO ₂ CH ₄	0.098 0.40	0.049 0.37		0.022 0.07	0.01 0.06	0.016 0.16	
CFADS45	G1301	24	Feb 2009	BIS	11, V: 52 % I: 47 %	CO ₂ CH ₄	0.035 0.24	0.043 0.33	0.038 0.31	0.017 0.12	0.05 0.08	0.016 0.10	
CFADS46	G1301	41	Feb 2009	MHD	54+, V: 96 % I: 4 %	CO ₂ CH ₄	0.052 0.68		0.051 0.56	0.051 0.15		0.033 0.32	
CFADS74	G1301	91	Dec 2009	OPE	35+ V: 82 % I: 18 %	CO ₂ CH ₄	0.04 0.22	0.059 0.24	0.051 0.25	0.054 0.40	0.01 0.06	0.02 0.08	0.029 0.15
CFADS95	G1301	76	Apr 2010	PUY	38+, V: 77 % I: 23 %	CO ₂ CH ₄	0.05 0.30	0.051 0.25	0.041 0.26	0.036	0.035 0.2	0.018 0.12	
CFADS87	G1301-m	26	Jan 2009			CO ₂ CH ₄	0.056 0.4	0.059 0.46		0.024 0.15	0.01 0.07	0.01 0.07	
CFADS2006	G2301		Jun 2012			CO ₂ CH ₄		0.027 0.41			0.01 0.06	0.03 0.36	
CFADS2122	G2301	54	Aug 2010	MHD	39+, V: 76 % I: 24 %	CO ₂ CH ₄	0.067 0.58	0.066 0.59	0.035 0.31	0.021 0.24	0.036 0.22	0.051 0.29	
CFADS2130	G2301	102	Nov 2010	PUJ	35+, V: 92 % I: 8 %	CO ₂ CH ₄	0.027 0.24		0.020 0.24	0.096 0.61		0.018 0.21	
CFADS2141	G2301	78	Nov 2010			CO ₂ CH ₄		0.031 0.33			0.02 0.22		
CFADS2142	G2301	75	Nov 2010	OPE	31, V: 84 % I: 16 %	CO ₂ CH ₄	0.019 0.18	0.022 0.26	0.021 0.22	0.025 0.12	0.01 0.21	0.024 0.20	
CFADS2164	G2301	93	Mar 2011	IVI	34+, V: 84 % I: 16 %	CO ₂ CH ₄	0.016 0.20	0.029 0.36	0.018 0.18	0.016 0.17	0.02 0.29	0.016 0.19	
CFADS2207	G2301	106	Sep 2011			CO ₂ CH ₄	0.036 0.30	0.026 0.30		0.024 0.32	0.02 0.23		
CFADS2213	G2301	111	Sep 2011	AMS	30+, V: 96 % I: 4 %	CO ₂ CH ₄	0.014 0.20	0.016 0.24	0.012 0.17	0.041 0.19	0.02 0.25	0.013 0.19	
CFADS2214	G2301	107	Sep 2011			CO ₂ CH ₄	0.024 0.27	0.047 0.38		0.021 0.20	0.03 0.13		
CFADS2304	G2301		Sep 2012			CO ₂ CH ₄	0.043 0.24	0.045 0.29		0.028 0.16	0.012 0.15	0.024 0.21	

Table 1. Continued.

Serial no.	Model	ICOS ID	Purchased date	Site	On the site for (months)	Species	Picarro	ATC	Field	Picarro	ATC	ATC	Field
							CMR	5 min/1 min		STR		LTR	
CFADS2314	G2301	172	Apr 2013			CO ₂	0.014	0.023		0.033	0.034	0.026	
						CH ₄	0.20	0.23		0.21	0.29	0.18	
CFADS2341	G2301		Jul 2013			CO ₂	0.017	0.019		0.03	0.008	0.012	
						CH ₄	0.20	0.38		0.30	0.10	0.20	
CFADS2343	G2301		Apr 2013			CO ₂	0.019	0.022		0.05	0.008	0.013	
						CH ₄	0.20	0.26		0.30	0.13	0.2	
CFADS2346	G2301		Nov 2013			CO ₂	0.018	0.021		0.027	0.006	0.015	
CKADS04	G1302	34	Jan 2010			CH ₄	0.20	0.30		0.20	0.10	0.21	
						CO ₂		0.052					
CKADS05	G1302	36	Jan 2010			CO		34.8					
						CO ₂	0.043	0.035		0.031		0.029	
CKADS07	G1302	35	Jan 2010			CO	9.6	7.5		2.6		4.2	
						CO ₂	0.035	0.057		0.012		0.014	
CKADS2032	G2302		Oct 2011			CO	47.9	8.1		7.4		1	
						CO ₂	0.031	0.033		0.031	0.01	0.02	
CFKADS2022	G2401	108	Oct 2011	TRN	13+, V: 76 % I: 4 %	CO	18.4	17.0		4.2	1.7	1	
						CO ₂	0.028	0.030	0.027	0.019		0.02	0.020
CFKADS2037	G2401	119	Nov 2011			CH ₄	0.24	0.29	0.20	0.23		0.22	0.22
						CO	5.5	5.6	6.8	2.4		1.7	2.2
CFKADS2041	G2401		Dec 2011			CO ₂	0.027	0.024		0.018		0.02	
						CH ₄	0.30	0.28		0.24		0.34	
CFKADS2071	G2401	148	Sep 2012	BIS	9+, V: 97 % I: 3 %	CO	5.7	5.3		0.8		0.4	
						CO ₂	0.070	0.034		0.063		0.02	
CFKADS2072	G2401	187	Sep 2012	OPE	5+, V: 42 % I: 58 %	CH ₄	0.60	0.32		0.30		0.17	
						CO	10.2	8.0		2.4		1.2	
CFKADS2077	G2401	148	Sep 2012	BIS	9+, V: 97 % I: 3 %	CO ₂	0.021	0.022	0.018	0.027	0.01	0.02	0.015
						CH ₄	0.20	0.23	0.17	0.3	0.15	0.34	0.26
CFKADS2072	G2401	187	Sep 2012	OPE	5+, V: 42 % I: 58 %	CO	5.4	5.5	5.3	1.0	0.4	2.2	1.9
						CO ₂	0.023	0.026	0.022	0.027	0.01	0.02	0.022
CFKADS2072	G2401	187	Sep 2012	OPE	5+, V: 42 % I: 58 %	CH ₄	0.20	0.28	0.19	0.20	0.15	0.22	0.19
						CO	6.5	5.7	5.5	1.2	1	1.7	3.4
CFKADS2073	G2401	150	Sep 2012			CO ₂	0.018	0.021		0.031	0	0.03	
						CH ₄	0.21	0.24		0.20	0.07	0.41	
CFKADS2073	G2401	150	Sep 2012			CO	4.9	5.1		0.1	0.5	0.6	
						CO ₂	0.023	0.028		0.037	0.01	0.01	
CFKADS2084	G2401	222	Jan 2013			CH ₄	0.20	0.31		0.40	0.11	0.23	
						CO	6.9	6.7		0.6	0.9	1.6	
CFKADS2095	G2401	236	Aug 2013			CO ₂	0.026	0.029		0.032	0.012	0.014	
						CH ₄	0.20	0.40		0.40	0.09	0.23	
CFKADS2095	G2401	236	Aug 2013			CO	8.8	9.0		23.6	0.6	1.0	
						CO ₂	0.021	0.023		0.031	0.012	0.016	
CFKADS2105	G2401		Jun 2013			CH ₄	0.20	0.30		0.30	0.20	0.24	
						CO	5.1	5.1		0.7	0.4	0.3	
CFKADS2111	G2401		Sep 2013			CO ₂	0.022	0.025		0.044	0.008	0.012	
						CH ₄	0.20	0.26		0.30	0.15	0.14	
CFKADS2111	G2401		Sep 2013			CO	7.0	7.5		1	0.5	0.7	
						CO ₂	0.022	0.025		0.033	0.009	0.016	
CFKADS2113	G2401		Sep 2013			CH ₄	0.30	0.27		0.40	0.16	0.23	
						CO	4.3	4.3		0.6	0.4	0.4	

Table 1. Continued.

Serial no.	Model	ICOS ID	Purchased date	Site	On the site for (months)	Species	Picarro	ATC	Field	Picarro	ATC	ATC	Field
							CMR	5 min/1 min		STR		LTR	
CFKADS2114	G2401		Sep 2013			CO ₂	0.018	0.025		0.025	0.008	0.012	
						CH ₄	0.20	0.30		0.2	0.15	0.18	
						CO	4.5	4.3		0.5	0.3	0.4	
CFKADS2120	G2401		Dec 2013			CO ₂	0.022	0.025		0.020	0.009	0.013	
						CH ₄	0.20	0.33		0.4	0.16	0.22	
						CO	5.8	42.3		0.9	0.4	0.5	
CFKADS2124	G2401	192	Dec 2013	LTO	3+, V: 99 % I: 1 %	CO ₂	0.02	0.022	0.018	0.033	0.007	0.01	0.019
						CH ₄	0.20	0.27	0.16	0.4	0.13	0.16	0.21
						CO	4.1	4.2	4.3	0.6	0.3	0.5	1.2
CFKADS2125	G2401		Nov 2013			CO ₂	0.021	0.028		0.024	0.008	0.01	
						CH ₄	0.20	0.34		0.40	0.14	0.18	
						CO	5.5	5.4		0.9	0.4	0.5	
CFKADS2127	G2401	221	Nov 2013	MLab		CO ₂	0.021	0.025		0.030	0.009	0.01	
						CH ₄	0.20	0.26		0.40	0.15	0.17	
						CO	0.2	4.8		0.9	0.2	0.3	
CFKADS2128	G2401	224	Nov 2013			CO ₂	0.025	0.026		0.022	0.008	0.01	
						CH ₄	0.20	0.29		0.2	0.13	0.23	
						CO	5.9	6.0		0.9	0.5	0.9	
CFKADS2129	G2401	223	Nov 2013			CO ₂	0.028	0.030		0.44	0.007	0.01	
						CH ₄	0.20	0.27		0.2	0.13	0.15	
						CO	7.4	6.5		1.4	0.5	0.6	
CFKADS2130	G2401		Nov 2013			CO ₂		0.022			0.01	0.013	
						CH ₄		0.25			0.19	0.16	
						CO		5.9			0.4	0.3	
CFKADS2131	G2401		Nov 2013			CO ₂	0.022	0.029		0.067	0.012	0.016	
						CH ₄	0.30	0.43		0.30	0.20	0.28	
						CO	6.5	6.5		1.1	0.5	0.8	
CFKBDS2001	G2401-m					CO ₂		0.025				0.01	
						CH ₄		0.25				0.17	
						CO		10.8				1.2	
CFKBDS2132	G2401-m		Jan 2014			CO ₂	0.025	0.029		0.063	0.006	0.015	
						CH ₄	0.20	0.31		0.30	0.11	0.25	
						CO	5.2	4.7		0.6	0.2	0.4	

study we compare the results of tests done at the factory before delivery of the analyzers, at the MLab before deployment in the field and at a selection of monitoring sites. Ideally, we would like to apply exactly the same protocols at the three locations, but due to time and design constraints all tests cannot be performed or the protocols have to be adapted. The different protocols are described hereafter. The common abbreviations used throughout the text and figures are redefined in Table A1 in the Appendix.

3.1 Factory tests

All instruments are tested before leaving the factory and a certificate of compliance is provided with the instrument. For all instruments and species (except H₂O), the continuous measurement repeatability, the short-term drift, the accuracy and the short-term repeatability are tested. The statistical parameters, used to characterize each instrument, are estimated

by the measurement of dry reference gases (target) and are defined as follows.

3.1.1 Continuous measurement repeatability

The continuous measurement repeatability (CMR, called precision in the certificate of compliance) is calculated as the average over 30 h of 5 min interval SD of raw data (frequency about 0.5 Hz).

3.1.2 Short-term drift

For the first two generations of instruments (ESP1000 and G1301), the short-term drift was the peak-to-peak amplitude of the 5 min averaged data over 30 h. From the third generation on, the drift is defined as the peak-to-peak amplitude of the 50 min averaged data over 30 h. For CO, the drift is again

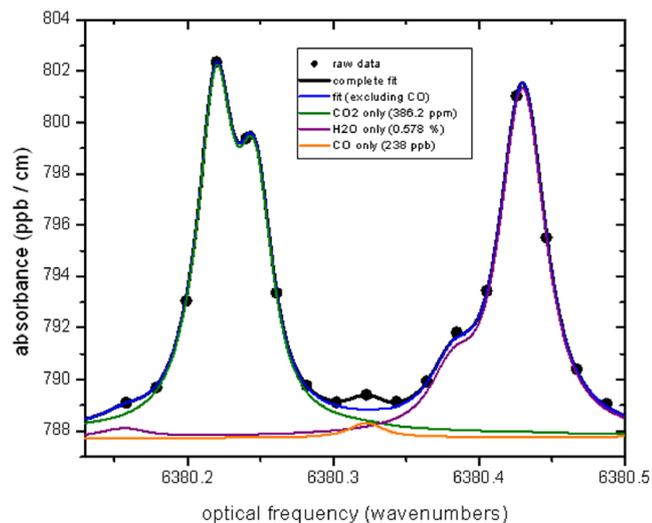


Figure 1. Spectral fits of absorption data (black solid dots) for CO_2 (four overlapping lines) only, H_2O (three lines) only, CO only, excluding CO , and all data. From Chen et al. (2013).

defined differently as the peak to peak amplitude of the 5 min averaged data over 24 h.

