



# Comparison of nitrous oxide (N<sub>2</sub>O) analyzers for high-precision measurements of atmospheric mole fractions

Benjamin Lebegue<sup>1</sup>, Martina Schmidt<sup>2</sup>, Michel Ramonet<sup>1</sup>, Benoit Wastine<sup>1</sup>, Camille Yver Kwok<sup>1</sup>, Olivier Laurent<sup>1</sup>, Sauveur Belviso<sup>1</sup>, Ali Guemri<sup>1</sup>, Carole Philippon<sup>1</sup>, Jeremiah Smith<sup>3</sup>, and Sebastien Conil<sup>4</sup>

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France

<sup>2</sup>Institut für Umweltp Physik, University of Heidelberg, Heidelberg, Germany

<sup>3</sup>Thünen Institut für Agrarklimaschutz, Braunschweig, Germany

<sup>4</sup>Observatoire Pérenne de l'Environnement, ANDRA, Bure, France

Correspondence to: Benjamin Lebegue (benjamin.lebegue@lsce.ipsl.fr)

Received: 9 September 2015 – Published in Atmos. Meas. Tech. Discuss.: 26 October 2015

Revised: 16 February 2016 – Accepted: 7 March 2016 – Published: 23 March 2016

**Abstract.** Over the last few decades, in situ measurements of atmospheric N<sub>2</sub>O mole fractions have been performed using gas chromatographs (GCs) equipped with electron capture detectors. This technique, however, becomes very challenging when trying to detect the small variations of N<sub>2</sub>O as the detectors are highly nonlinear and the GCs at remote stations require a considerable amount of maintenance by qualified technicians to maintain good short-term and long-term repeatability. With new robust optical spectrometers now available for N<sub>2</sub>O measurements, we aim to identify a robust and stable analyzer that can be integrated into atmospheric monitoring networks, such as the Integrated Carbon Observation System (ICOS). In this study, we present the most complete comparison of N<sub>2</sub>O analyzers, with seven analyzers that were developed and commercialized by five different companies. Each instrument was characterized during a time period of approximately 8 weeks. The test protocols included the characterization of the short-term and long-term repeatability, drift, temperature dependence, linearity and sensitivity to water vapor. During the test period, ambient air measurements were compared under field conditions at the Gif-sur-Yvette station. All of the analyzers showed a standard deviation better than 0.1 ppb for the 10 min averages. Some analyzers would benefit from improvements in temperature stability to reduce the instrument drift, which could then help in reducing the frequency of calibrations. For most instruments, the water vapor correction algorithms applied by companies are not sufficient for high-precision atmospheric measurements,

which results in the need to dry the ambient air prior to analysis.

## 1 Introduction

Nitrous oxide (N<sub>2</sub>O) is a greenhouse gas with an atmospheric lifetime of 131 years and a global warming potential that is approximately 300 times that of CO<sub>2</sub> at a 100-year time horizon (Prather et al., 2012). At present, the N<sub>2</sub>O emissions are the most important factor for stratospheric ozone depletion, and they are expected to remain the largest factor for this century (Ravishankara et al., 2009; Wuebbles, 2009).

Global observations of the atmospheric N<sub>2</sub>O mole fraction from networks such as Advanced Global Atmospheric Gas Experiment (AGAGE), National Oceanic and Atmospheric Administration's Earth System Research Laboratory (NOAA/ESRL), Commonwealth Scientific and Industrial Research Organisation's Global Atmospheric Sampling Laboratory (CSIRO GASLAB) and Réseau Atmosphérique de Mesure des Composés à Effet de Serre (RAMCES) showed a mean value of approximately 328 ppb for the Northern Hemisphere and 326.5 ppb for the Southern Hemisphere in 2014 (Lopez et al., 2012; Schmidt et al., 2014; Thompson et al., 2013). The amplitude of the seasonal cycle is smaller in the Southern than in the Northern Hemisphere, with a value of approximately 0.4 ppb at Cape Grim, Tasmania, compared with 0.89 ppb at Mace Head, Ireland (Nevison et al., 2007).

The N<sub>2</sub>O growth rate in the atmosphere over the last 5 years is, on average, 0.75–0.78 ppb per year. Overall, gradients over the continent or between maritime and continental air are small (less than 0.6 ppb) and need to be measured precisely.

Even closer to the sources, at European semi-rural stations, only slight variations in atmospheric N<sub>2</sub>O were found on the timescale of days. Lopez et al. (2012) showed that the mean diurnal cycle at the semi-urban station Gif-sur-Yvette (France, measurements 7 m a.g.l.) has an amplitude of 0.96 ppb, whereas at Traînou tall tower (rural area, measurements up to 180 m a.g.l.) the mean amplitude is only 0.32 ppb. At the 17th WMO/IAEA meeting, 10–13 June 2013 in Beijing, an expert group on CO<sub>2</sub> and other greenhouse gases from the World Meteorological Organization Global Atmosphere Watch (WMO/GAW) recommended an N<sub>2</sub>O inter-laboratory comparability goal of ±0.1 ppb ([http://www.wmo.int/pages/prog/arep/gaw/documents/Final\\_GAW\\_213\\_web.pdf](http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_213_web.pdf)). This ambitious goal has not yet been reached, as shown recently by Bergamaschi et al. (2015), who found biases between in situ gas chromatography (GC) measurements and flask sampling at different European stations of up to 0.7 ppb. These biases, even when corrected, limit the precision of N<sub>2</sub>O emission estimates by inverse models.

High-precision atmospheric N<sub>2</sub>O measurements in flask measurement networks and at in situ stations are traditionally measured by GC using an electron capture detector. Methods incorporating this technique have achieved a typical short-term continuous measurement repeatability (CMR) of 0.1 to 0.3 ppb (Lopez et al., 2012; Nevison et al., 2011; Popa et al., 2010; Schmidt et al., 2001). Over the last few years, new analytical techniques have become commercially available for obtaining high-precision measurements of atmospheric N<sub>2</sub>O. Hammer et al. (2013) described the Fourier transform infrared (FTIR) absorption, which can reach a long-term repeatability (LTR) for N<sub>2</sub>O of 0.04 ppb over a 10-month period. More recently, laser-based systems, e.g., cavity-enhanced off-axis integrated cavity output spectroscopy (OA-ICOS), cavity ring-down spectroscopy (CRDS), quantum cascade tunable infrared laser differential absorption spectroscopy (QC-TILDAS) and difference frequency generation (DFG)-based systems, were developed and commercialized by different companies. In this study, we present the first assessment of the performance of seven N<sub>2</sub>O analyzers and compare these techniques to routine instruments, including a GC analyzer (Lopez et al., 2012) that is used at LSCE (Laboratoire des Sciences du Climat et de l'Environnement) for ambient air measurements since 2001 and a FTIR that has been running since 2012. All of the tested instruments have been characterized for repeatability, long-term stability, linearity, temperature dependency and spectroscopic cross-interferences with water vapor. The instrument evaluations were all performed at the ICOS Atmospheric Thematic Center Metrology Laboratory (ATC MLab) hosted

at LSCE in Gif-sur-Yvette, but not all of the instruments were tested at the same time. Because the ATC MLab was in its creation phase, there has been an evolution toward the best practices in the test protocol, which will be detailed below. The evaluations have been performed between October 2012 and January 2014 over three periods: November–December 2012, May–June 2013 and December 2013–January 2014. While the tests have been carried out in the frame of ICOS, the results are valid for all groups and networks doing high-precision atmospheric N<sub>2</sub>O measurements.

## 2 Instrument descriptions

### 2.1 Gas chromatograph instrument, Agilent

The LSCE laboratory at Gif-sur-Yvette is equipped with an automated GC system (HP-6890, Agilent, coupled to a PP1, Peak Performer Laboratories, up to May 2013; since then, HP-7890A was used with the same PP1 and similar specificities and performances) to analyze the CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, SF<sub>6</sub>, CO and H<sub>2</sub> mole fractions. This instrument has been contributing to the RAMCES monitoring network since 2001 and is used to obtain in situ measurements at Gif-sur-Yvette station for the analysis of flask samples and the calibration of working standards. It also serves as a reference instrument for international comparison programs, such as the WMO Round Robin or the European Cucumbers intercomparison program (<http://cucumbers.uea.ac.uk/>). Detailed descriptions of the GC system for N<sub>2</sub>O analysis is given by Lopez et al. (2012). The GC is equipped with a flame ionization detector and a nickel catalyst to determine the CH<sub>4</sub> and CO<sub>2</sub> and with an electron capture detector (ECD) for N<sub>2</sub>O and SF<sub>6</sub> analysis. We use a 10 mL sample loop, two Hayesep-Q columns and Ar/CH<sub>4</sub> as a carrier gas to separate N<sub>2</sub>O and SF<sub>6</sub> from the other compounds of air. Each analysis takes less than 6 min and allows two to six measurements of air samples on an hourly basis. For the small range of N<sub>2</sub>O mole fractions in ambient air (324–334 ppb), the ECD can be corrected for nonlinearity, applying a two-point calibration strategy with two working standards (322 and 338 ppb). A calibration frequency of 30 to 45 min was chosen to reach a N<sub>2</sub>O repeatability of 0.2 ppb. Lopez et al. (2012) described an interference between SF<sub>6</sub> and N<sub>2</sub>O measurements when SF<sub>6</sub> mole fraction exceeds 15 ppt. Therefore, during the comparison period, all of the N<sub>2</sub>O measurements made by our routine GC are flagged as not valid when SF<sub>6</sub> exceeds 15 ppt. The standard deviation of the quality control gas (i.e., the “target” gas), which was injected every 2 h, was 0.3 ppb over the comparison period (October 2012 to January 2014). The GC is used in this study as the routine reference instrument to compare the instruments' performance regarding CMR and drift assessments and for ambient air comparisons.