3.1.3 Short-term repeatability

The short-term repeatability (STR) is measured by cycling two gases at 10 min intervals for 2 h. The data from 8 min 50 s to 9 min 50 s are averaged and the SD of the averages calculated.

3.1.4 Accuracy

The accuracy against the factory internal scale was first evaluated with the mean of the raw data over 30 h then later over 30 min only. The results from this test are not discussed in this study.

3.2 MLab tests

Since the beginning of the MLab operation in 2008, the protocols and metrics used to evaluate the instruments have evolved. However, for our analysis, most of the data sets could be reanalyzed and the latest version of the protocols applied.

For the present protocol, the MLab keeps the instrument for about 1 month to perform all tests using dry reference gases calibrated in agreement with WMO scales, comparison to reference instruments and drying/humidifier system to evaluate the sensitivity to water vapor content. A detailed report is provided for each instrument. The cylinders are aluminum tanks (for the older ones used only as test cylinders: LUXFER, UK aluminum alloy 6061 with VTI Ventil Technik GmbH stainless steel valves, for all the newer ones: LUXFER, UK aluminum alloy 6061 with Rotarex membrane

valve (D200 type with PCTFE seal) with brass or stainless steel body) with brass or stainless steel pressure regulator from Air Liquide America Specialty Gases LLC (previously Scott). They are either filled with dry natural air or with dry synthetic air. The isotopic composition of these last cylinders is controlled to correct for any bias compared to natural air. Indeed, as the CRDS instruments are sensitive only to the major isotopologue ($^{12}\text{C}^{16}\text{O}_2$ or $^{12}\text{CH}_4$), a bias in the isotopic composition would lead to a bias in the mixing ratio. The synthetic air cylinders are used for linearity and calibration purposes, and their gas concentrations span the range of ambient air concentrations and are calibrated against the international WMO scales. The natural air cylinders are filled at the MLab using a RIX Industries oil-free compressor and used as test cylinders; their concentrations are close to the average ambient air concentrations and allow us to evaluate the performances of the instruments as defined hereafter. The gases in these cylinders are referred to as target gases.

For the first instruments, in order to develop meaningful tests, the stabilization time for each instrument was evaluated in order to know how long a cylinder had to be measured before being stable. This stabilization period has been determined to vary between 5 and 15 min depending on the water content of the previous sample and the length of its analysis (not shown). It also depends on the length of the sampling lines and on the design of the whole system (dead volume, flush volume, etc.), which is relatively uniform in the MLab but can vary from site to site. These first tests have helped to define the length of the measurement intervals for the now standard protocols. Also, inlet pressure tests have been realized to evaluate the influence of the inlet pressure onto the measurements. Results (not shown) highlight the importance of having a difference below 0.4 bar between the calibration gases and the rest of the samples to avoid significant inlet pressure influence which leads to systematic biases.

As part of the latest MLab protocol, 11 criteria are now evaluated for each instrument. These are defined below. Target gases are used to evaluate most of the metrics except the calibration, linearity and comparison with reference instruments.

3.2.1 Continuous measurement repeatability

The continuous measurement repeatability is evaluated with the SD of the continuous measurements of a cylinder over 24 h as described above.

3.2.2 Short-term drift

The short-term drift is defined as the peak-to-peak amplitude of the same measurements.

These two metrics are evaluated for different integration times (typically, raw data, 1 min and 1 h). Usually, in the synthesis report, we provide values for 1 min and 1 h averages.

3.2.3 Allan deviation

The Allan deviation, which shows the stability as a function of the integration time and informs about the optimal integration time, is also calculated and provided in the synthesis report.

These three metrics are illustrated in the Appendix in Fig. A1.

3.2.4 Short-term repeatability

The short-term repeatability is defined as the repeated measure of a sample over a short period of time (about 3 h). In the laboratory, a target gas is measured 10 times in 15 min sequences bracketed by 5 min of wet ambient air measurements. For each measure, only the last 9 min are averaged. The repeatability is then expressed through the mean and SD of these averaged measures.

3.2.5 Long-term repeatability

The long-term repeatability (LTR) is comparable to the short-term repeatability but on a longer timescale (3 days). In the laboratory, a target gas is measured for 30 min bracketed by around 5 h of wet ambient air over 72 h of total measurements. For each measure, only the last 10 min are averaged. The long-term repeatability is then expressed through the SD of these averaged measures. Typically, several 3-day exercises are performed and the results compared and aggregated at the end of the 1-month duration of the instrument test period. In Fig. A2 shows an example of short-term and long-term repeatability. For each species the mean, the SD and the drift are calculated.

3.2.6 Ambient temperature and pressure dependence

For the latest instruments, since 2013, the temperature and pressure dependencies have also been tested at the MLab. For the pressure, we plot the target gas measurements realized during the long-term repeatability test against the atmospheric pressure over several days and evaluate the correlation between the two. For the temperature dependence, the room temperature was until now varied using the room air conditioning system and we plot the target gas measurements against this varying temperature. Plans have been made to acquire a temperature-controlled chamber. As for the pressure, the correlation between the two is calculated. Two examples are shown in Fig. A3 with for each case, the linear regression and the correlation coefficients calculated.

3.2.7 Water vapor correction

An important test for the CRDS instruments is the water vapor correction evaluation. The applied factory correction is the same for all instruments even if not all of them have the same response to water vapor. It is also not always possible

to measure only dry air when in the field. Over the years, different tests have been applied to evaluate this correction, from comparing two instruments measuring the same air, one with a drying system, the second without, to the latest tests that progressively increase the humidity of the measured gas stream. The complete methodology and the results of these tests will be treated in a separate publication.

3.2.8 Calibration

For an operational network, it is crucial to report not only precise but also accurate data linked to each other by a common scale. If each instrument is usually calibrated at the factory, the calibration scale used is not linked to the international WMO standards. Moreover, a regular calibration allows us to correct for long-term drift in the instrument. In the laboratory, at least four calibration sequences are done to determine the calibration function that links the measured values to the assigned values. Three to four standard gases are measured one after the other at least four times for 30 min each calibration sequence (each set of the four cylinders measurement is hereafter called a cycle; see Fig. A4). Then the calibration function using a linear fit is calculated. The calibration standards are themselves calibrated against the international primary scale of each species (WMO X2007 for CO₂, NOAA04 for CH₄ and WMO CO X2004 for CO, Zhao and Tans, 2006; Dlugokencky et al., 2005; Novelli et al., 1994). Since 2008, six different secondary scales have been used. These calibrations are used for the comparison tests at the MLab.

3.2.9 Linearity

The linearity of the instrument is also evaluated. For the first instruments, the same cylinders as for calibration (four cylinders) were used. Then, two cylinders (low and high concentrated cylinders; see Fig. A5) were added to the set. The residuals from the fit are calculated, and their concentrations along with the correlation coefficient allow us to judge of the linearity of the instrument against the calibration scale. It is important to note that the validity of this test depends strongly on the proper assignation of the concentrations from each calibration cylinders, hence the importance of the link to international scales and the regular recalibrations of the MLab calibration cylinders against a “master” set of cylinders provided by the central calibration laboratories.

3.2.10 Comparison with reference instruments

Finally, ambient air measurements from each instrument are compared with other reference instruments maintained by the MLab. The MLab is located in Gif-sur-Yvette, about 50 km southwest of Paris. We are thus sampling suburban air with large variability as we are looking at 1 min averages. Initially, the CRDS analyzers were compared to the gas chromatograph system and if available to another CRDS ana-

lyzer in test. Since the end of 2011, most of the instruments have been tested against the same CRDS reference instrument for CO₂ and CH₄ (CFCD503). For CO, since the end of 2013, a CRDS reference instrument has also been chosen (CFKADS2127). The tested instrument measures wet and dry air and is compared to the MLab reference instrument which measures ambient air dried through a cryogenic water trap. This allows the checking of the factory and MLab water vapor correction and the estimation of the biases. In Fig. A6, the comparison for CH₄ is shown. The H₂O and target gas measurements allow a quality check of the tests. The histogram can point out outliers if the distribution is strongly not Gaussian. The difference between the wet corrected air and the dry air in the left panel (about 1.2 ppb on average compared to -0.03 ppb for both instruments measuring dry air) is due to the automatic H₂O correction, which is here not sufficient to correct all the bias introduced by H₂O.

3.3 Field tests

In the field, to estimate instrument performance, we use the calibration and target gases. Usually, one or two target gases are measured regularly for quality control purposes. If only one cylinder is used, then this cylinder is a so-called short-term target and is measured once to twice a day. If two cylinders are in use, then the second one will be the long-term target and will be measured at the same time as the calibrations, usually every 2–4 weeks. The short-term target tank lasts about 1–2 years; the long-term targets in principle last 10 years or more, allowing continuity throughout the station lifetime. It is important to note that the analysis chain setup has an influence that can be hard to separate from the instrument performances. At the sites in this study, the ambient air sampling lines are usually Synflex 1300 (EATON) outside of the shelter and stainless steel tubing inside. Ambient air can be dried using either a cryocooler or a Nafion membrane. A Valvo (Vici) valve is used to distribute the different gases to the instrument. In the next tests, we however try to evaluate the influence of the setup, either by looking at specific data or by using the MLab tests as a comparison. We investigate the stabilization time of the instrument for each sample measurement interval and for each calibration sequence. We also look at the evolution of the calibration equation and its residuals which give insights into the linearity and stability of the instrument. We estimate the field continuous measurement repeatability, short-term repeatability and the long-term repeatability using the target gas measurements. Finally, we study the instrumental drift as a function of gas pressure or temperature.

3.3.1 Stabilization time within one measurement interval

For the stabilization time for one measurement, we select the last measurement interval of the last tank of the calibration

sequence to avoid the influence of water from potentially humid ambient air samples and to ensure the flush and equilibration of the tank pressure regulators. We are indeed trying to look only at the performances of the instrument itself independently of the analysis chain setup. We calculate the minute averages within the interval and then the difference of the averages to the last minute of analysis.

3.3.2 Stabilization time within one calibration sequence

For the stabilization time within one calibration sequence, we compare the average of the last 10–15 min of each interval for the last cylinder to the last measurement interval.

3.3.3 Instrument long-term stability

We also look at the stability of the instrument by looking at the evolution with time of the calibration equation and evaluate whether the periods between the calibration allow us to capture this evolution. Finally, we look at the evolution of the linear fit residuals to investigate the linearity of the instrument over time.

3.3.4 Field continuous measurement repeatability equivalent

The target gases in the field are not measured continuously for 24 h. However, the short-term target is measured at least once a day for 20 to 30 min. Here, as an equivalent of the continuous measurement repeatability, we calculate the monthly average of the SDs of raw data over 1 min intervals.

3.3.5 Field long-term repeatability

For this value, we calculate the SD of the averaged target measurement intervals over 3 days as in the MLab; then we calculate monthly average of this number for graphical visibility.

3.3.6 Pressure vs. temperature as a source of instrumental drift

By studying the drift of the calibration constants of the instruments over time, we have an opportunity to study the behavior of this population of instruments over time. The following quantities are evaluated:

$$\text{CO}_{2\text{frac}} = \frac{a_{\text{CO}_2} \cdot 390 + b_{\text{CO}_2}}{390}, \quad (1)$$

$$\text{CH}_{4\text{frac}} = \frac{a_{\text{CH}_4} \cdot 1900 + b_{\text{CH}_4}}{1900}, \quad (2)$$

with a and b the slope and intercepts from the calibration fits, 390 and 1900 ppb reference ambient air mixing ratios and CO_{2frac} and CH_{4frac} the fractional change of CO₂ and

CH₄ concentrations compared to the reference mixing ratios. We consider two different sources of drift: gas pressure and gas temperature. Both quantities affect both the number density of molecules in the CRDS optical cavity, as well as the shape of the spectral line from which the mole fraction is derived. To measure the effect of temperature and pressure on the fractional change in carbon dioxide and methane, experiments have been performed at the factory on a single CRDS instrument (model G2301) in which temperature and pressure of the gas sample were changed and the fractional change in carbon dioxide and methane measured, to obtain the following values:

$$\frac{\Delta\text{CO}_{2\text{frac}}}{\Delta T} = -4.6 \times 10^{-3} \text{ } ^\circ\text{C}^{-1}, \quad (3)$$

$$\frac{\Delta\text{CO}_{2\text{frac}}}{\Delta P} = +1.3 \times 10^{-3} \text{ Torr}^{-1}, \quad (4)$$

$$\frac{\Delta\text{CH}_{4\text{frac}}}{\Delta T} = -5.3 \times 10^{-3} \text{ } ^\circ\text{C}^{-1}, \quad (5)$$

$$\frac{\Delta\text{CH}_{4\text{frac}}}{\Delta P} = +3.5 \times 10^{-3} \text{ Torr}^{-1}. \quad (6)$$

We see that temperature drift has a fundamentally different character than pressure drift. That is to say, when the temperature drifts, the fractional change of CH₄ vs. the fractional change in CO₂ has a slope of $5.3/4.6 = 1.15$; for pressure drifts, this ratio is $3.5/1.3 = 2.7$. This difference in ratios means that we discriminate between the two mechanisms, by looking at the slope for each instrument data set.

4 Results

4.1 Allan deviation

The first test in the MLab is the continuous measurement repeatability measurement which allows us to draw the Allan deviation vs. the averaging time and already gives good insight into the stability of the instrument. In Fig. 2, the Allan variance is plotted for all instruments and for the three species. Averaging times of 1 min and 1 h are plotted as vertical solid black lines. For all instruments, the Allan deviation decreases and on average reaches its minimum around 1 h of averaging time. We generally observe that the first generation of instruments or the first instruments of the next generation perform less well for the smaller averaging time. For example, we see for CO₂ and CH₄, that below 1 min averaging time, ESP1000 and G1301 have a higher Allan deviation than the G2301 except for one G2301 which happens to be the first purchased and tested in the laboratory. In the same way, the first G2401 performed less well than the next instruments. For CO, we see the same pattern with the first CO instruments (G1302) performing less well than the next generation. For the two instruments that were upgraded for water–CO₂–CO cross-talks only after the tests, we also found that for CO₂ they performed less well than the previous generation (Allan deviation above 0.03 at 2 s). Looking at the

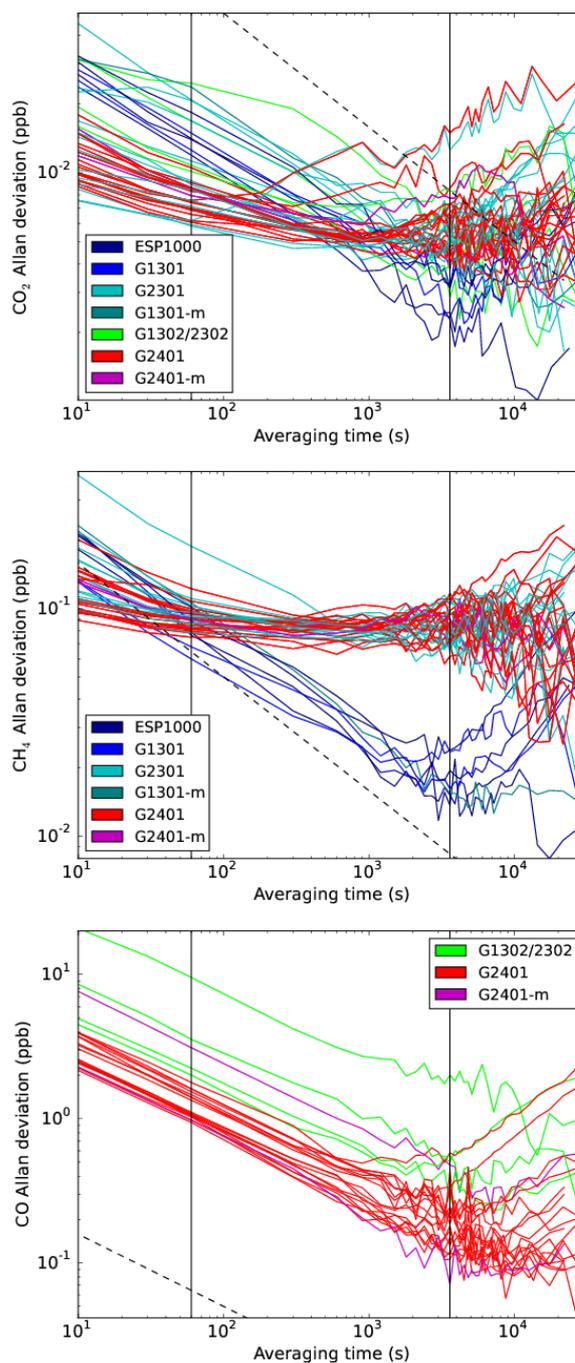


Figure 2. Allan deviation for CO₂, CH₄ and CO. The vertical solid black lines show 1 min and 1 h averaging times. X and Y axes are in logarithmic scales. The black dotted line shows the white noise $\left(\frac{1}{2\sqrt{\text{time}}}\right)$.

upgraded instrument (Allan deviation of 0.02 at 2 s), we see that this has probably been corrected for with the upgrade. For a longer averaging period between 1 min and 1 h, the results are much less different, except – as said before – for the two G1302 instruments tested before upgrade. It is also in-

interesting to note that for CH₄ the first models ESP1000 and G1301 performed better than the following models, which is thought to be due to a change in the electronic design. Since this study, a thorough investigation concerning the electronic design changes and the effect on CH₄ averaging has been conducted by the manufacturer. The issue has been identified and corrective actions have been taken, restoring the expected performance in terms of CH₄ averaging according to the manufacturer. This will have to be confirmed when testing new instruments. It is important to note that for a given model the performances are very consistent with very few outliers, especially for the latest models.

4.2 Comparison of the results from the factory, the MLab and the field

In this section, we look at the performances of the groups of instruments using some of the metrics defined in the previous sections. Results from the factory, the MLab and the field sites are shown in Table 1 and in Figs. 3–6. Only the MLab continuous measurement repeatability using raw data is presented in order to compare the three values (first panel of Figs. 3–5). The short-term repeatability cannot be calculated in the field but is evaluated at the factory and at the MLab (second panel). Then, the long-term repeatability is not evaluated at the factory but is calculated at the MLab and in the field (third panel). Finally, the fourth panel highlights the dry ambient air comparison (mean difference) between the instruments tested at the MLab and the reference instruments. For the continuous measurement repeatability, we observe that for CO₂ and CH₄ the factory values are usually slightly lower than in the MLab but with a negligible difference (−0.003 ppm for CO₂ and −0.06 ppb for CH₄ on average when the three first outliers are excluded). There are three outliers for the first three ESP1000 where the factory values are about 2 times higher than the MLab values. For CO, the sign of the difference is not systematic and the difference is small on average (0.14 ppb) if we exclude one outlier (one G1302). For the short-term repeatability, we observe higher values at the factory than the MLab (0.024 for CO₂, 0.15 for CH₄ and 0.5 ppb for CO (with one outlier excluded for CO) on average), which can be due to the different protocols (6 times 10 min with only 1 min kept against 10 times 15 min with 9 min kept). The difference is indeed relatively constant over the instruments especially for CO₂ where this difference is about twice the noise of the instruments as measured at the MLab. For the long-term repeatability, the field results are very close to the MLab results, showing that the instruments kept their performances over time on average and that the linear regression for the field calibrations stays valid with time.

Finally, in terms of bias to the reference instrument, the average of the difference is usually within the WMO compatibility goals.

For all species, and especially CO₂ and CO, we can see that the manufacturer did improve the performances of its instruments, which is shown in continuous measurement, short-term and long-term repeatability tests. Moreover, the performances are more reproducible with smaller dispersion of the results. In general, the precision has been improved by a factor of 2 to 3 as can be seen from Fig. 6, which presents the synthesis of the instrument performances averaged by model.

4.3 Ambient temperature and pressure dependence

Most of the instruments tested at the MLab show a limited sensitivity to temperature and pressure. However, some instruments present a higher dependence. In the case of the temperature changes, if these changes are slow and within the range guaranteed by the manufacturer (10–35 °C), the instruments are temperature independent. However in case of rapid temperature changes, like in an airplane, a temperature dependence can appear for CO₂ and CH₄. It is caused by the fact that the cavity cannot regulate its temperature as fast as the outside changes. As the mixing ratios are calculated assuming a fixed temperature, small changes in this temperature lead to biases. Of all the 29 instruments tested for temperature dependence, only 2 presented a significant temperature dependence ($R^2 > 0.5$) using rapid changes. This is shown on one mobile instrument (CFKBDS2132) tested at the MLab as described above and using a temperature-controlled chamber that allows us to set different ramps and steps (see Fig. 7). In the shown experiment (lower panel of Fig. 7), a target gas was measured for 16 h with the temperature varying from 10 to 35 °C. The temperature was increased by 5 °C / 30 min ramps with 1 h steps. We observe a clear dependence ($R^2 > 0.5$) for CO₂ and CH₄ with 0.01 and 0.03 ppb °C^{−1} respectively. With longer ramps (not shown), the dependence was smaller but still the species were highly correlated with the room temperature. With the usual rapid change test (upper panel of Fig. 7), this dependence was also clear and comparable.

In the case of atmospheric pressure variations, CO₂ and CH₄ are not significantly affected, but for some instruments CO presents a significant dependence. At the MLab, out of 16 instruments measuring CO that undertook the pressure dependence tests, 5 presented a dependence higher than 4 ppb hPa^{−1}. One of them (CFKADS2084) presented a dependence of more than 7 ppb hPa^{−1} at the first test and was then sent back to the factory for upgrade. After upgrade, the dependence was significantly reduced to 0.04 ppb hPa^{−1} as can be seen in Fig. 8.

4.4 Calibration and linearity

To evaluate the linearity of the instruments, we have to be confident in the assignation of our cylinders. In the past, we have used different instruments (GC, Fourier transform in-

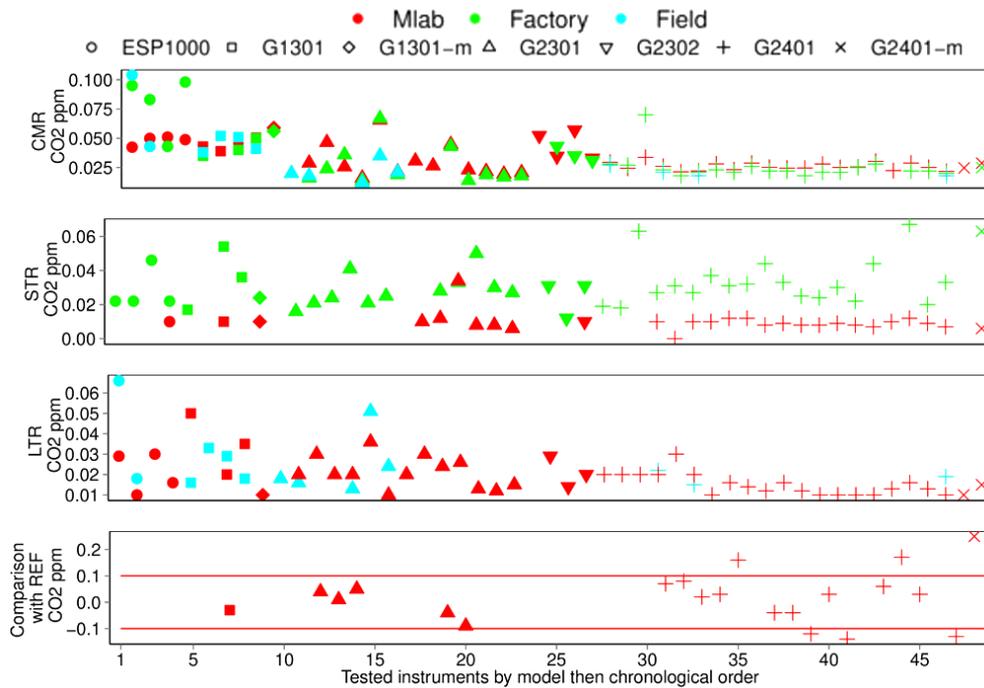


Figure 3. Results for CO₂. First panel: continuous measurement repeatability (CMR) at the factory (as defined), at the MLab (on the raw data) and in the field (as defined). Second panel: short-term repeatability (STR) at the factory and the MLab. Third panel: long-term repeatability (LTR) at the MLab and in the field. Fourth panel: comparison with reference instrument, average difference at the MLab. The horizontal lines show the WMO compatibility goals.

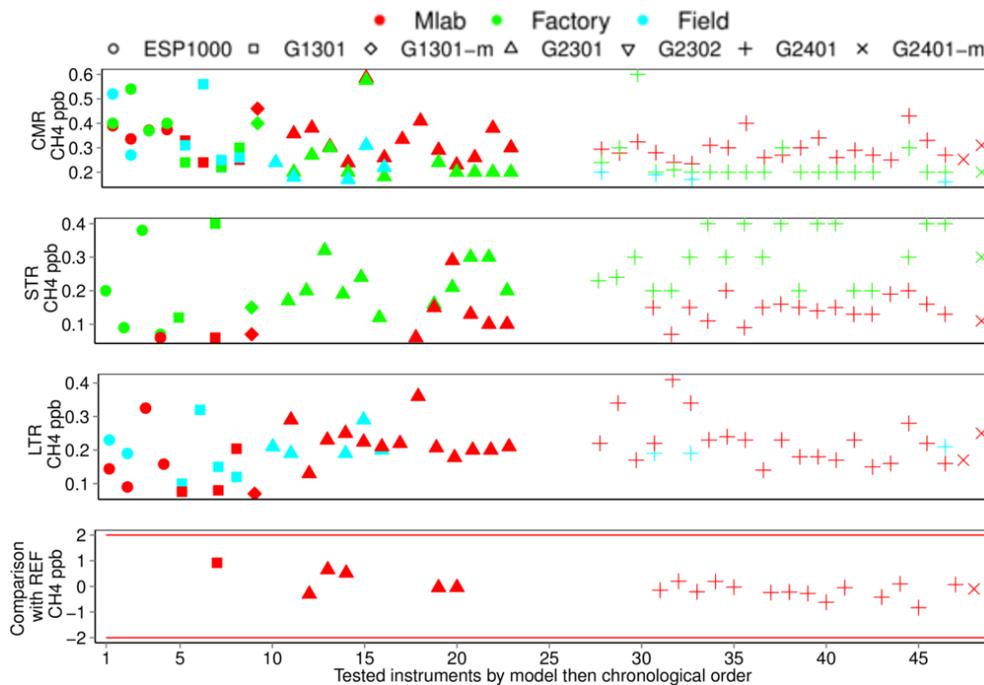


Figure 4. Results for CH₄. First panel: continuous measurement repeatability (CMR) at the factory (as defined), at the MLab (on the raw data) and in the field (as defined). Second panel: short-term repeatability (STR) at the factory and the MLab. Third panel: long-term repeatability (LTR) at the MLab and in the field. Fourth panel: comparison with reference instrument, average difference at the MLab. The horizontal lines show the WMO compatibility goals.