## 2.2 Fourier transform infrared spectrometer, Ecotech

FTIR spectroscopy is based on the absorption of infrared radiation (Beer–Lambert law). A polychromatic infrared beam from an infrared source first passes through a Michelson interferometer, and this modulated beam then traverses the sample cell. The resulting time-modulated signal is converted into an infrared spectrum through Fourier transformation. One of the main advantages of an FTIR analyzer is its ability to record a spectrum over a broad IR range (1800 to 7500  $\text{cm}^{-1}$ ), thereby offering the possibility to measure a large number of species simultaneously. Spectra are stored and can be analyzed at a later date with a different method to obtain better data or study new species.

Here, we briefly describe the instrument configuration used during the comparison. The LSCE purchased this analyzer (built by Ecotech, Australia) in 2011 from the University of Wollongong (Australia). Detailed descriptions of similar analyzers used in the atmospheric community are presented by Griffith et al. (2012) and Hammer et al. (2013). The instrument used in our laboratory consists of a commercially available FTIR interferometer (IRcube, Bruker Optics, Germany) with a  $1\text{ cm}^{-1}$  resolution coupled to a 3.5 L multi-pass glass cell with a 24 m optical path length (PA-24, Infrared Analysis, USA). The cell and the interferometer are assembled on an optical bench inside a temperature-controlled chamber. An in situ J-type thermocouple, to monitor the cell temperature, and a pressure sensor (HPM-760s, Teledyne Hastings, USA) are included in the multi-pass cell. We used high-purity nitrogen (at least 99.995 vol %) to slowly purge the interferometer housing and the transfer optics between the cell and the interferometer. A drying system composed of a 24 in. counter-flow Nafion dryer (Permapure, Toms River, USA) followed by a chemical dryer ( $\text{Mg}(\text{ClO}_4)_2$ ) is installed upstream of the cell. The flow is provided by a pump (MV2NT, Vacuubrand, Germany). The pressure of the cell is controlled using a built-in mass flow controller mounted at the outlet of the cell, and the flow is controlled by a mass flow controller installed upstream of the cell and downstream of the drying system. Our instrument uses the absorption bands located between 2097 and 2242  $\text{cm}^{-1}$  to provide the mole fraction of  $\text{N}_2\text{O}$  and CO. A calibration with five working standards is carried out every 2 weeks, and a quality control gas is analyzed every 5 h for 40 min. Due to its size, it takes 5 to 10 min to empty and flush the cell when changing the type of sample. The sample flow rate was regulated at  $1 \pm 0.05\text{ L min}^{-1}$  and the cell pressure and temperature were regulated at  $1100 \pm 0.02\text{ hPa}$  ( $1.1 \times 10^5 \pm 2\text{ Pa}$ ) and  $32 \pm 0.03\text{ }^\circ\text{C}$ , respectively. The sample measurement intervals for the target and air measurements are 1 and 3 min, respectively. The quality control gas showed a  $\text{N}_2\text{O}$  standard deviation of 0.08 ppb for 1 min measurements during the first and third periods of the comparison (December 2012 and January 2014) and a standard deviation

of 0.12 ppb for 1 min measurements during the second period (May 2013).

## 2.3 Cavity ring-down spectrometer/quantum cascade laser (CRDS-QCL) instruments G5101-I, Picarro

The two CRDS instruments tested in this study were a loaned prototype, tested from November to December 2012, and the commercialized model of the G5101-i unit from Picarro (Picarro Inc., CA, USA) bought by the LSCE in May 2014. The thermal regulation was not yet fully optimized for the prototype, and it did not possess a water vapor correction. The CRDS system made by Picarro uses a QCL as a source to measure  $\text{N}_2\text{O}$ ,  $\delta^{15}\text{N}_\alpha$ ,  $\delta^{15}\text{N}_\beta$  and  $\text{H}_2\text{O}$  in the mid-infrared region (4.55  $\mu\text{m}$ ). The CRDS technique uses an optical cell (48 mL) with three highly reflective mirrors (Crosson, 2008). Light is injected into the cavity at the required wavelength from the QCL through a highly reflective mirror and is measured by a photodetector through a second highly reflective mirror. The path length inside the cell is about 8 km. The intensity of the light inside the cell builds up over time by resonance due to a third mirror mounted on a piezoelectric device, which allows for inter-mirror distance adjustment. Then, the laser is switched off, and the time constant of the light intensity decrease is measured. From this cavity decay time, the concentration is retrieved when knowing the absorption cross section of the species at the laser wavelength. A measurement interval of less than 10 s is obtained. During the comparison campaign at LSCE, the sample flow rate is approximately  $50\text{ mL min}^{-1}$ . The cell pressure is regulated at  $100 \pm 0.001\text{ Torr}$  ( $1.33 \times 10^4 \pm 0.13\text{ Pa}$ ), and the cell temperature is set at  $40 \pm 0.001\text{ }^\circ\text{C}$ . A calibration with four working standards is performed every 10 days, and a quality control gas is analyzed every 5 h for 30 min.

## 2.4 IRIS 4600, Thermo Fisher Scientific (DFG)

The Thermo IRIS 4600 was lent by Thermo Fisher Scientific for our test campaign from November to December 2012. The instrument measures  $\text{N}_2\text{O}$  and water vapor. It uses DFG laser technology, which consists of combining two near-infrared telecom lasers into a nonlinear frequency conversion crystal to reach the mid-infrared region. The laser continuously sweeps the absorption bandwidth at a rate of 500 Hz (Scherer et al., 2013). This spectrometer measures in the 4.6  $\mu\text{m}$   $\text{N}_2\text{O}$  and  $\text{H}_2\text{O}$  bands, and each sweep provides a near instantaneous measurement of the two gases. The measurement interval is adjustable between 0.1 and 10 s. The cell is 40 cm long for a cell size of approximately 80 mL, which provides an optical path length of 5 m. The sample flow rate through the cell is  $300\text{ mL min}^{-1}$ . The cell pressure and temperature are regulated at  $175 \pm 0.002\text{ mbar}$  ( $1.75 \times 10^4 \pm 0.2\text{ Pa}$ ) and  $37.5 \pm 0.002\text{ }^\circ\text{C}$ , respectively. A calibration with four standard gases is performed every week, and a quality control gas is measured every 5 h for 30 min.

## 2.5 OA-ICOS-QCL instruments, Los Gatos Research (OA-ICOS)

Three OA-ICOS-QCLs made by Los Gatos Research (LGR, USA) were provided to us by the French National Radioactive Waste Management Agency (ANDRA) for a performance assessment. These three instruments measure N<sub>2</sub>O, CO and H<sub>2</sub>O in the 4.6 μm region. Instruments based on OA-ICOS use a tunable laser and an optical cavity (408 mL). However, unlike other cavity methods such as CRDS, this technique is based on directing the laser beam off axis through the cavity, which, when combined with highly reflective mirrors, provides a long effective optical path that spatially sweeps the cavity volume due to spatially separated multi-reflections within the cavity before the reentrant condition of the optical beam is fulfilled. Moreover, in contrast to CRDS, the laser beam is not locked on each cavity mode but is swept over the gas absorption line. The mole fraction can then be obtained from the measured spectra integrated over the entire absorption feature, the cell pressure and temperature, the effective optical path length and the absorption line parameters for each species. These instruments have a measurement rate of up to 1 Hz with the internal pump and up to 10 Hz with an optional pump.

One instrument is the standard model of the analyzer, referred to here as ICOS-SD. The two others are enhanced performance models that incorporate an improved temperature control of the cavity, referred to here as ICOS-EP38 and ICOS-EP40. The sample flow rate is set for the three analyzers at 300 mL min<sup>-1</sup>. The cell is regulated at 85 ± 0.007 Torr (1.13 × 10<sup>4</sup> ± 0.93 Pa) and 27 ± 0.2 °C for the standard model and 45 ± 0.005 °C for the enhanced models. For the ICOS-SD, a calibration is conducted once a week. For the two enhanced models, the calibration frequency occurs every 2 weeks. For all instruments, a quality control gas is analyzed every 5 h for 30 min for the ICOS-SD and every 6 h for 30 min for the two ICOS-EP models.

## 2.6 QCL Mini Monitor, Aerodyne (QC-TILDAS)

The tested Aerodyne's QCL Mini Monitor (Aerodyne Research Inc., USA) was provided by the Thünen Institut für Agrarklimaschutz (Braunschweig, Germany) for our test campaign. This instrument is currently used for eddy covariance measurements.