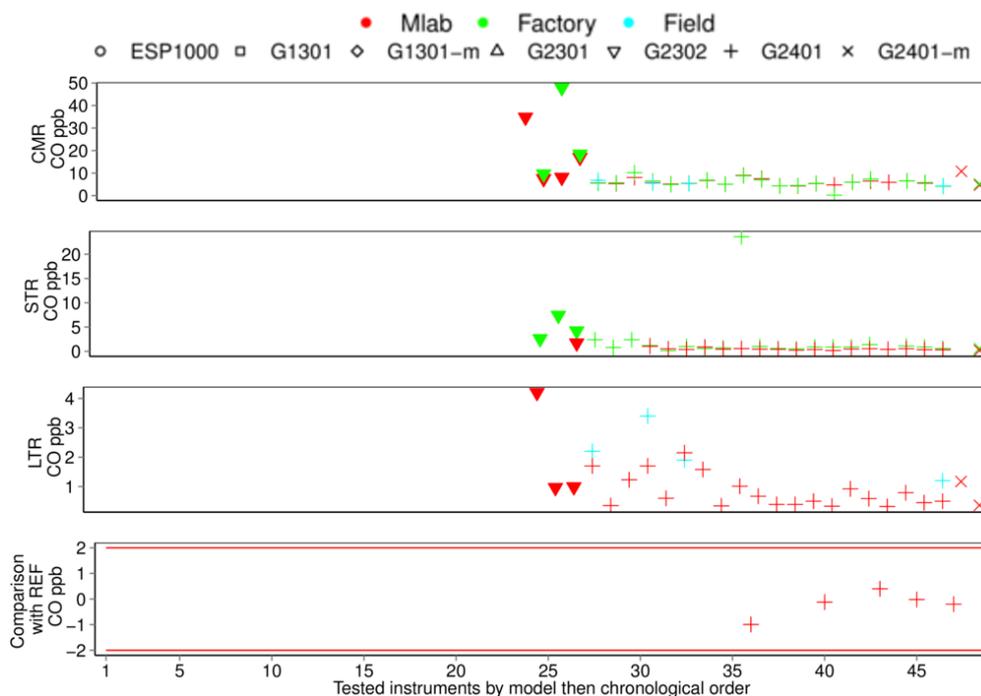


Figure 5. Results for CO. First panel: continuous measurement repeatability (CMR) at the factory (as defined), at the MLab (on the raw data) and in the field (as defined). Second panel: short-term repeatability (STR) at the factory and the MLab. Third panel: long-term repeatability (LTR) at the MLab and in the field. Fourth panel: comparison with reference instrument, average difference at the MLab. The horizontal lines show the WMO compatibility goals.

frared spectroscopy) and set of cylinders to assign the values of the MLab calibration sets. We are still in the process of reevaluating these values when possible. Here to assess whether the amplitude of the observed residuals are due to the instruments or to the calibration scales, we have plotted the residuals of each cylinder for the 14 instruments that used the latest set of calibration cylinders. In Fig. 9, we see that the residuals behave the same way for every instrument. For each assigned concentration, we found similar residuals for every tested instrument. Either all instruments behave exactly the same, or (the most probable hypothesis) the MLab calibration cylinders are not perfectly assigned. However, for this scale at least, we see that the residuals are about the same as the instrument precision especially for CO and CH₄. This allows us to be confident in the linearity of the instruments, as the difference between the linear fit and the assigned values (the residuals) is on the order of magnitude of the precision of the instruments, and that part of the amplitude of the residuals is due to the assignment of the calibration cylinders.

4.5 Evolution of the metrics with time in the field

Out of the 47 instruments tested at the MLab, we present results for 13 that have been installed in the field on sites instrumented for at least 1 year (see Tables 1 and A1, Fig. 10) plus two more instruments that were not tested in the laboratory. One is running in parallel at MHD with one tested instru-

ment, and the other one has been installed at a site (PUJ) for several years without major troubles. The complete measurement setup (length of sampling lines, buffer volume, drying system, measuring time) is not uniform and can differ from site to site. However all instruments are regularly calibrated with cylinders linked to the WMO scales and use at least one target tank. Also all the raw data have been processed with the same algorithm (Hazan et al., 2015).

The earliest data set begins in September 2008, and we end our study on 30 September 2014. As detailed in Table 1, we have analyzed field data sets from 11 three-species and 4 four-species instruments. These last instruments have been installed for a year or less to replace previous instruments. The instruments have been running at nine different stations: four in France, three in Europe, one in Africa and one in the Indian Ocean. For two stations, MHD and OPE there are two instruments running in parallel, while at the other stations, there is either only one instrument since the beginning (AMS, IVI, PUY, TRN, PUJ) or either several instruments replaced one after the other (three instruments at BIS, two at LTO). At OPE, one of the three instruments was replaced.

In Fig. 10, we see that we regularly have invalid data. At the day scale, the invalid data are mostly due to the flushing time of the measurement lines, especially when the site samples at several heights. However, in this graph, we are plotting daily average. If more than 50 % of the

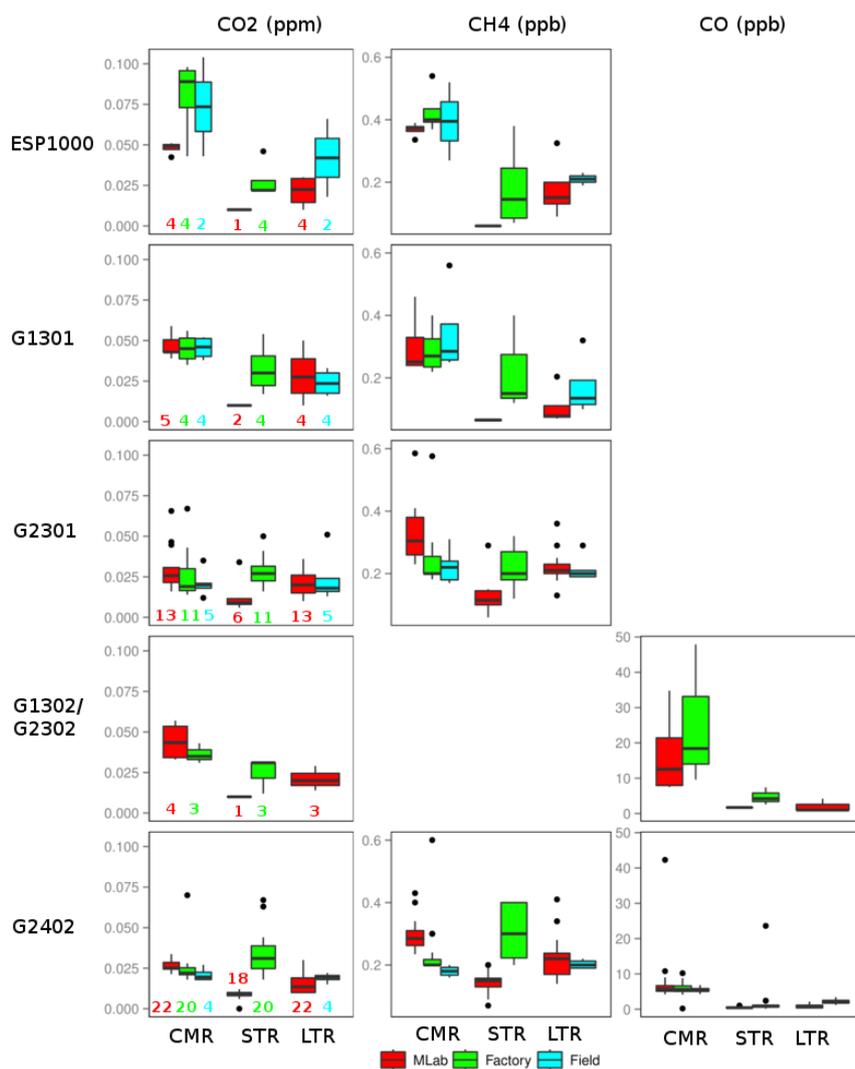


Figure 6. Whisker boxplots summarizing the CMR, STR and LTR tests for CO₂ (first column), CH₄ (second column) and CO (third column) by models. The mobile instruments are grouped with their main model. G1302 and G2302 are grouped together. The middle horizontal shows the median, the limits of the box the 25th and 75th quartiles, and the end of the whiskers the lowest datum still within 1.5 interquartile range (IQR – difference between the first and third quartiles) of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. The numbers under the boxplots in the CO₂ column indicate the number of instruments tested in each case.

data are valid, then the day is flagged as valid; if not, then the day is flagged invalid. With this way of calculating, problems will be the most likely cause of the invalid data. This can be caused by several factors (leaks in the lines, frozen water traps, local contamination, etc.) but also by an instrument failure. Usually, small periods of invalid data will be due to problems in the setup while longer invalid periods will be linked to an instrument failure. We have listed in Table 2 the main instrument failures or problems that happened for the shown field instruments as well as for the MLab references, the models to which it happened and the usual solution. Failures from other instruments, with their proposed solutions, are compiled in the ICOS-INWIRE Report GA N313169 (<http://www.icos-inwire.lsce.ipsl.fr/>

documents/10179/19756/ICOS-INWIRE+report+D2.1/). For the instruments installed on site for more 6 months (11 of them), on average, more than 80 % of the data are valid.

In the next figures the data from these instruments are compiled. We use first data from calibration gases. Data from the same site are in the same shade of color. In Fig. 11, we investigate the measurement interval and calibration cycle stability of the instruments. In the database, each instrument measurement protocol is configured according to the station setup. For example, the first minutes (usually half the time of the length of the measurement interval – so 10 to 15 min) are invalidated; for a calibration sequence, the first cycle is invalidated and at least three cycles are needed to have a valid calibration. In the left panel of Fig. 11, we show the minute

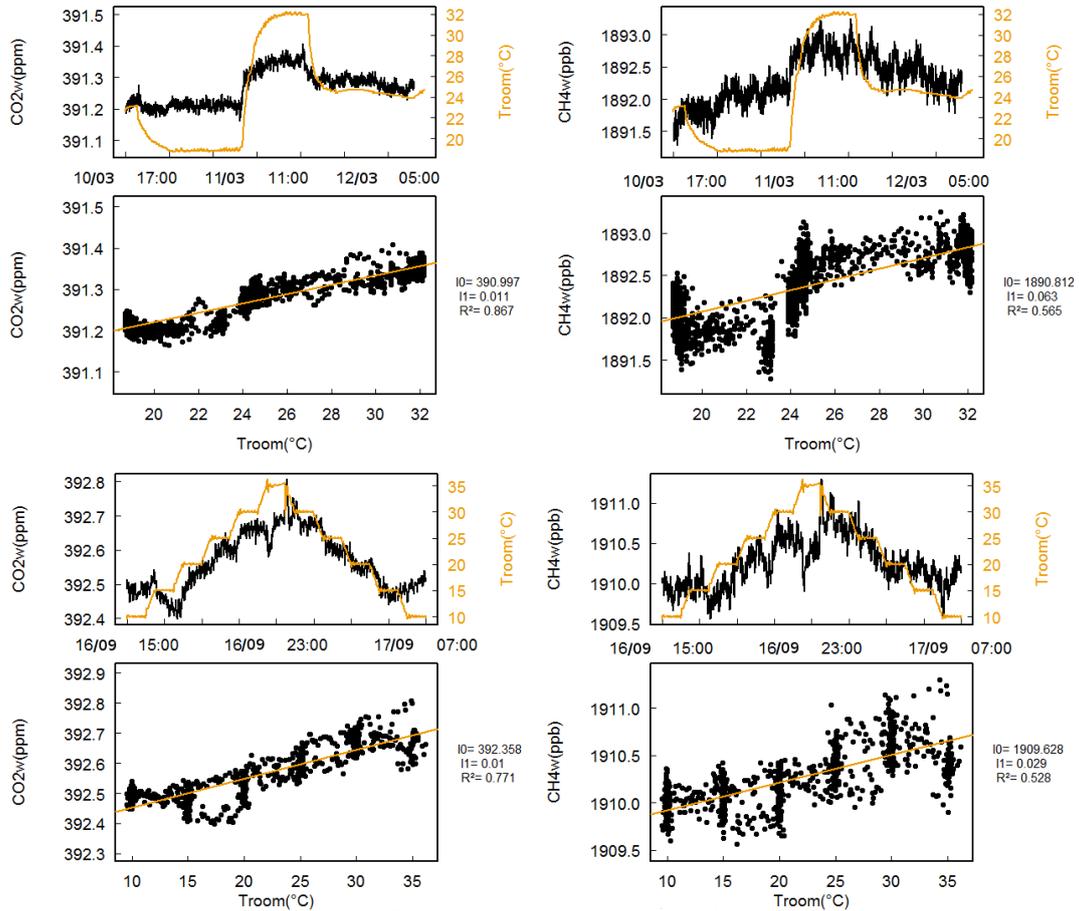


Figure 7. CO₂ and CH₄ temperature dependence of the same mobile instrument (CFKDBS2132) tested at the MLab (top four panels) and in a temperature-controlled chamber (bottom four panels) using in both cases rapid variations of temperature. Among the tested instruments, another one was also showing a strong temperature dependence.