The instrument uses a quantum cascade tunable infrared laser differential absorption spectroscopy technique. One QC laser beam is sent through a Herriott astigmatic multi-pass cavity (0.5 L) with a fixed optical path length of 76 m. Then, the light is received by a thermoelectrically cooled infrared detector. This instrument works in the mid-infrared domain (4.54 μm). The instrument performs an advanced type of wavelength sweep integration before dropping the laser current below a threshold to determine the voltage of the detector for zero light. The instrument can then determine con-

centrations by fitting the measured spectrum with the HITRAN database using the cell temperature and pressure. The instrument can run periodically an auto-background, which consists of acquiring a spectrum with the sample cell filled with dry nitrogen and is used to normalize future spectra. This option was not used during this campaign. The instrument tested had no active control of the cell pressure. During the tests in our laboratory, we set the sample flow rate at 1 L min<sup>-1</sup> and the pressure at 33 ± 0.2 Torr (4.4 × 10<sup>3</sup> ± 27 Pa) by using a valve at the outlet of the cavity and a needle valve at the inlet. The cell's temperature is regulated at 22 ± 0.03 °C. A calibration is conducted every 2 weeks, and a quality control gas is measured every 5 h.

## 3 Instrument tests

### 3.1 Laboratory description

All tests were performed at the ICOS ATC MLab located at Gif-sur-Yvette, 20 km southwest of Paris. The MLab purpose is to test and validate atmospheric analyzers, instrumental setups and related components and consumables. The laboratory is air conditioned and equipped with calibration cylinders, target gases and inlet lines coupled with drying systems for ambient air comparisons. The water dependency, which was first tested using the droplet test, is now evaluated with a dedicated humidifier bench. The ambient air inlet is located on the roof of the laboratory, 7 m above ground level (a.g.l.).

For ambient air comparisons, the air is dehumidified, by passing through 335 mL glass traps cooled in an ethanol bath using a cryogenic cooler (Thermo Neslab CC-65 or HAAKE EK 90). The cooling traps are filled with glass beads to increase the surface area for water vapor condensation. Depending on the weather conditions, the cooling traps are typically changed once or twice per week. This setup dries the air sample down to less than 15 ppm of water.

In total, we use four sets of calibration cylinders (laboratory standards) to calibrate all of the N<sub>2</sub>O analyzers presented in this study. For the GC measurements, we use two calibration cylinders (Luxfer aluminum cylinders) filled by Deuste Steininger (Mühlhausen, Germany) in a synthetic matrix of N<sub>2</sub>, O<sub>2</sub> and Ar. These cylinders have been calibrated against laboratory primary standards purchased from NOAA/CMDL and are reported for N<sub>2</sub>O on the NOAA-2006a scale. For all optical instruments, three other sets of calibration cylinders are used during the comparison with a certified concentration of N<sub>2</sub>O (Table 1) among others gases such as CO, CH<sub>4</sub> or CO<sub>2</sub>. The first calibration set consists of five aluminum cylinders (Luxfer), which were filled with ambient air, spiked and calibrated with a GC-ECD by the Max Planck Institute of Jena, Germany, on the NOAA-2006a scale (Hall et al., 2007) spanning a range from 320 to 345 ppb N<sub>2</sub>O. Because this calibration set has been routinely used by the FTIR and to test other analyzers in the MLab, these cylin-

**Table 1.** The time period during which the tests were performed for all instruments and working specificities. Most instruments used two sets of calibration cylinders; here, we indicate the most frequently used: the set filled and calibrated by Max Planck Institute (MPI), spanning a range from 320 to 360 ppb of N<sub>2</sub>O, or the set filled by Deuste Steiningering and calibrated at LSCE (DS), ranging from 320 to 345 ppb N<sub>2</sub>O.

Instrument	Test period	Cell size (mL)	Cell flow rate (mL min <sup>-1</sup> )	Cell temperature (°C)	Cell pressure (Pa)	Main calibration set
FTIR	Oct 2012–Jan 2014	3500	1000 ± 50	32 ± 0.03	1.1 × 10 <sup>5</sup> ± 2	MPI
CRDS	Nov 2012–Dec 2012	48	< 50	40 ± 0.001	1.33 × 10 <sup>4</sup> ± 0.13	MPI
DFG	Nov 2012–Dec 2012	80	300	37.5 ± 0.002	1.75 × 10 <sup>4</sup> ± 0.2	MPI
ICOS-SD	Nov 2012–Dec 2012	408	300	27 ± 0.2	1.13 × 10 <sup>4</sup> ± 0.93	MPI
ICOS-EP38	May 2013–Jun 2013	408	300	45 ± 0.005	1.13 × 10 <sup>4</sup> ± 0.93	DS
ICOS-EP40	May 2013–Jun 2013	408	300	45 ± 0.005	1.13 × 10 <sup>4</sup> ± 0.93	DS
QC-TILDAS	Dec 2013–Jan 2014	500	1000	22 ± 0.03	4.4 × 10 <sup>3</sup> ± 667	MPI

ders were not always available during the second test period and had to be replaced by another set of calibration cylinders. This second set consists of six aluminum cylinders filled with a synthetic matrix of 21.0 ± 1 vol % of O<sub>2</sub>, 0.93 ± 1 vol % of Ar and a balance of N<sub>2</sub> (Deuste Steiningering) and calibrated by our FTIR. This set spans a range from 320 to 360 ppb N<sub>2</sub>O. When the QC-TILDAS was tested, a third calibration set was performed once by the instrument. This set consists of four aluminum cylinders, spanning a range from 335 to 355 ppm N<sub>2</sub>O, filled with a synthetic matrix of 21.0 ± 1 vol % of O<sub>2</sub>, 0.93 ± 1 vol % of Ar and a balance of N<sub>2</sub> (Deuste Steiningering) and calibrated by our FTIR. These calibration sets were analyzed at least every 2 weeks on each analyzer. Calibration sequences were made by measuring each cylinder for at least 15 min, including 10 min for flushing the inlet line and instrument cell. The whole calibration set was analyzed at least three times, with the first run systematically rejected to ensure a proper flushing of the system. The target cylinders used for various tests are filled with dry ambient air, using an oil-free compressor (RIX) and coalescent filters associated with magnesium perchlorate drying cartridges. These cylinders are analyzed prior to and after use by the GC system. All calibration and test cylinders are equipped with the same type of two-stage nickel-plated brass pressure regulators (Model 14, Scott Speciality Gases, Breda, the Netherlands).

All of the tests were performed in a temperature-controlled room (22 ± 1 °C), except for the temperature-dependence tests, for which the laboratory temperature was deliberately modified. In general, all analyzers were tested using similar procedures and during the same time span. We would have preferred to test all of the instruments at the same time, but due to constraints in their availability, tests were spread over time from October 2012 to January 2014. During this time span, the test procedures were improved towards the ICOS standard test protocol, which is now used to test new instruments in the MLab for the European ICOS atmospheric measurement network. In some cases, the time period was just too short to perform all tests. In the paragraphs below, we describe the test procedures in detail and state when an indi-

vidual test or analysis diverged from the standard protocol. Most of the tests presented in the following sections have been already described by Yver Kwok et al. (2015).

### 3.2 Continuous measurement repeatability and drift assessment

To determine the CMR (often called precision by the manufacturer) of the instruments, a single target gas tank filled at the MLab with dry natural air is measured continuously over a time period of at least 30 h. The raw data measurement interval of the analyzers tested varied from 5 min for the GC to 1 s for the QC-TILDAS analyzer (Table 2). We calculate the standard deviations of 30 h of continuous measurements of a target gas. Table 2 presents these values for the raw data at the frequency given by all instruments for 1 min averaged data (when available), 10 min averaged data and 1 h averaged data. During this 30 h sequence of target gas measurement, no calibrations were performed and no drift correction was applied, except for the GC, which automatically corrects the data with calibration cylinders every 45 min.

From this experiment, we also calculate the drift for each instrument (Table 2). The drift is calculated using a linear regression with the data from the 30 h test. The slope of the regression represents the drift of the instrument (in ppb h<sup>-1</sup>, which has been converted to ppb d<sup>-1</sup>) over this 30 h period.

For high-frequency measurements (1 to 2 s), the ICOS-EP and QC-TILDAS analyzers show the best CMR (0.08 to 0.10 ppb). For 1 min averaged data, the QC-TILDAS, CRDS and ICOS-EP models show very similar standard deviations of approximately 0.05–0.07 ppb, whereas the standard deviations of the ICOS-SD, DFG and FTIR are between 0.12 and 0.16 ppb. For the 1 h averaging data, the GC, FTIR and CRDS are the most precise, with the two ICOS-EP models and QC-TILDAS 2 times less precise and the ICOS-SD and DFG 5 times less precise. Apart from the GC, whose drift is corrected with the working standards, the FTIR has the smallest drift, with the CRDS and QC-TILDAS drifts being slightly higher. The other instruments present a significant drift of 0.1 ppb day<sup>-1</sup> or more. For these instruments, a stan-

**Table 2.** Continuous measurement repeatability calculated as 1 standard deviation over more than 30 h for different averaging times (raw data, 1 min, 10 min and 1 h averaging). The short-term drift is estimated with a linear regression over 30 h.

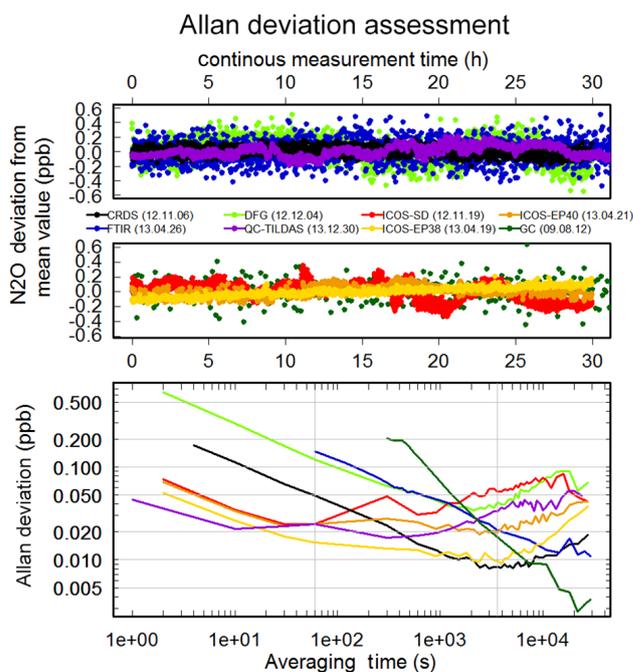
Instrument	$1\sigma$ (ppb, raw)	$1\sigma$ (ppb, 1 min)	$1\sigma$ (ppb, 10 min)	$1\sigma$ (ppb, 1 h)	Drift (ppb day <sup>-1</sup> )
GC	0.16 (5 min)	–	0.113	0.016	–
FTIR	0.15 (1 min)	0.149	0.055	0.026	0.017
CRDS	0.17 (4 s)	0.055	0.026	0.023	–0.034
DFG	0.66 (2 s)	0.159	0.107	0.097	–0.108
ICOS-SD	0.14 (2 s)	0.124	0.114	0.106	–0.185
ICOS-EP38	0.08 (2 s)	0.064	0.062	0.061	0.151
ICOS-EP40	0.10 (2 s)	0.068	0.060	0.054	0.070
QC-TILDAS	0.09 (1 s)	0.075	0.070	0.066	0.046

**Table 3.** Short-term repeatability assessment. A target tank is measured 10 times for 15 to 20 min (30 min for the FTIR), alternating with ambient air (for 5 min). An N<sub>2</sub>O mean value is calculated by taking the last 5 min of each analysis. The repeatability is expressed as the standard deviation ( $1\sigma$ ) of these 10 injections. The peak-to-peak value is the difference between the lowest and the highest values of the 10 analysis.

Instrument	Repeatability (ppb, $N = 10$ )	Peak to peak (ppb)
FTIR	0.09	0.26
CRDS	0.03	0.11
DFG	0.17	0.55
ICOS-SD	0.02	0.04
ICOS-EP38	0.02	0.06
ICOS-EP40	0.02	0.06
QC-TILDAS	0.02	0.05

standard gas measured several times a day could be used to correct this drift.

The optimal averaging time can be estimated by using Allan standard deviation plots. These plots can also be used, with LTR assessment, to estimate the stability of an instrument and decide on a calibration strategy. Figure 1 shows the time series of the 30 h target test for each instrument (two upper panels), and the Allan deviation plotted against the averaging times using a logarithmic scale (lower panel). With the Allan standard deviation assessment, we can define two main categories. First is the category of instruments with a high precision for high-frequency measurements (ICOS-SD, ICOS-EP or QC-TILDAS instruments). They present their best averaging time for intervals shorter than 5 min and higher variability over longer averaging times. The other category regroups the instruments with better stability over longer averaging intervals; the best averaging time is from 10 min to 1 h or higher (CRDS, FTIR, DFG and ICOS-EP38). Some instruments, such as the ICOS-EP38, have strong performances in both categories: high precision for high frequencies and good stability. These two types of performances



**Figure 1.** Allan deviation assessment for all instruments. The two upper panels present the time series for the eight instruments over at least 30 h. The lower panel presents the Allan deviation for all instruments from 1 s to 3.10<sup>4</sup> s (logarithmic scale). The color codes for the instruments and the test periods are given in the legend. The two vertical lines in the lower panel correspond to an averaging time of 1 min and 1 h.

will interest different research communities: the high precision for high frequencies will interest anyone working on short time phenomena (< 1 min), such as eddy covariance studies. The second category will interest communities working on typically 10 min to hourly averaged data, which is the case of atmospheric background monitoring stations, such as the ICOS atmospheric network. It should be noted that all of the instruments tested at the MLab for this study achieved the CMR specifications given by the manufacturers.

**Table 4.** Long-term repeatability computed from the standard deviation ( $1\sigma$ ) of the mean values over the last 5 min of 30 target measurements ( $N$ ). The peak-to-peak value is the difference between the lowest and the highest values of the  $N$  target measurements. The calibration frequency gives the mean time between two calibrations. The calibrations were applied as drift corrections.

Instrument	$N$	Period (min)	$1\sigma$ (ppb)	Peak to peak (ppb)	Mean calibration frequency
GC	1 year		0.29		30 to 45 min
FTIR	30 (7 days)	40	0.07	0.27	20 days
CRDS	30 (12 days)	20	0.07	0.28	11 days
DFG	30 (12 days)	20	0.21	0.86	8 days
ICOS-SD	30 (7 days)	20	0.32	1.00	9 days
ICOS-EP38	30 (7 days)	30	0.25	0.70	13 days
ICOS-EP40	30 (7 days)	30	0.29	0.60	13 days
QC-TILDAS	27 (6 days)	30	0.14	0.44	30 days

### 3.3 Short-term repeatability (STR) assessment

Because the CMR test is an assessment of the precision of the instrument over continuous measurements, the STR assessment quantifies the ability of one instrument to always reach the same value for a target gas when alternated with a different sample. For this test, a target gas is measured 10 times for 15 to 20 min alternating with dry ambient air measurements for 5 min. From our experience with other analyzers, 15 to 20 min should be appropriate for all instruments to stabilize and to provide at least 5 min of stable measurements. Similar to the CMR assessment, no calibration or drift corrections are applied. A  $N_2O$  mean value is then calculated for each injection of the target by taking the last 5 min of each analysis. The repeatability is expressed as the standard deviation ( $1\sigma$ ) of the 10 injections, and the results are presented in Table 3.

The STR is approximately the same for all instruments ( $\approx 0.02$  ppb). Only the FTIR and DFG instruments show higher STR of 0.09 and 0.17 ppb, respectively. Part of the difference between the FTIR and DFG and the other instruments can be explained with the CMR, as the FTIR and DFG are the least precise instruments for small averaging time (1 to 5 min). As a consequence, when measuring calibration gases, FTIR and DFG owners would need to increase the measurement time to 20 to 30 min and then keep the last 10 min to reach a better STR.

### 3.4 Long-term repeatability

The LTR assessment tests quantify the stability of an analyzer over periods of several days. For each instrument, a target gas was measured regularly (at least twice a day) alternating with ambient air for several days in our temperature-controlled laboratory. Depending on the instrument type and the test period, the target measurements were performed for a period of 20 min for the instruments that were compared during the first campaign, 30 min for the second and third campaigns and 40 min for the FTIR due to its cell size, which needs more time for the stabilization of the physical param-

eters. For all instruments, a mean value was calculated over the last 5 min of each analysis. A calibration was performed every week or 14 days and was applied as a drift correction, with a linear interpolation between the bracketing calibrations. Table 4 shows the standard deviation ( $1\sigma$ ) over 30 measurements for all tested instruments.

The two instruments showing the best LTR are the FTIR and CRDS with a standard deviation of 0.07 ppb. They are the only two instruments that can reach the compatibility goal recommended by the WMO. The QC-TILDAS, with a precision of 0.14 ppb, is just above the recommendations, but the instrument tested had no pressure control and its pressure needed regular adjustment during this test. We can expect an improvement of the LTR for the QC-TILDAS when using a pressure controller. The three ICOS-QCL instruments and the DFG instrument present a LTR between 0.21 and 0.32 ppb. To meet the WMO recommendations, the calibration frequency may need to be increased to one to several calibrations per week. To test this point, the ICOS-EP40 was re-tested from November to December 2014. During this period, a sequence of analysis of 1 h of air alternated with 15 min of a target gas was used. The target gas measurements were separated into two data sets. One was used as a target gas, and the other was used as a calibration gas, to correct the first data set as a one-point calibration. Different LTRs were calculated by choosing different frequencies for the calibration data set. Without any calibrations, the LTR was 0.85 ppb (over 3 weeks), and with a calibration every 2 days, the LTR was 0.28 ppb. For a calibration every 12 h, the LTR improved to 0.07 ppb, and for every 2.5 h (one target gas alternated with one calibration gas), the LTR reached 0.03 ppb. Thus, to reach a LTR better than 0.10 ppb for the ICOS-EP40, a calibration frequency of twice a day is necessary.