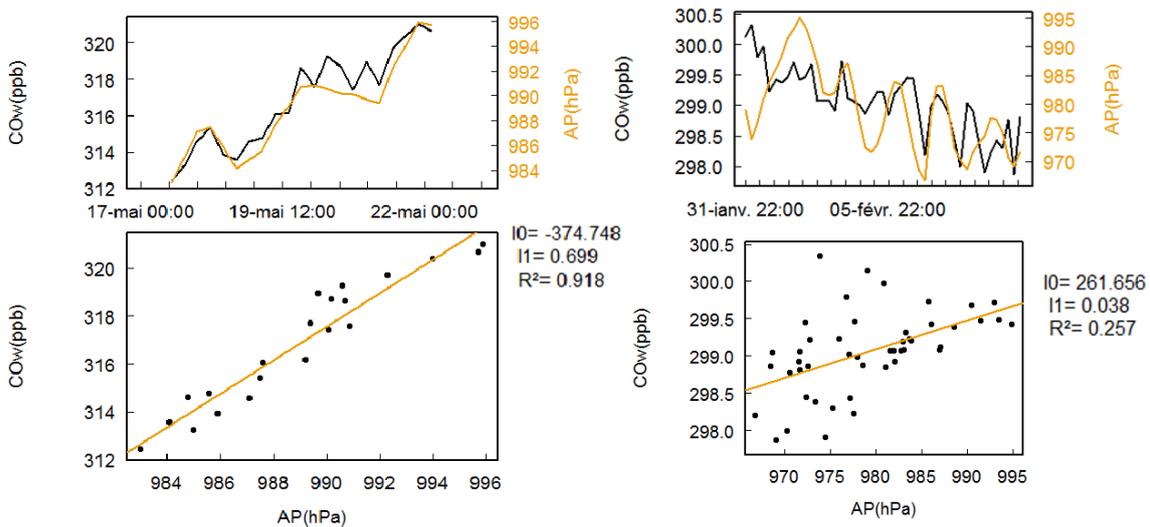


Figure 8. CO pressure dependence before (left) and after (right) repair at the factory for CFKADS2084. Among the tested instruments, four others were showing a dependence higher than 4 ppb hPa⁻¹.

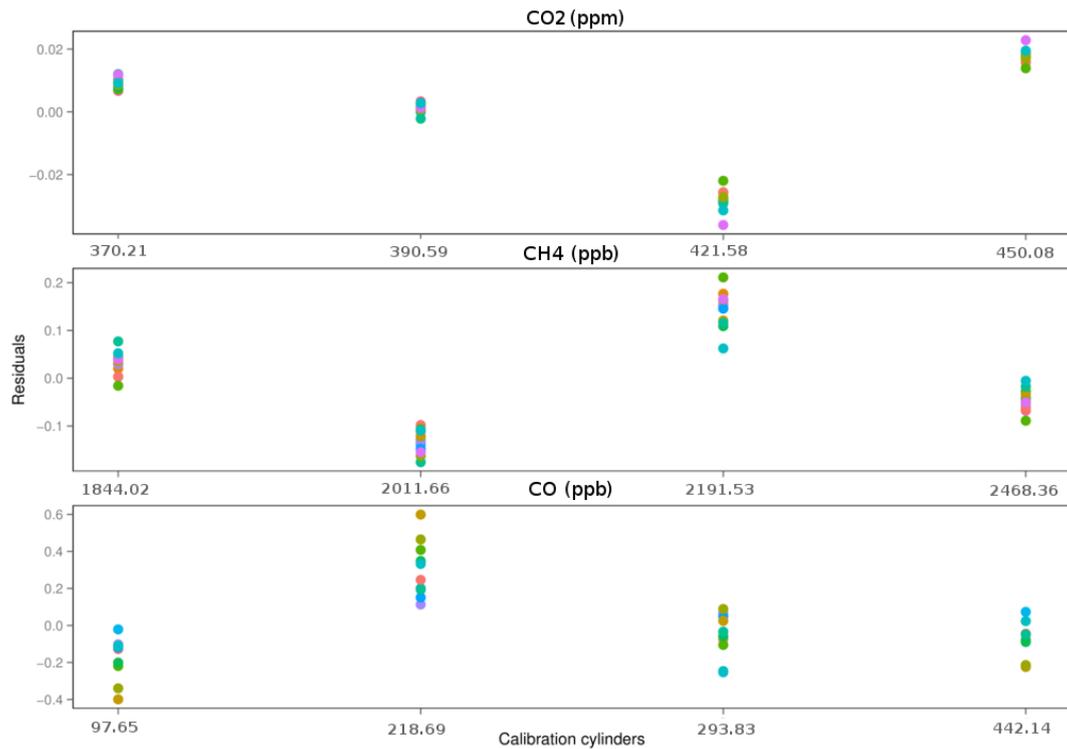


Figure 9. Calibration residuals for one set of cylinders measured on 14 different instruments. Each point is the average residual for each instrument for one calibration cylinder. For each cylinder, the reference concentration is indicated.

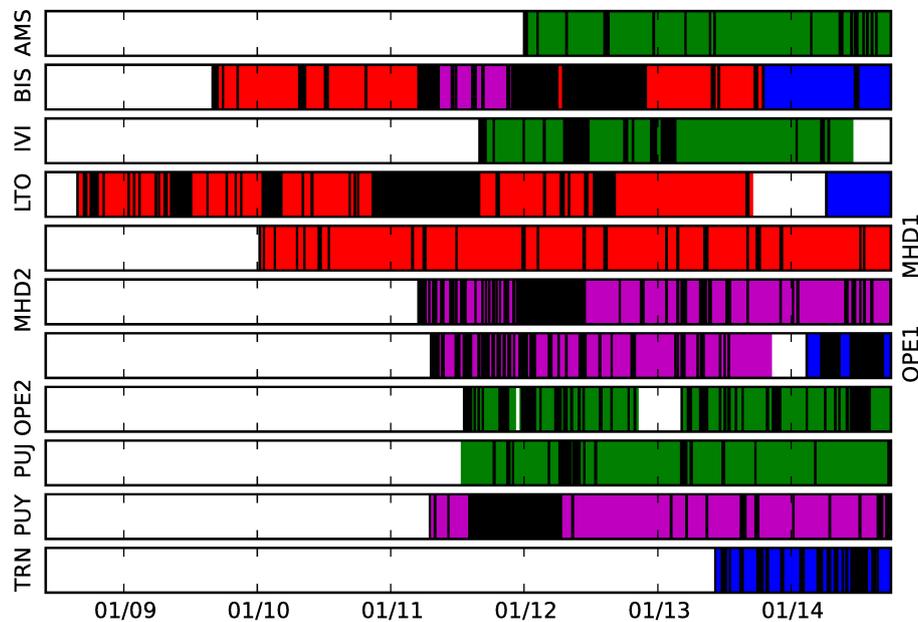


Figure 10. Daily average of data availability at each site (if 50 % of data are valid for 1 day, then the day is valid; if less, the day is invalid). Valid data are in color; invalid data are in black (% are indicated in Table 1). ESP1000 are in red, G1301 in magenta, G2301 in green and G2401 in blue.

averages of the mean difference to the last minute. The first point is at 3 min after the beginning of the interval. We see that, for CO₂ already after 5 min, the difference to the last

minute is below 0.05 ppm. For CH₄, this difference is below 0.2 ppb. For CO, however, the difference does not always decrease with time and can be above 2 ppb after 20 min of mea-

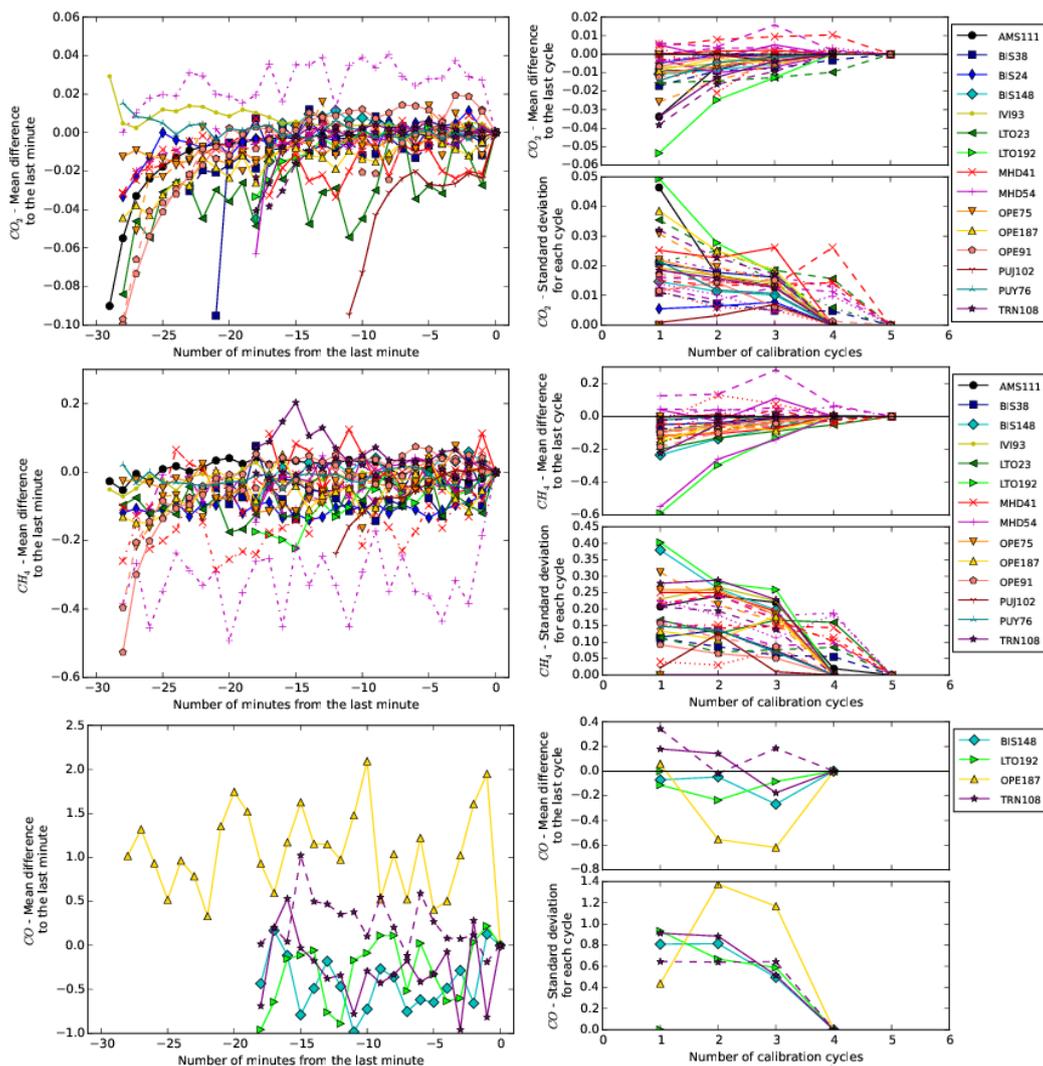


Figure 11. Top: CO₂ (ppm). Middle: CH₄ (ppb). Bottom: CO (ppb). Left: mean difference to the last minute for one measurement interval. Right: mean difference to the last cycle and SD of a calibration sequence. Dashed, dotted or dash-dotted lines show a different set of calibration cylinders for a same site.

measurements. It seems that, for this species, the noise is greater than the flushing influence. Indeed, in the MLab, the typical CMR is around 5 ppb.

In the right panel of Fig. 11, we look at the calibration cycles. Each point is the average of the validated minutes of one measurement interval, with one interval lasting 15 to 30 min depending on the site. For CO₂ and CH₄, we see a clear decrease of the difference with the cycles while for CO, as before, this is not as clear. At the second cycle, the difference for CO₂ is below 0.03 ppm and for CH₄ below 0.3 ppb. For CO, the difference is variable but stays under 0.6 ppb for any cycle. This indicates that to lower costs (less gas consumption) or at sites where there are several heights to sample from, to sample more ambient air, we could use fewer calibration cycles for the same quality. For example, to be under

0.02 ppm error for CO₂ and 0.2 ppb error for CH₄, in most of the cases, two cycles are enough.