### 3.5 Linearity assessment

For each instrument, linearity assessments were made using calibration tanks with known  $N_2O$  mole fractions. As explained in Sect. 3.1, three calibration sets of four to six

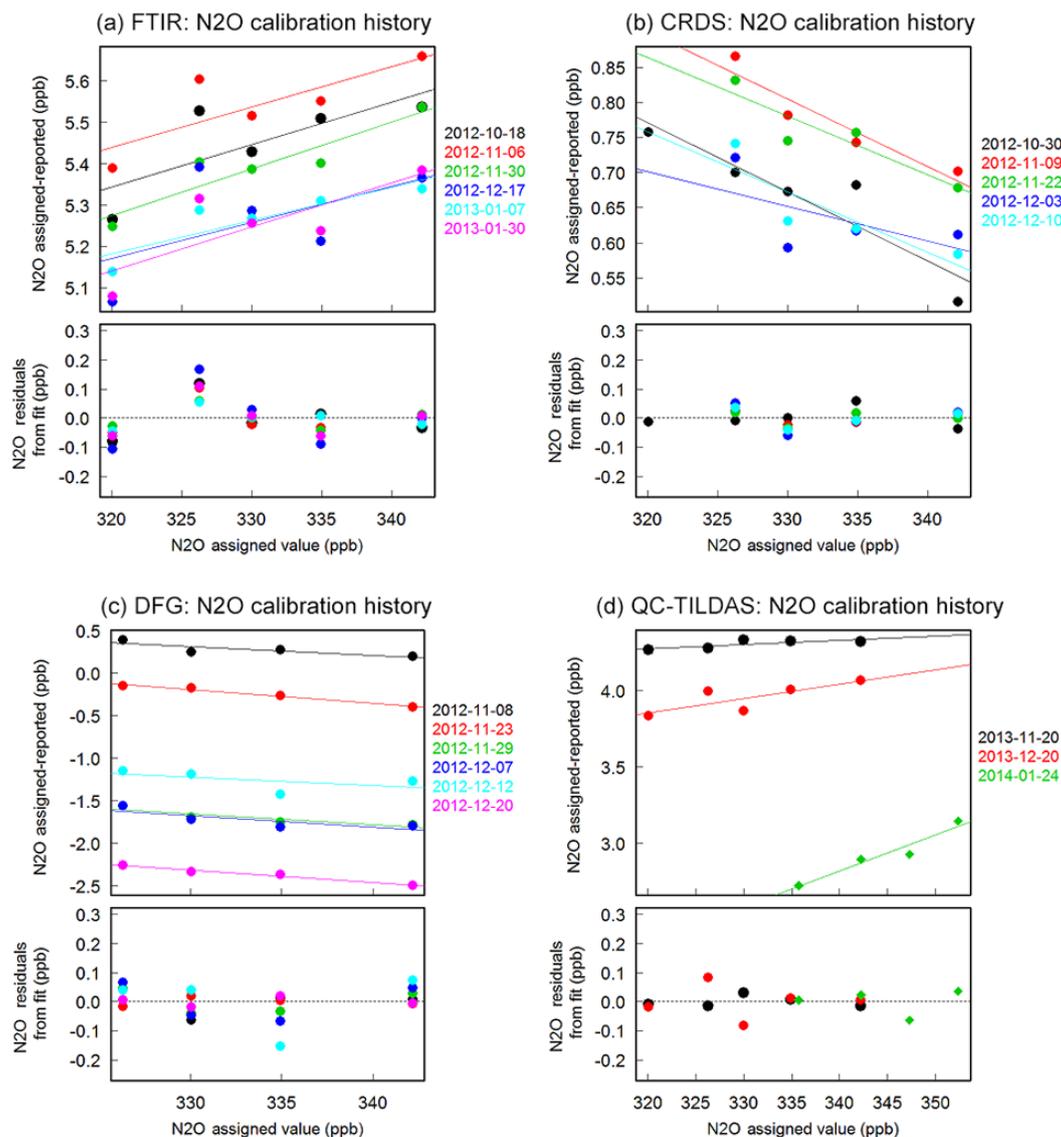


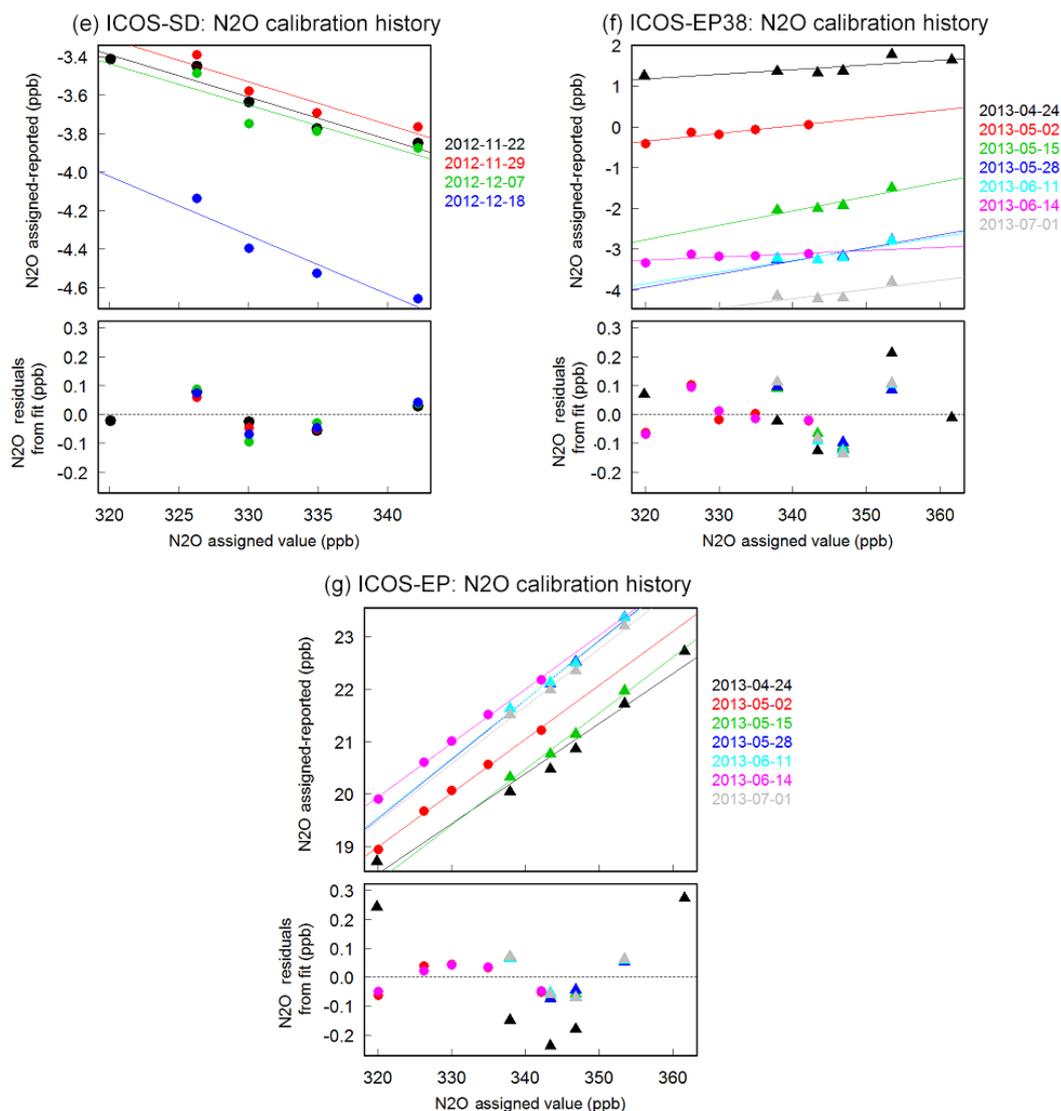
Figure 2.

**Table 5.** Long-term drifts: the drifts between two consecutive calibrations from the same calibration set are normalized over a time span of 10 days. The drifts from all consecutive calibrations are then averaged to obtain a mean drift for all analyzers.

Instrument	Mean drift for 10 days (ppb)	Highest drift (ppb)
FTIR	0.12	0.23
CRDS	0.07	0.19
DFG	1.02	2.53
ICOS-SD	0.30	0.71
ICOS-EP38	0.76	1.62
ICOS-EP40	0.31	1.08
QC-TILDAS	0.12	0.16

different tanks were used during the campaigns. The mole fraction measured by the instrument compared with the assigned mole fraction was used to assess the linearity of the instrument. The linearity assessment for each instrument is displayed in Fig. 2.

All of the analyzers show a linear response curve, which can be described by a linear fit using several calibration cylinders. To reduce the errors in the assessment of the calibration cylinders, we recommend the use of at least three calibration gases, spanning the full atmospheric range. During our linearity assessment, we looked at the deviation of individual tanks from the fit curve (Fig. 2, lower panel for each instrument) and used this as a measure of the linearity of an analyzer. We found typical residuals of up to  $\pm 0.15$  ppb for the



**Figure 2.** Linearity tests for each instrument are plotted in individual graphs. The upper panel presents the difference between the certified values of the calibration scale and the values measured for all the calibrations made during the tests. The cylinders from the MPI scale are represented with circles, and those from the DS scale are represented by triangles. Squares for the fourth calibration set are only used once by the QC-TILDAS. The lower panel presents the residuals from the fit. The color code for the calibration dates is given in the legend.

FTIR, ICOS-QCL, and DFG and  $\pm 0.05$  for the CRDS and QC-TILDAS analyzers.

The linear fit of the differences between assigned values minus measured values plotted against the assigned value (upper panel) show different slopes, depending on the instrument. Even with the same analyzer model, such as the ICOS-EP38 and ICOS-EP40, the slopes differ considerably. From the time evolution of the linear fit function, we can extract further information about the long-term stability and calibration frequency needed. For each calibration cylinder, we measure the drift between the consecutive calibration runs. Then these drifts per day are normalized to drifts per 10 days. Finally we average the drifts from all calibration cylinders to

extract the mean and maximum drift (Table 5). Overall, this study confirms the results from the much shorter 30 h test presented in Table 2. The FTIR, CRDS and QC-TILDAS show a mean drift of approximately 0.1 ppb per 10 days, which justifies a calibration frequency of 10–14 days. The ICOS-SD, ICOS-Eps and DFG show a mean drift over 10 days between 0.3 and 1 ppb, which suggests that they should be calibrated at least every 3 days or daily to obtain an equivalent correction of the drift.

### 3.6 Stabilization time

Another important parameter is the time necessary for the instrument to reach a stable value when changing the sam-

**Table 6.** Stabilization time: time necessary to reach the final value (calculated over the last 5 min of an analysis) at either  $\pm 0.1$  or  $2\sigma$  ppb (from CMR test; 1 min value) of the final value. The stabilization time is averaged over at least 24 injections of cylinders from calibration sets.

Instrument	$\pm 0.1$ ppb of final value		$\pm 2\sigma$ of final value	
	Stab. time (min)	Not reached* (%)	$2\sigma$ (ppb)	Stab. time (min)
FTIR	–	70	0.298	$10 \pm 7$
CRDS	$11 \pm 5$	6	0.11	$10 \pm 6$
DFG	$17 \pm 2$	9	0.318	$2 \pm 3$
ICOS-SD	$2 \pm 1$	0	0.248	$1 \pm 1$
ICOS-EP38	$2 \pm 2$	0	0.128	$2 \pm 2$
ICOS-EP40	$2 \pm 1$	0	0.136	$2 \pm 0$
QC-TILDAS	$1 \pm 1$	0	0.125	$1 \pm 0$

\* The “not reached” value is the percent of runs that did not reach  $\pm 0.1$  ppb of the final value.

ple analyzed. This test is made by using the calibration runs. The calibration sets the  $\text{N}_2\text{O}$  mole fraction differences between the different samples ranging from 3 to 16 ppb. For each analysis of a calibration cylinder, the raw data are first averaged over 1 min intervals, and the final values are calculated by averaging the last 5 min of a 15 to 20 min sequence. We estimated the stabilization time by examining the time from which all the 1 min averaged data stay within  $\pm 0.1$  ppb or  $\pm 2\sigma$  ppb (see CMR test for 1 min averaged data, Table 2) of the final value. For all instruments, the inlet system consisted in pressure regulators (SCOTT MODEL 14 M-14C, nickel-plated brass) installed on each cylinder, connected to a Valco multi-port valve (VICI) using 2 to 4 m of either 1/4 in. OD Synflex 1300 (EATON) tubing for the FTIR and QC-TILDAS or 1/16 in. OD stainless steel tubing for the other instruments. A short length of similar tubing was used to connect the Valco valve to the inlet of the instruments. It should be noted that such an inlet system did not impact the time of stabilization as there are nearly no dead volumes and the volume to flush (mainly the tubing) is not significant in regards to of the flow rates (short residence time). The stabilization time is a function of the cell volume and design, dead volume and sample flow rate. The results found in this study are only valid for the sample flow rates that were considered and for our inlet systems. Other inlet systems should be mindful of any possible dead volumes or the influence of the tubing length. We used the flow rates recommended by the manufacturers, which are documented in Table 1. The amplitude of the concentration change compared to the sample analyzed previously could also influence the stabilization time, but despite concentration changes ranging from 3 to 16 ppb no correlation was found between the two. The values in Table 6 are the stabilization times of all the instruments obtained by averaging the stabilization times calculated for at least 24 cylinder runs.

When choosing 0.1 ppb as the criterion for reaching stabilization, the instruments can be classified into two cate-

gories: in the first category, the stabilization is reached after 1 to 2 min (ICOS and QC-TILDAS); in the second category, the stabilization is reached later or never (FTIR, CRDS and DFG). These last results can be easily explained by the CMR test (Table 2) because for some instruments, the  $\pm 0.1$  ppb criterion cannot be reached for 1 min averaged data. To make a meaningful comparison, a criterion of  $\pm 2\sigma$  ppb of the final value was chosen. In this case, the ICOS, DFG and QC-TILDAS instruments rapidly reach the final values (under 3 min), but the FTIR and CRDS instruments require much more time to achieve stabilization (more than 10 min). As a consequence, instrument owners should be mindful of the time required to reach stabilization to keep only the relevant data.

### 3.7 Temperature dependence

All tests and measurements described previously were performed in a laboratory with temperature variations of less than  $\pm 1^\circ\text{C}$ . However, the working conditions at stations where the analyzers will be installed might not always be as stable. Temperature-dependence tests were conducted to characterize the sensitivity of the instruments to room temperature variations. While continuously measuring a target tank, the temperature of the laboratory was changed. From the laboratory working conditions ( $22 \pm 1^\circ\text{C}$ ), the temperature was varied between a low temperature (15 to  $20^\circ\text{C}$ ) and a high temperature (28 to  $35^\circ\text{C}$ ) before returning to the normal working temperature. The low and high temperatures were maintained for several hours to allow for stabilization. Depending on the season or time period when the instrument was tested, the span of the variation differs between 10 to  $17^\circ\text{C}$ . Due to the high gas consumption of the FTIR, the target tank measurements for this instrument were analyzed not continuously but rather every 6 h at different temperatures for 3 days, and the last 7 min of each measurement were kept for this test.

In Fig. 3, a two-panel plot for each instrument is presented to describe the  $\text{N}_2\text{O}$  response to the room temperature changes. Table 7 summarizes these results with the room temperature change applied to the instrument, the type of temperature dependence and its slope when a linear dependence was found.

Most of the instruments show a significant sensitivity to room temperature variations. Only the QC-TILDAS instrument and the ICOS-EP38 do not show significant temperature dependence for  $\text{N}_2\text{O}$ , with variation below 0.1 ppb for five degree variations. It should be noted that for the QC-TILDAS test, the high-frequency variations of  $\text{N}_2\text{O}$  at the beginning and near the end are due to pressure variations inside the cell. The FTIR and CRDS instruments show a linear dependence to the temperature of  $-0.04$  and  $0.05$  ppb per  $^\circ\text{C}$ , respectively. The CRDS instrument tested was a prototype, and thus no correction for temperature was applied at this stage. Such correction is now built in the commer-