For CO₂, we also see that there is no large difference between the instruments except for LTO192, TRN108 and AMS111 that show a longer stabilization time. For LTO192, this can be explained by the setup of the station, with all the gases going through a Nafion dryer, which dries to 0.1 % H₂O. The first measurement interval of calibration gases is thus still wet, and it was shown in the MLab that, after the H₂O correction, there was still a bias. For AMS111, the short-term repeatability is of 0.04 ppm so the difference to the last cycle is in the instrument STR. For TRN108, we observe the same long stabilization time for both sets of cylinders; as the difference for the first cycle is twice the instrument STR, this seems to be due to the setup, but there is no easy explanation for LTO. For CH₄, we observe longer stabi-

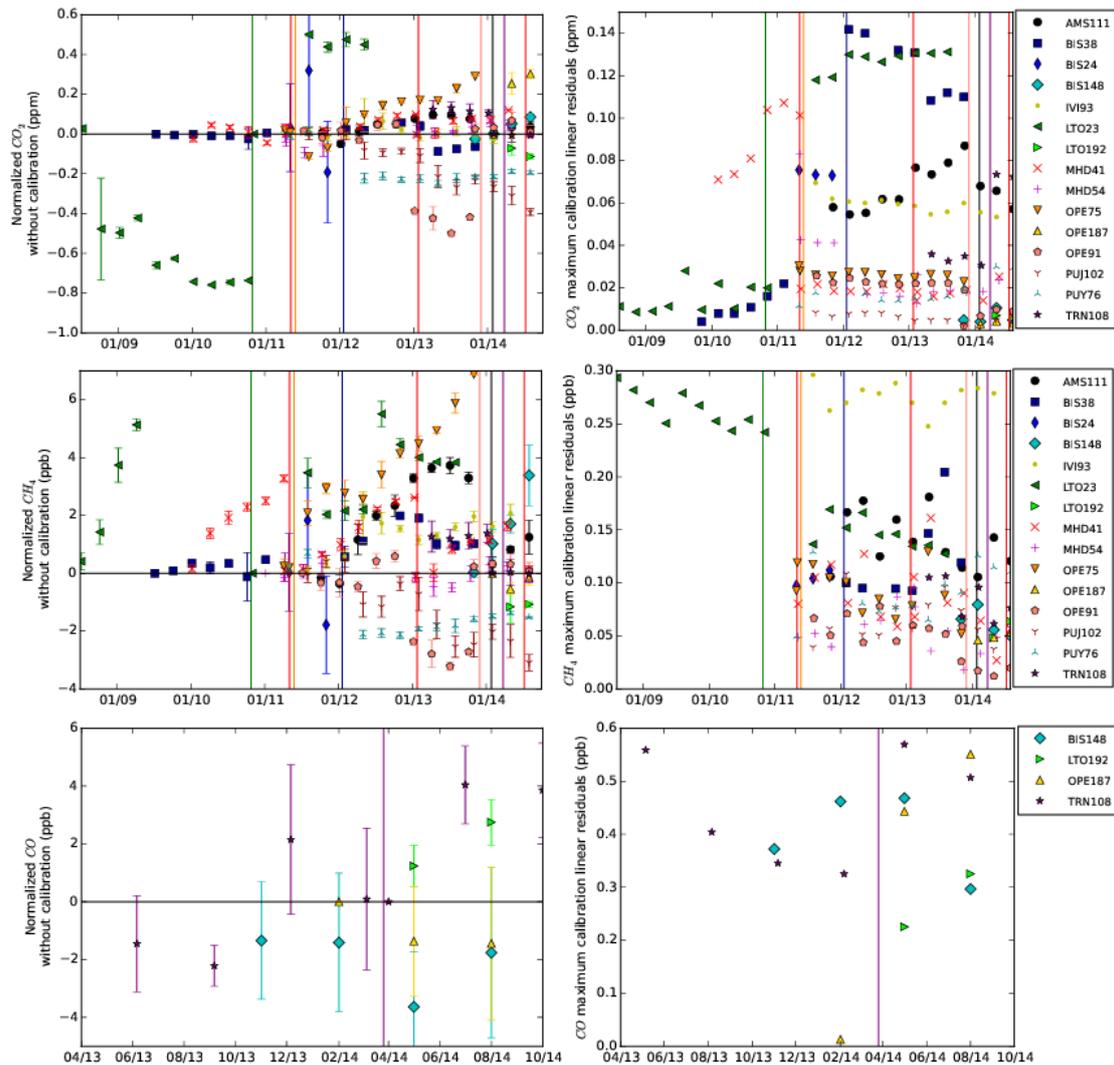


Figure 12. Top: CO_2 . Middle: CH_4 . Bottom: CO . Left: temporal drift of the instrument calculated using a virtual tank with a fixed value after calibration and the calibration equations – the concentrations are normalized using the first calibration. Right: temporal evolution of the residuals from the calibration linear fit. The vertical solid lines show a change of calibration scales. For all the panels, each data point is a 3-month average.

lization time again for LTO192 as could be expected and also for MHD54 but only for one set of cylinders. For this last instrument, the longer stabilization time is most probably due to a leakage in the tubing that was fixed when changing the cylinders. For CO , we also see that one instrument (OPE187) performs less well than the others both for the interval and the cycle stabilizations. We do not see the effect of the Nafion dryer on CO at LTO, and indeed in the case of CO no bias was observed after H_2O correction. For all three species, we see that the stabilization time seems to depend more on the setup than on the initial performances of the instruments as could be measured by the STR.

In Fig. 12, we plot on the left panels the instrumental setup drift over time and on the right panels the evolution of the residuals from the linear fit of the instruments. The first plot

allows us to investigate the need for more or less frequent calibrations while the second checks if the instrument response stays linear over time. For legibility, we have excluded one site where the values were so different that the scale was too large to see the trends (BIS24 for CO_2).

To evaluate the drift, using the calibration equation, we calculate the raw values of a virtual cylinder that once calibrated would have a fixed value (390 for CO_2 , 1900 for CH_4 and 150 ppb for CO ; see Eq. 7).

$$C_{\text{raw}} = \frac{C_{\text{cal}} - b}{a}, \quad (7)$$

with C_{raw} , the raw values, C_{cal} , the calibrated values we assigned, a , the slope of the linear regression and b the intercept. We normalized the concentrations by subtracting

Table 2. Instrument failures, models concerned and solutions for the field instruments of this study.

Failure	Models	Solutions
Decrease of the data sampling frequency due to laser aging	ESP1000, G2401, G2402	Laser tuning
Delay in recording the data that can lead to loss of data if not caught up by the instrument computer, usually due to old age, sometimes with freezing of the computer	ESP1000, G1301	Reboot the instrument with possible loss of data, change of the hard drive
Failure of temperature regulation	ESP1000, G1301	Back to factory
Loose electric connectors that lead to contact failure, effects depend on which connector is loose	G2402	Change connectors, lengthen the cables to lessen the tension
Software bugs with erroneous messages	All	Usually solved by a complete restart

the first value. Over time, for most of the instruments, the variations are slow and regular but still significant (up to 0.15 ppm yr^{-1} for CO_2 for OPE75, up to 2.2 ppb yr^{-1} for CH_4 for the same instrument; for CO , there is not enough data to observe trends), and the regular calibrations allow us to follow these changes. These rates are comparable but lower than rates from other studies such as Richardson et al. (2012) and Karion et al. (2013), which found drift rates of 0.15 ppm yr^{-1} for CO_2 in the first case and 0.25 for CO_2 and 3.4 ppb yr^{-1} for CH_4 in the second study. For some instruments however, there were some rapid changes and we can question the calibrated values for these periods. For example, for OPE91, we observe a sharp decrease in January 2013. Looking at Fig. 10, we see that this instrument was not providing data before this modification. Indeed, the instrument had a failure; after repair, the calibration response was different. It is notable that, even after a simple restart, the instrument response can differ from before due to some startup processes that are not reinitialized. The sharp changes in the LTO23 instrument are also usually linked to problems either in the instrument or the setup that modified the calibration response. It is important to note that the drift is not only due to the instrument but could be due to leaks in the lines, drifts in the calibration tanks, etc. However, we observe a general trend toward increase of the concentrations for CH_4 , which seems to indicate an instrumental drift for this species.

On the right panels of Fig. 12, we plot the largest residuals from the linear regression to evaluate the linearity of the instruments over time and whether the linear approximation stays valid for the tested instruments. We see that for all three species, for most of the sites, the answer is yes, with stable residuals over time. However, for AMS111, BIS38 and MHD41, we observe a significant drift over time. After the change of the calibration scales, the drift is corrected for MHD41. For BIS38, after the change, there are only three calibrations before being sent back to the factory and three

after. It seems however that, after repair, there was no drift anymore. For AMS111, there are also only three calibrations which present no or an inverse drift. It seems then that the observed drifts are most probably due to a drift in the calibration scale and not in the linearity of the instruments.

In Fig. 13, we show the average field continuous measurement repeatability and the long-term repeatability. For this, we use data from the target gas measurements. For all species, the CMR and LTR show very few variations over time (not shown), and thus we have chosen to show the data using boxplots summarizing the instrument performances over the whole period of measurement. For comparability, we have added the MLab continuous measurement repeatability and the long-term repeatability. We observe that, for most of the sites, the laboratory values and the field values agree well. The field continuous measurement repeatability tends to be higher than the long-term repeatability as can be expected.

In Fig. 14, we plot the fractional change in the CH_4 concentration vs. the fractional change in the CO_2 concentration. The data from the different instruments are normalized to the first measurement in the time sequence and are offset horizontally from one another for clarity. We only considered continuous sequences when the calibration standards were not changed. Changing calibration standards caused jumps in the observed ratios, which is due to differences in assigned values for the standards. Whenever the standards were changed, a new data set was generated (labeled “a”, “b”, etc. in the figure legend). In addition, we created a new data set in instances when the instrument was sent back to the manufacturer for repair or was repaired on site, since the repair work can affect the instrument calibration.

We see that the fractional change is on the order of 0.001. In other words, the drifts of the CRDS instruments are typically about 0.1 % over a year. The magnitude of the fractional change is larger for methane than for carbon dioxide,

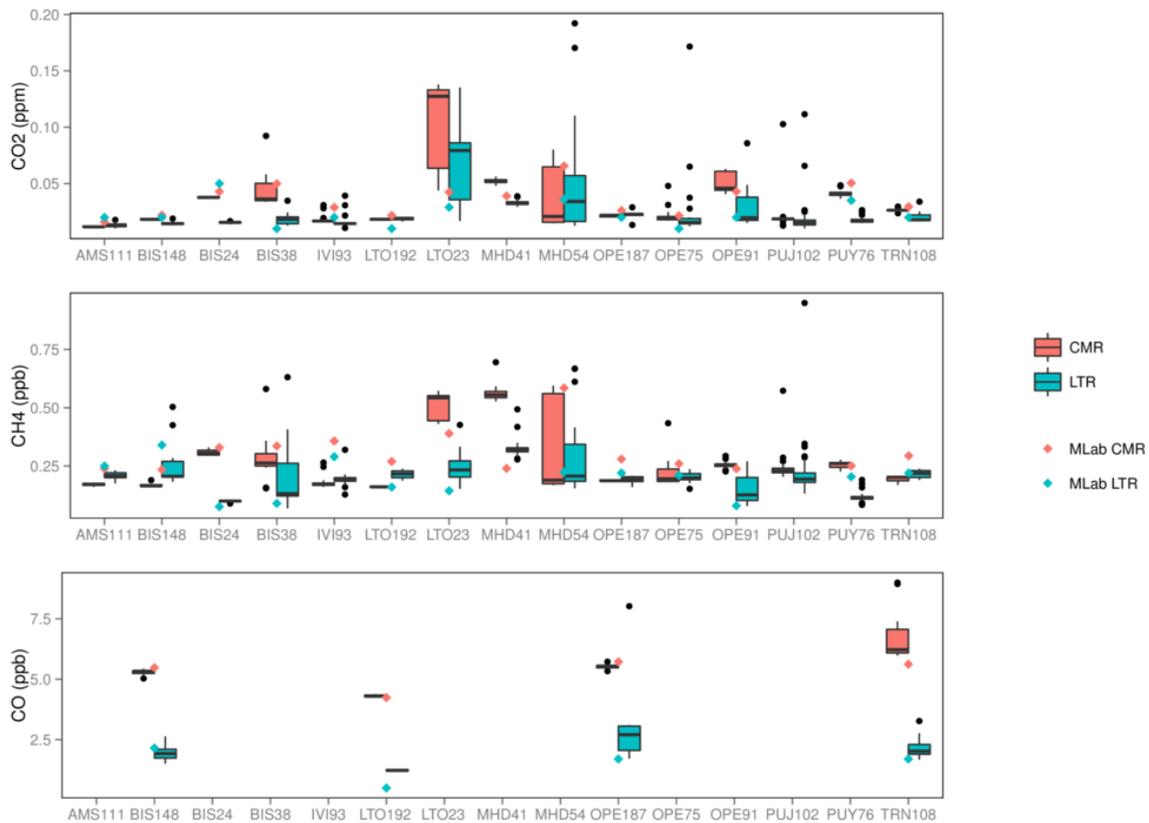


Figure 13. Top: CO₂. Middle: CH₄. Bottom: CO. For each field instrument, we show the statistics as defined in Fig. 6 for the CMR and LTR during the measurement period. The MLab initial values are added with diamond shapes.

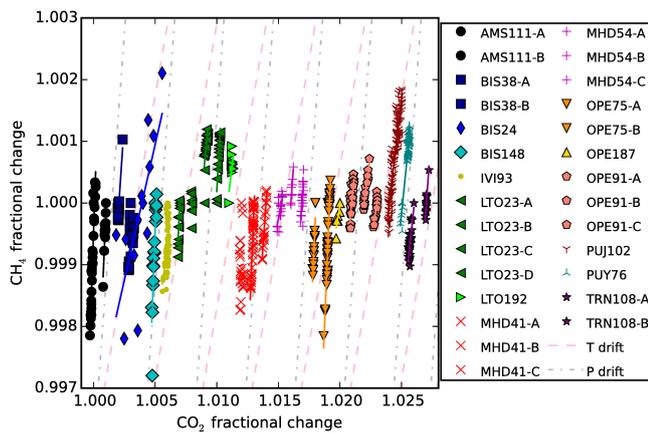


Figure 14. Data showing the fractional change in CH₄ vs. the fractional change in CO₂, as derived from instrument calibrations over time. Each data set has been offset horizontally for clarity, and normalized to the first calibration data. The linear fit is plotted over the data. The pink dashed lines indicate the slopes corresponding to temperature-dependent drift, and grey dashed-dotted lines to pressure-dependent drift.

same period of time, the mean fractional change is 0.13 %, or 2.5 ppb. These values are on the order of the compatibility target for these two gases. In nearly all cases, the drift in the fractional change in methane is well-correlated to the drift in the fractional change in carbon dioxide. We can see this in the linear fits to the individual data sets, which have a median R^2 of 0.79. This high degree of correlation implies that the drift is caused by drift in some quantity or set of quantities that affect carbon dioxide and methane proportionally. Finally, looking at the ratios, the grey dashed lines follow the pressure-derived slope of 2.7, and the pink dashed lines follow the temperature-derived slope of 1.15. It is clear from the figure that the data tend to follow the pressure-derived slope rather than the temperature-derived slope; this observation is supported by the fact that the mean slope from the linear fits in the figure is 2.4, which is close to the predicted value of the slope for a pressure drift. A more careful analysis of the data, by vector decomposition along the pink (temperature) and grey (pressure) axes, implies that 85 % of the observed drift in this population of instruments is due to pressure-dependent drift. It is important to note that this analysis is not conclusive; other drifts other than pressure or temperature might be at play, which might invalidate these results. However, these results point to the tantalizing possibility that

by about a factor of 2.7. For CO₂, the mean fractional change is 0.05 %, or 0.2 ppm; for the same instruments over the

improving the performance of the pressure sensor could lead to a much higher degree of stability in these instruments. It is to be noted that, from the dependences evaluated at the factory and the instrument data shown, the mean pressure drift rate on a yearly basis is $0.3 \text{ Torr year}^{-1}$, which is well within the expected drift rates for the sensing technology used in the instrumentation.