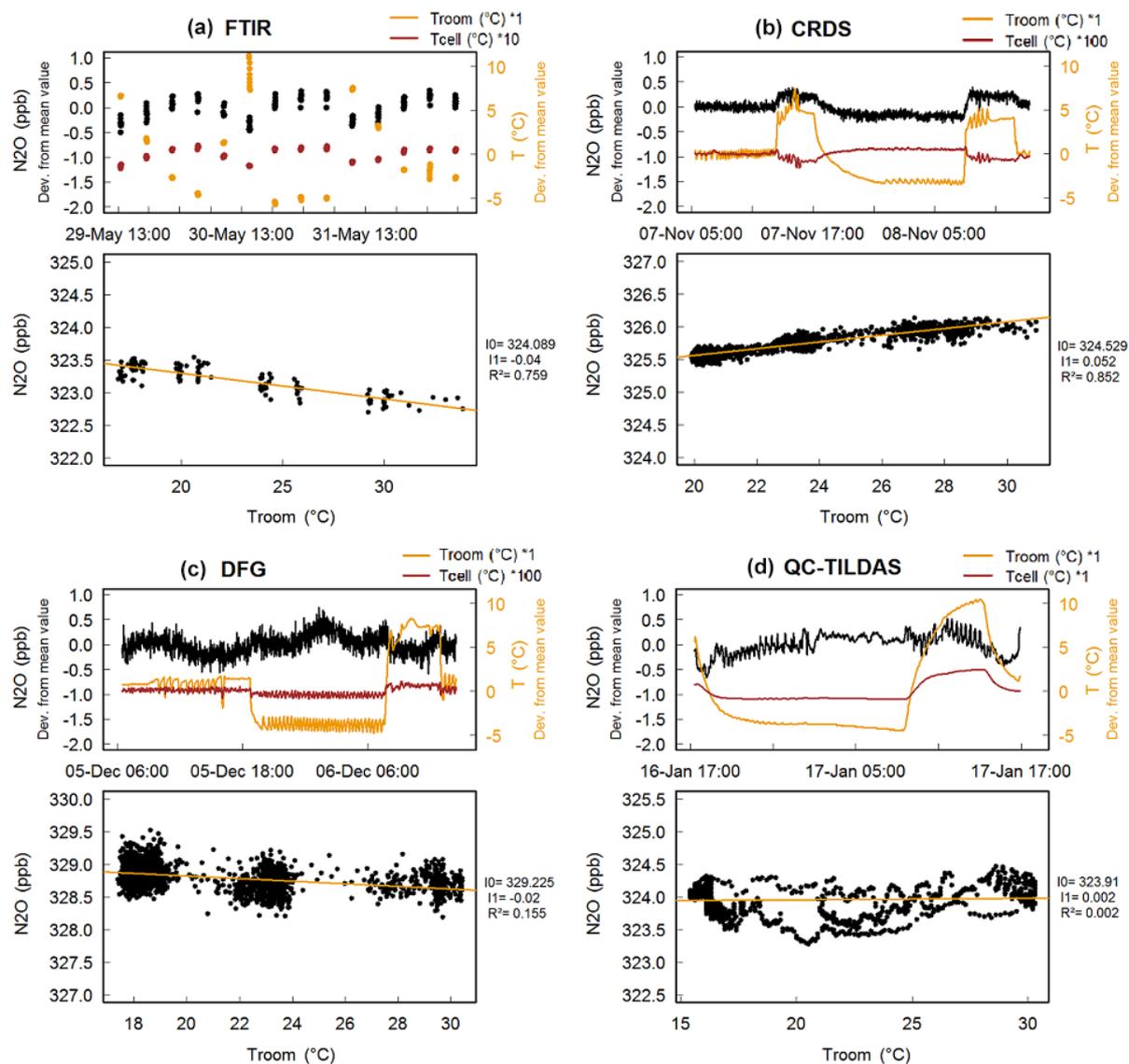


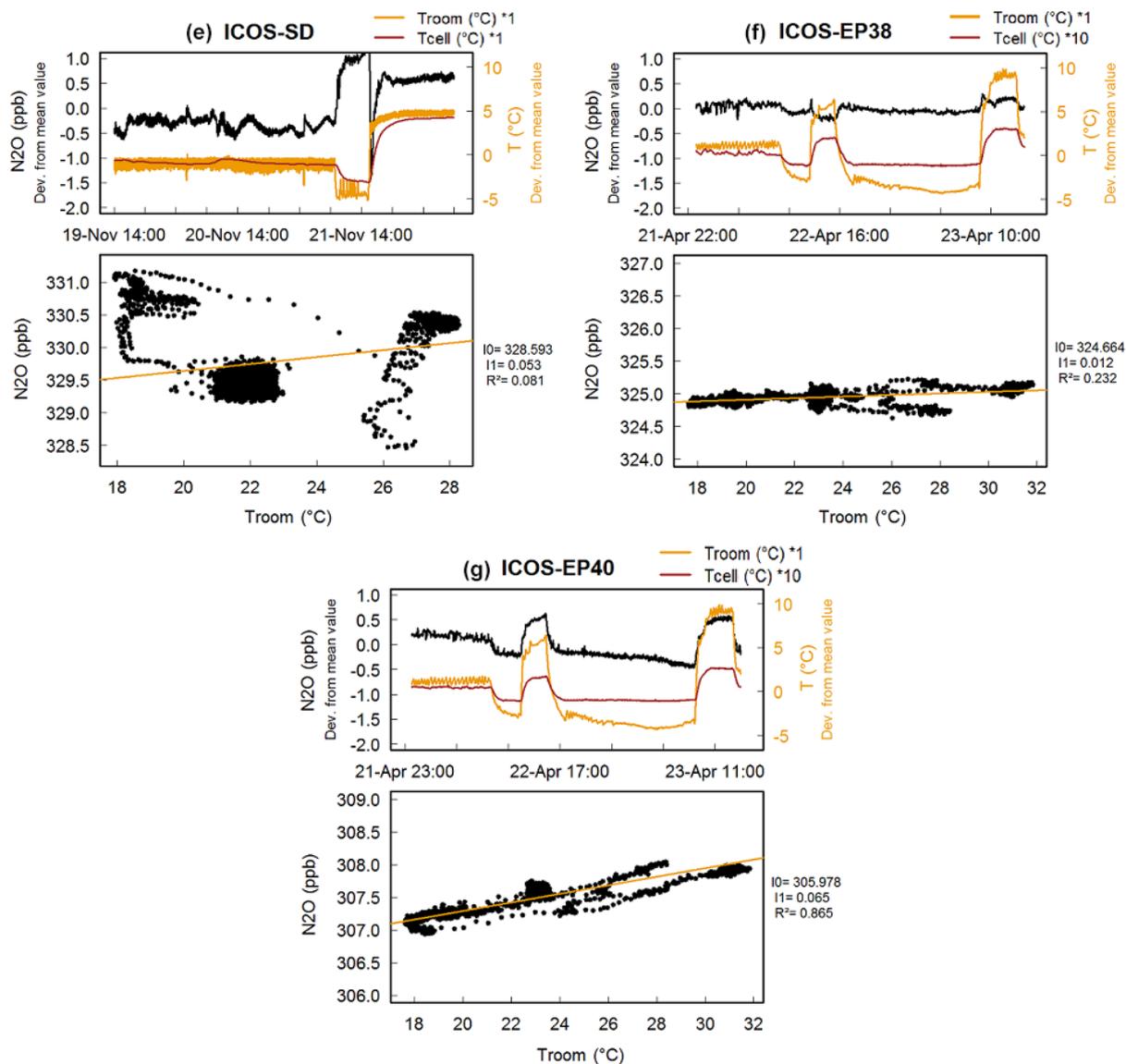
Figure 3.

cialized version, and in May 2014 we had the opportunity to test a newly purchased CRDS analyzer with temperature correction in the MLab. It shows an improved behavior to room temperature changes, with a sensitivity to temperature of less than 0.02 ppb of N<sub>2</sub>O per °C. The DFG instrument presents a dependence that is not significant compared with the relatively large noise. A larger temperature sensitivity was found for the ICOS-SD with approximately 2 ppb N<sub>2</sub>O changes (peak to peak), but the nonlinear relationship makes it impossible to apply a correction. The ICOS-EP model does improve the temperature control compared with the standard model, but it is important to highlight the difference between the instruments: although instrument ICOS-EP38 presents no significant temperature influence, instrument ICOS-EP40 shows a temperature dependence of 0.07 ppb N<sub>2</sub>O per °C. To

reach the best attainable performance, most instruments need a temperature-controlled environment, especially the FTIR and ICOS-SD. If an instrument presents a linear dependence, it is also possible for the user to add an instrumental specific correction that could be applied to the final data. In this case, the room temperature needs to be monitored precisely, and the temperature dependence needs to be determined accurately by repeating the temperature test two to three times.

### 3.8 Water vapor correction

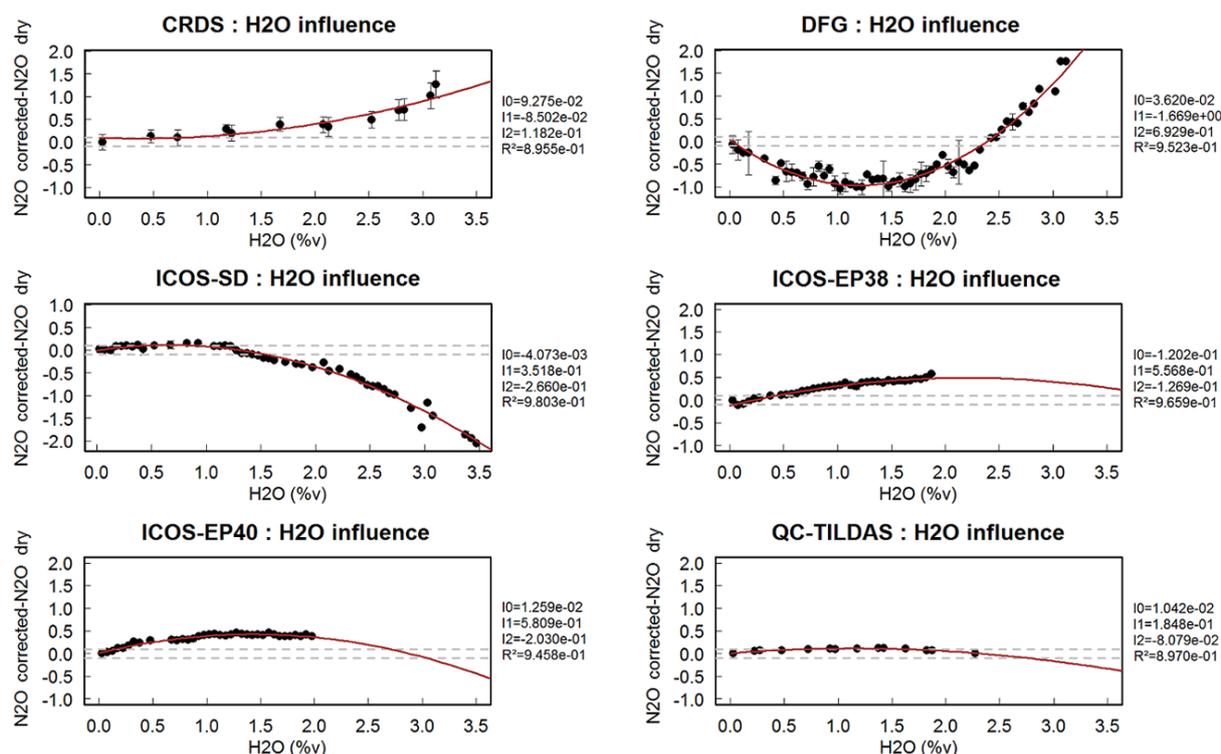
Water vapor in the atmosphere can vary from a few ppm to several percent of volume. Usually, the N<sub>2</sub>O measurements are presented as a dry mole fraction, and a drying system is needed for ambient air measurement. Several of the instru-



**Figure 3.** The results of the temperature-dependence test for each instrument tested. The top panel presents the time series of concentration (black), the room temperature in the laboratory (orange) and the temperature in the cell (red). For some instruments, the temperature in the cell was multiplied by either 10 or 100 to make the variations visible on the same scale as the room temperature (right axis). The concentration of N<sub>2</sub>O is plotted against the room temperature in the lower panel. On the right of the lower panel,  $I_1$  is the slope,  $I_0$  is the intercept and  $R^2$  is the coefficient of determination of the linear regression.

ments tested provide water vapor measurements and a correction function to transfer wet ambient air measurements to the dry mole fraction. This correction accounts for dilution and spectroscopic effects such as pressure broadening (Chen et al., 2013). In this study, we test the water vapor correction applied by the manufacturer of the different instruments. This test was not performed for the FTIR and GC because the FTIR has its own built-in drying system, which removes the water vapor to 2–4 ppm, and because the GC is required to measure dry air only.