5 Conclusions

Between 2008 and fall of 2014, 47 non-isotopic Picarro instruments were tested at what is now the MLab. The goals of this work were to give insight into the MLab testing procedures that are also applied to the other instruments as well as to provide an evaluation of the tested instruments. We show that over time the instruments tend to have more reproducible performances. However, the first instruments of a new model tend to differ from one another than the last instruments of the previous model. This conclusion holds for CO even though its measurement is challenging. We also see that the results from the factory, at the MLab and in the field generally agree well with each other; in the case of the field, the performances stay relatively unchanged over time. The laboratory test can then be used to prioritize the location of the instruments according to their performances and the needs of the stations before installation. This also shows that, for instruments that could not be tested at the laboratory, the field estimate could be an acceptable proxy if measured on a long enough period of time using the same protocols as described here. We can conclude that the instruments tested are well designed for field study (with an average of more than 80 % of valid data over the instruments tested in the field in this study). The troubleshooting list provided in this study is representative only of the observed failures for the tested instruments. Within SNO-RAMCES, a troubleshooting log-book is developed to allow every station to consult and add failures and solutions. This could be extended to the whole ICOS network. We would also like to add a short list of recommendations for use of these analyzers in the field. These recommendations are most likely valid for other instruments as well.

- Instruments should be tested in the laboratory before being on site. Indeed, some important tests such as the temperature, pressure and water vapor dependence tests are only done at the laboratory. It is also convenient to be able to verify the performances of an instrument with a standardized and recognized protocol.
- Measurement interval duration should be at least 10 min to allow for stabilization and should be in any case tested for the specific setup of the station as we have shown that this seems to be mostly station-specific and not instrument-specific. Indeed, to be able to reach the WMO comparison goals, we need biases as small as possible for every source of bias. Here, we aim for a difference of less than 0.05 for CO₂, 0.2 for CH₄ and 1 ppb for CO.
- Pressure difference between the different samples should not exceed 0.4 bar to avoid significant inlet pressure influence, which leads to systematic biases.
- Calibration sequences should have at least two cycles, but in most cases this could be enough.
- Calibrations need to be run regularly to follow the instrument and setup drift. Especially after each restart of the instrument, calibrations have to be run. Each station setup being different, we cannot recommend a specific frequency for calibration, but we recommend that during the first 6 months these calibrations are run at least every 2 weeks. Then after analysis of the data, the frequency should be optimized.
- Despite these findings, we highly recommend carrying out a thorough test of the instrument at the station to take into account specificities that would lead to a needed higher number of calibration cycles or a longer interval time.

Appendix A

Table A1. Abbreviations.

AMS	Amsterdam Island, France, Indian Ocean
BIS	Biscarrosse, France
CMR	continuous measurement repeatability
IVI	Ivittuut, Greenland
LTO	Lamto, Côte d'Ivoire
LTR	long-term repeatability
MHD	Mace Head, Ireland
OPE	Observatoire Pérenne de l'Environnement, France
PUJ	Puijo, Finland
PUY	Puy de Dôme, France
STR	short-term repeatability
TRN	Trafnou, France

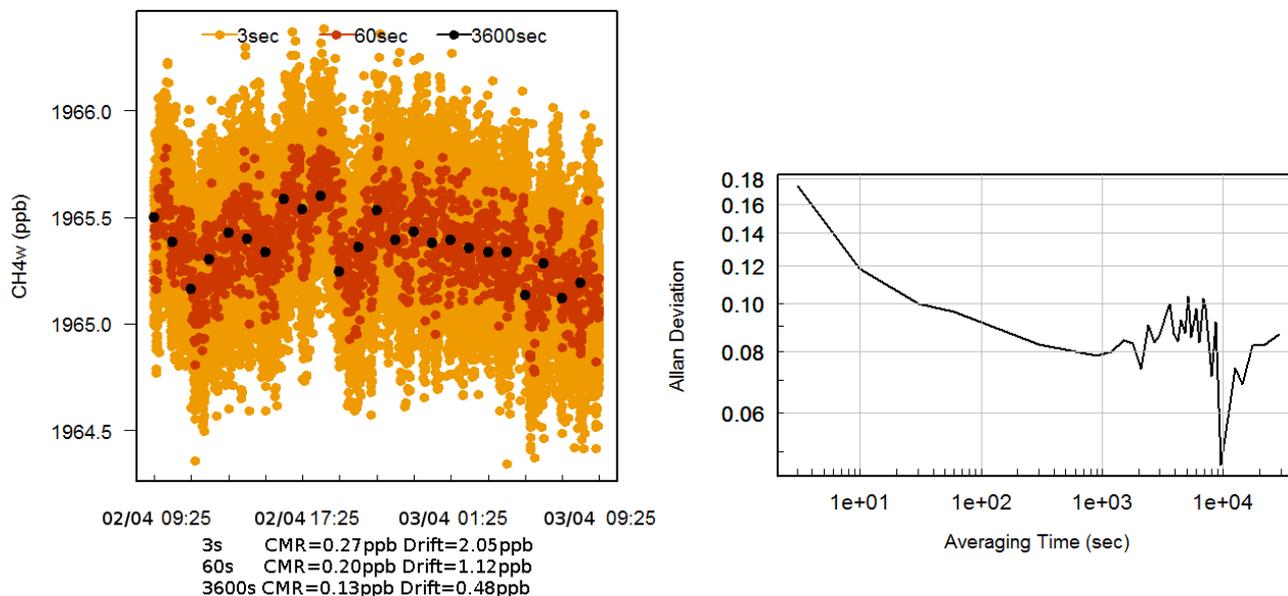


Figure A1. CH₄ continuous measurement repeatability. First panel: measurements averaged over different time intervals. Second panel: Allan deviation.

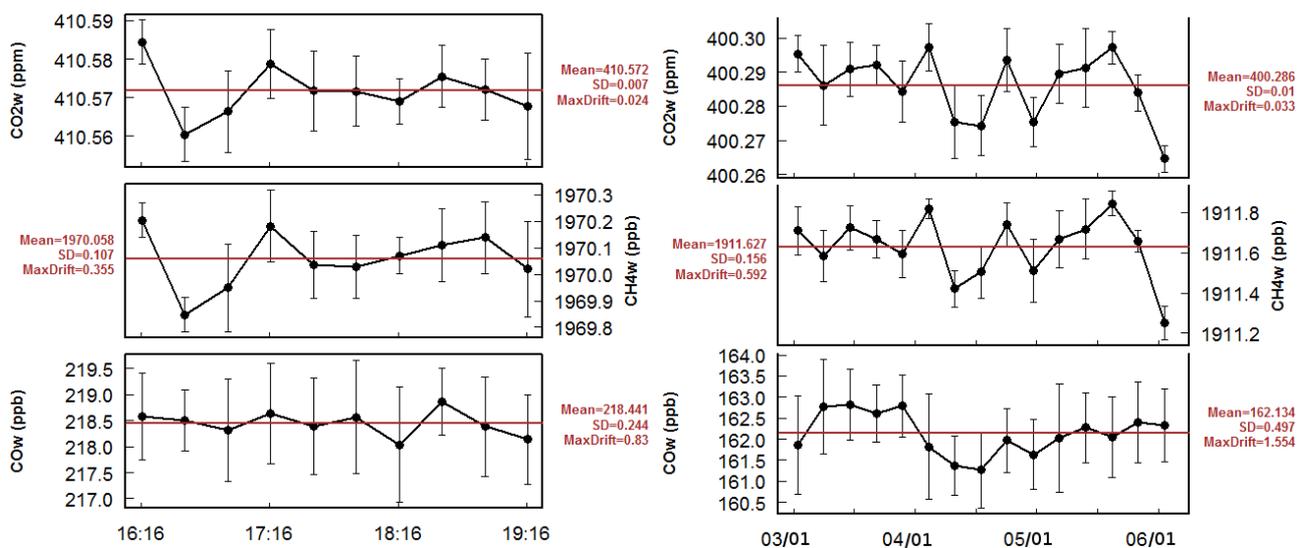


Figure A2. Short-term and long-term repeatability for the three species. First panel: short-term repeatability. Second panel: long-term repeatability.

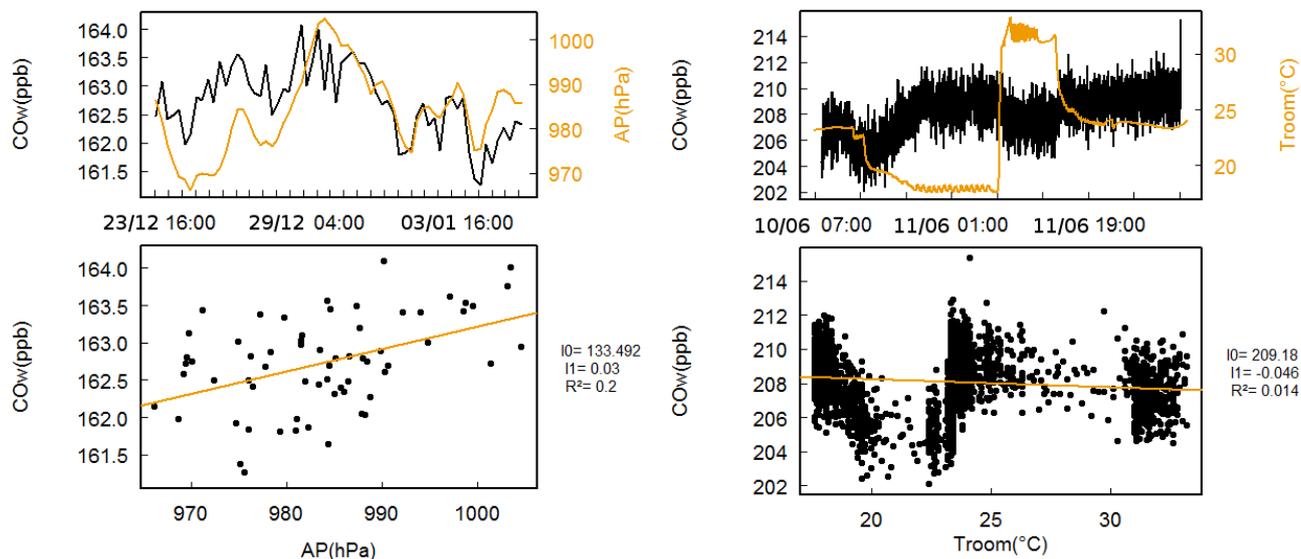


Figure A3. CO pressure and temperature dependency. First panel: pressure dependency. Second panel: temperature dependency. On the right of the lower plot, the slope (I_1), intercept (I_0) and the coefficient of correlation (R^2) are indicated.

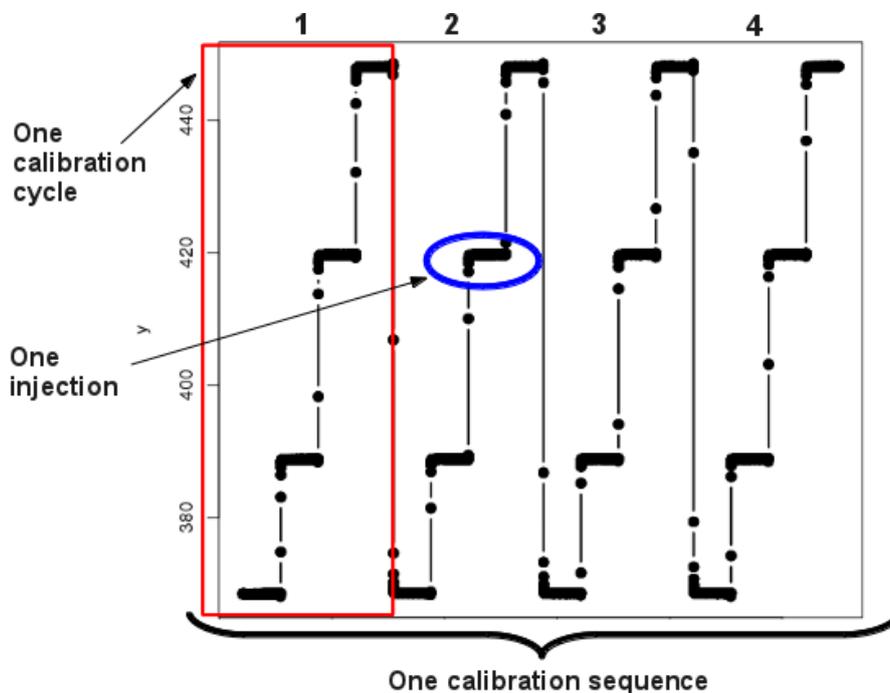


Figure A4. Schematics of the calibration procedure at the MLab. With a measurement interval time of 20–30 min, a full sequence lasts between 5 and 8 h.

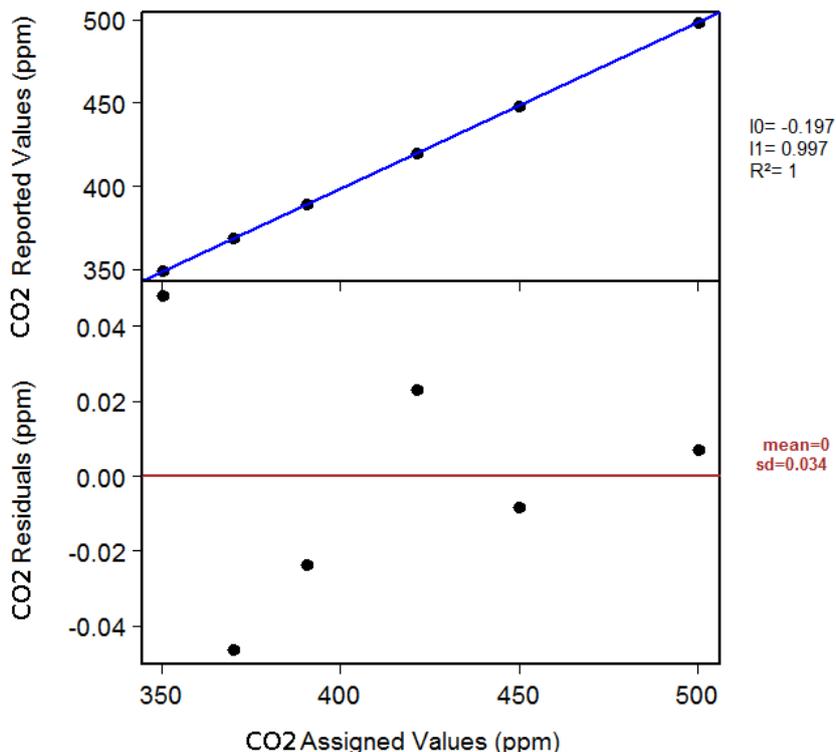


Figure A5. Linearity test for CO₂. On the top panel, the measured values vs. the assigned values. On the right of this panel, the slope (I1), intercept (I0) and the coefficient of correlation (R^2) are indicated. On the lower panel, the residuals, i.e., the difference between the assigned values and the values calculated using the linear equation, are shown.

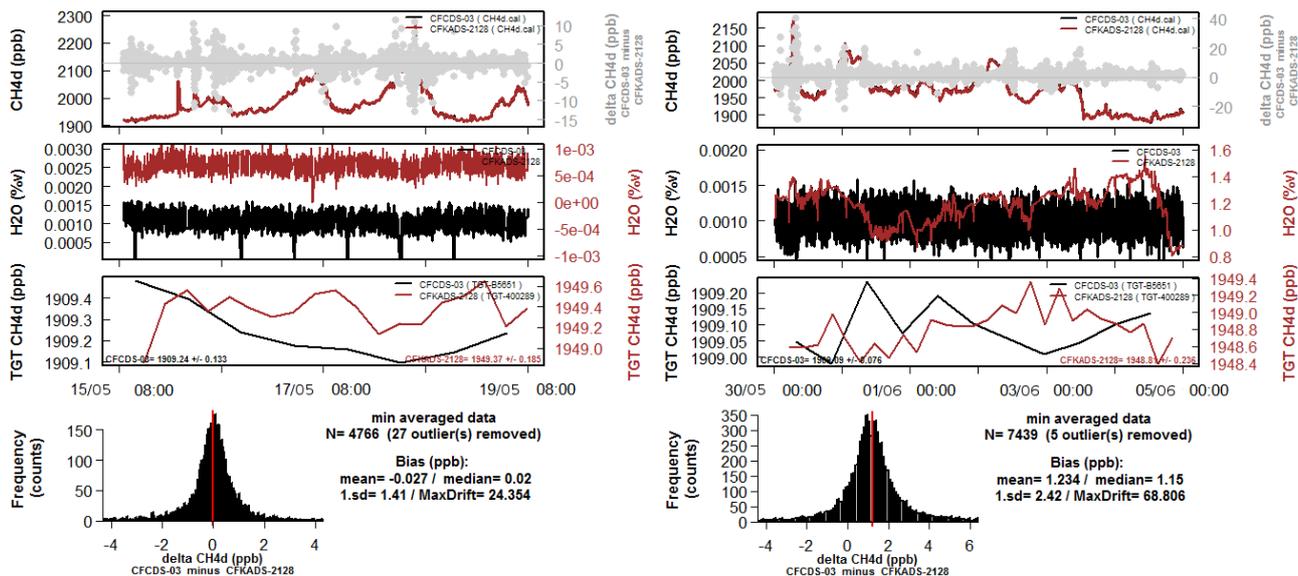


Figure A6. Comparison with the reference instrument for CH₄. First panel: dry air vs. dry air. Second panel: wet air corrected for H₂O vs. dry air. From top to bottom, the concentrations for both instruments and the difference of the two are plotted, then the water vapor concentrations for both instruments, then the evolution of the target for both instruments and finally a histogram of the distribution of the differences along with statistics.

Acknowledgements. We would like to thank the technicians in the field that carry out the regular maintenance and the station PIs who control the quality of the data. For AMS, we thank the French civil service volunteers, Olivier Jossoud, Delphine Combaz and Laurine Hégo. For BIS, we thank Yvan Rush, Patricia Salin and Christophe Martin from DGA. For IVI, we thank the GLK Danish navy and the Kommuneqarfik Sermersooq (Greenland). For LTO, we thank Palmer Yao, Ismael Coulibaly, Bi Danko Raphaël Zouzou and Adama Diawara from Abidjan-Cocody University. For MHD, we thank Gerry Spain, Mick Geever and Damien Martin from Galway University and Ireland's EPA and Simon O'Doherty from Bristol University. For OPE, we thank Sébastien Conil and Maxime Simon from ANDRA. For PUJ, we thank Tuomas Laurila, Juha Hatakka and Harri Portin from FMI. For PUY, we thank Jean-Marc Pichon, Mickaël Ribeiro, David Picard and Aurélie Colomb from LaMP. Finally, for TRN, we thank Eric Parmentier from IPGP.

We also would like to thank Lynn Hazan and Amara Abbaris, who develop and maintain the station database. This study was funded by the European Commission under the EU Seventh Research Framework Programme through ICOS (grant agreement no. 211574), and ICOS-INWIRE (grant agreement no. 313169).

Edited by: D. Griffith

References

- Andrews, A. E., Kofler, J. D., Trudeau, M. E., Williams, J. C., Neff, D. H., Masarie, K. A., Chao, D. Y., Kitzis, D. R., Novelli, P. C., Zhao, C. L., Dlugokencky, E. J., Lang, P. M., Crotwell, M. J., Fischer, M. L., Parker, M. J., Lee, J. T., Baumann, D. D., Desai, A. R., Stanier, C. O., De Wekker, S. F. J., Wolfe, D. E., Munger, J. W., and Tans, P. P.: CO₂, CO, and CH₄ measurements from tall towers in the NOAA Earth System Research Laboratory's Global Greenhouse Gas Reference Network: instrumentation, uncertainty analysis, and recommendations for future high-accuracy greenhouse gas monitoring efforts, *Atmos. Meas. Tech.*, 7, 647–687, doi:10.5194/amt-7-647-2014, 2014.
- Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E. J., Brunke, E.-G., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D. A., Krummel, P. B., Langenfelds, R. L., Ramonet, M., Schmidt, M., Steele, L. P., Szopa, S., Yver, C., Viovy, N., and Ciais, P.: Source attribution of the changes in atmospheric methane for 2006–2008, *Atmos. Chem. Phys.*, 11, 3689–3700, doi:10.5194/acp-11-3689-2011, 2011.
- Broquet, G., Chevallier, F., Bréon, F.-M., Kadyrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguá, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P.: Regional inversion of CO₂ ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, *Atmos. Chem. Phys.*, 13, 9039–9056, doi:10.5194/acp-13-9039-2013, 2013.
- Carouge, C., Rayner, P. J., Peylin, P., Bousquet, P., Chevallier, F., and Ciais, P.: What can we learn from European continuous atmospheric CO₂ measurements to quantify regional fluxes – Part 2: Sensitivity of flux accuracy to inverse setup, *Atmos. Chem. Phys.*, 10, 3119–3129, doi:10.5194/acp-10-3119-2010, 2010a.
- Carouge, C., Bousquet, P., Peylin, P., Rayner, P. J., and Ciais, P.: What can we learn from European continuous atmospheric CO₂ measurements to quantify regional fluxes – Part 1: Potential of the 2001 network, *Atmos. Chem. Phys.*, 10, 3107–3117, doi:10.5194/acp-10-3107-2010, 2010b.
- Chen, H., Winderlich, J., Gerbig, C., Hofer, 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 (CO₂ and CH₄) using the cavity ring-down spectroscopy (CRDS) technique, *Atmos. Meas. Tech.*, 3, 375–386, doi:10.5194/amt-3-375-2010, 2010.
- Chen, H., Karion, A., Rella, C. W., Winderlich, J., Gerbig, C., Filges, A., Newberger, T., Sweeney, C., and Tans, P. P.: Accurate measurements of carbon monoxide in humid air using the cavity ring-down spectroscopy (CRDS) technique, *Atmos. Meas. Tech.*, 6, 1031–1040, doi:10.5194/amt-6-1031-2013, 2013.
- Crosson, E. R.: A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon dioxide, and water vapor, *Appl. Phys. B*, 92, 403–408, doi:10.1007/s00340-008-3135-y, 2008.
- 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 CH₄ mole fractions to a gravimetrically prepared standard scale, *J. Geophys. Res.-Atmos.*, 110, D18306, doi:10.1029/2005JD006035, 2005.
- Hazan, L., Tarniewicz, J., Ramonet, M., Laurent, O., Abbaris, A., and Rivier, L.: Automatic near real time CO₂ and CH₄ mole fraction data processing, in preparation, 2015.
- Karion, A., Sweeney, C., Wolter, S., Newberger, T., Chen, H., Andrews, A., Kofler, J., Neff, D., and Tans, P.: Long-term greenhouse gas measurements from aircraft, *Atmos. Meas. Tech.*, 6, 511–526, 2013.
- Novelli, P. C., Collins, J. E., Myers, R. C., Sachse, G. W., and Scheel, H. E.: Reevaluation of the NOAA/CMDL carbon monoxide reference scale and comparisons with CO reference gases at NASA-Langley and the Fraunhofer Institut, *J. Geophys. Res.-Atmos.*, 99, 12833–12839, 1994.
- Peltola, O., Hensen, A., Helfter, C., Beletti 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 covariance flux measurements: the InGOS inter-comparison field experiment, *Biogeosciences*, 11, 3163–3186, doi:10.5194/bg-11-3163-2014, 2014.
- Ramonet, M., Cais, P., Aalto, T., Aulagnier, C., Chevallier, F., Cipriano, D., Conway, T. J., Haszpra, L., Kazan, V., Meinhardt, F., Paris, J.-D., Schmidt, M., Simmonds, P., Xueref-Remy, I., and Necki, J. N.: A recent build-up of atmospheric CO₂ over Europe, Part 1: observed signals and possible explanations, *Tellus B*, 62, 1–13, 2009.
- 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.

- Richardson, S. J., Miles, N. L., Davis, K. J., Crosson, E. R., Rella, C. W., and Andrews, A. E.: Field testing of cavity ring-down spectroscopy analyzers measuring carbon dioxide and water vapor, *J. Atmos. Oc. Technol.*, 29, 397–406, 2012.
- Schmidt, M., Lopez, M., Yver Kwok, C., Messenger, C., Ramonet, M., Wastine, B., Vuillemin, C., Truong, F., Gal, B., Parmentier, E., Cloué, O., and Ciais, P.: High-precision quasi-continuous atmospheric greenhouse gas measurements at Trainou tower (Orléans forest, France), *Atmos. Meas. Tech.*, 7, 2283–2296, doi:10.5194/amt-7-2283-2014, 2014.
- Xueref-Remy, I., Bousquet, P., Carouge, C., Rivier, L., and Ciais, P.: Variability and budget of CO₂ in Europe: analysis of the CAATER airborne campaigns – Part 2: Comparison of CO₂ vertical variability and fluxes between observations and a modeling framework, *Atmos. Chem. Phys.*, 11, 5673–5684, doi:10.5194/acp-11-5673-2011, 2011a.
- Xueref-Remy, I., Messenger, C., Filippi, D., Pastel, M., Nedelec, P., Ramonet, M., Paris, J. D., and Ciais, P.: Variability and budget of CO₂ in Europe: analysis of the CAATER airborne campaigns – Part 1: Observed variability, *Atmos. Chem. Phys.*, 11, 5655–5672, doi:10.5194/acp-11-5655-2011, 2011b.
- Yver, C., Schmidt, M., Bousquet, P., Zahorowski, W., and Ramonet, M.: Estimation of the molecular hydrogen soil uptake and traffic emissions at a suburban site near Paris through hydrogen, carbon monoxide, and radon-222 semicontinuous measurements, *J. Geophys. Res.*, 114, D18304, doi:10.1029/2009JD012122, 2009.
- Zhao, C. L. and Tans, P. P.: Estimating uncertainty of the WMO mole fraction scale for carbon dioxide in air, *J. Geophys. Res.-Atmos.*, 111, D08S09, doi:10.1029/2005JD006003, 2006.