This test consists of measuring a high-pressure tank filled with dry natural air and then injecting a droplet of Milli-Q water (0.2 mL) on a hygroscopic filter (M&C LB1SS) to humidify the stream. This water droplet humidifies the gas at approximately 3 % vol of water depending of the room temperature and sample pressure. After this, the dry natural air from the high pressure tank dries slowly the filter (droplet evaporation). With this method, the tanks of dry ambient air were humidified at varying levels, up to 2–3 % vol of water vapor. However, this method, although easy to implement, does not offer a steady drying rate over all the H<sub>2</sub>O range, re-



**Figure 4.** Water vapor correction for all instruments except the FTIR. All the data were averaged over 30 s and separated into bins of 0.05 % of H<sub>2</sub>O. The dashed lines represent  $\pm 0.1$  ppb of the dry value. On the right-hand side of the panels,  $I_0$ ,  $I_1$  and  $I_2$  are the coefficients, and  $R^2$  is the coefficient of determination of the polynomial regression.

**Table 7.** Influence of room temperature on N<sub>2</sub>O.

Instrument	Temperature dependence	Temperature range (°C)	Temperature dependence (ppb °C <sup>-1</sup> )	Peak to peak (ppb)
FTIR	Linear	17 to 34	-0.04	0.84
CRDS	Linear	20 to 31	+0.05	0.73
DFG	Linear	17 to 30	-0.02	1.33
ICOS-SD	No linear dependence	18 to 28	NA*	2.70
ICOS-EP38	No significant dependence	17.5 to 32	NA*	0.60
ICOS-EP40	Linear	17.5 to 32	+0.07	1.11
QC-TILDAS	No significant dependence	15 to 30	NA*	1.19

\* Denotes cases where it was not possible to give a value because either there was no dependence or it was not linear.

sulting in few measurement data over part of the H<sub>2</sub>O range. In order to get a better statistical weight on these H<sub>2</sub>O range parts, the method is repeated at least three times for all of the instruments. The assessment of the water vapor correction is made by comparing the values of the wet target found by the instrument to its dry value. When the QC-TILDAS and the commercialized model of the CRDS were tested, a new method to characterize the water vapor correction had been implemented by the MLab. A humidifying bench is composed of one thermal mass flow controller (F-201CV, Bronkhorst), to regulate the flow of a tank filled with dry

natural air, one liquid mass flow controller (Mini Cori-Flow M12, Bronkhorst), to regulate the quantity of Milli-Q water injected in the sample line, and one controlled evaporator mixer (Bronkhorst) to humidify the target gas by evaporating the water at 40 °C while mixing it with the gas. This setup enables a precise control of the water vapor percentage in the sample analyzed. The target gas can now be humidified at different H<sub>2</sub>O levels (up to 5 % vol of water vapor) with a suitable stability (H<sub>2</sub>O standard deviation of 100 ppm) as long as required, allowing long data set averaging and thus

improving the representativeness of the results, especially for noisy analyzers.

The manufacturers Picarro, Los Gatos, Thermo Fisher and Aerodyne provided a water vapor correction for their instruments. The correction was not yet implemented in the CRDS prototype tested; therefore, we did not include the results for this instrument. Figure 4 shows the difference between water vapor corrected and the dry  $\text{N}_2\text{O}$  mole fraction against the concentration of  $\text{H}_2\text{O}$  for each instrument. All data were averaged in 30 s intervals.

This test demonstrates the difficulty of most analyzers to provide a correct water vapor correction when measuring wet air. Of all of the instruments, only the QC-TILDAS supplies an accurate water vapor correction: its corrected wet measurements of  $\text{N}_2\text{O}$  did not exceed 0.1 ppb compared with the dry mole fraction. The ICOS-SD supplies a correction that results in a  $\text{N}_2\text{O}$  difference to the dry value below 0.2 ppb for  $\text{H}_2\text{O}$  not exceeding 1–1.5 % vol. For higher  $\text{H}_2\text{O}$  values, the correction shows larger differences of up to 2 ppb. The two ICOS-EP corrections are not sufficient, with a  $\text{N}_2\text{O}$  difference to the dry value of up to 0.5 ppb for high water vapor concentrations. The DFG instrument correction is clearly not suitable, with a difference in the  $\text{N}_2\text{O}$ 's dry/wet values between  $-1.0$  and  $+2.0$  ppb. Finally, the commercialized version of the CRDS supplies a correction that results, similar to the ICOS-SD, in a  $\text{N}_2\text{O}$  difference to the dry value below 0.2 ppb for  $\text{H}_2\text{O}$  not exceeding 1 % vol. However, for higher values of water vapor, the  $\text{N}_2\text{O}$  difference increases to 1.5 ppb. As a result, to achieve the best performances for high-precision atmospheric  $\text{N}_2\text{O}$  measurements, most instruments will need a drying system prior to the inlet or a careful evaluation of the water vapor dependence, with the exception of the FTIR, which has a built-in drying system. While the QC-TILDAS tested here showed a good water correction, users of this instrument should still test the water correction.

Here, we can only recommend using a drying system for high-precision  $\text{N}_2\text{O}$  measurements with all of the instruments tested. However, if some stations or laboratories are sufficiently equipped to make their own instrument-specific water vapor dependency test on a regular basis, wet air measurements could then be performed.

### 3.9 Ambient air measurement comparisons

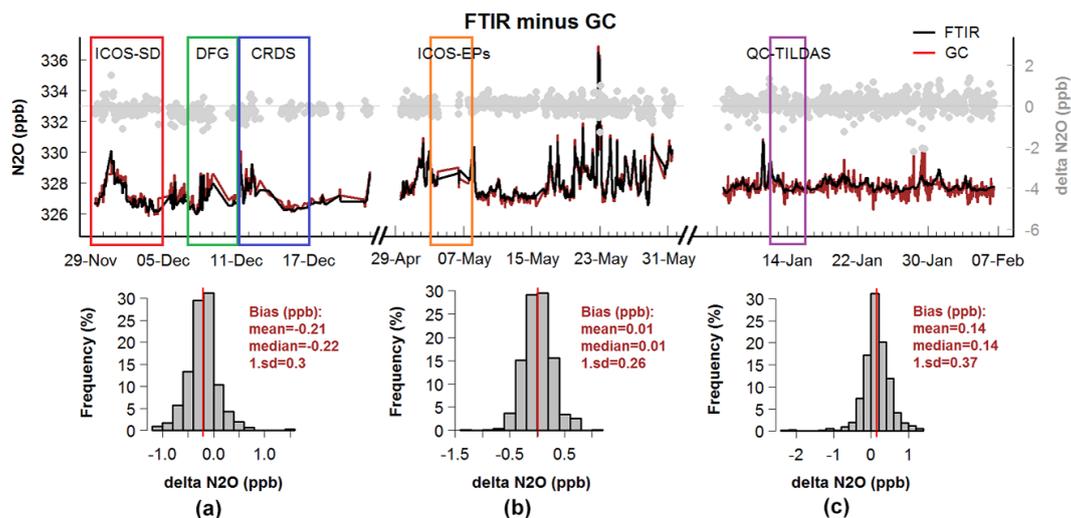
All of the instruments that were tested measured at least 100 h of ambient air during the testing period. The measurements were made at the MLab, as described in Sect. 3.1. Pumps were used to reduce the residence time in the air line to avoid time differences between the measurements of the different instruments. For all instruments, the measurements were hourly averaged to allow for meaningful comparisons and to reduce the influence of short time variations. For the three test periods, the GC and the FTIR were the only instruments that were always present. Figure 5 presents the comparison between these instruments over the three peri-

ods. It can be observed that although the  $\text{N}_2\text{O}$  mole fraction ranged from 325 to 338 ppb, most of the peaks were less than 2 to 3 ppb in height. Although the mean difference between the instruments is different in each period ( $-0.21$  ppb for the first, 0.01 ppb for the second and 0.14 ppb for the third), it was constant during each period, with a standard deviation between 0.26 and 0.37 ppb. Because the FTIR showed a smaller standard deviation than the GC during these periods, it was chosen as the reference instrument for the comparison with all of the other instruments.

During the first test period (CRDS, ICOS-SD and DFG), a water trap was used to dry the air (see Sect. 3.1.), and the dry air measurements were then compared. During the second period (ICOS-EP38 and ICOS-EP40), there was not enough common dry air data for the ICOS-Eps and the FTIR to conduct the comparison. Therefore, we were only able to compare the wet air measurements, which were corrected for the water vapor by the correction algorithms provided by the manufacturers (between 0.7 and 1.6 % of water vapor during the period). For the third period (QC-TILDAS), the instrument only measured wet air, so the comparison was conducted on the values of the QC-TILDAS with the manufacturer's water vapor corrections applied (1 % of water vapor or less during the period). For all comparisons, the 100 h periods were chosen among the most stable consecutive data available (see Fig. 5 for the period chosen for each instrument). For all instruments measuring wet air, we attempted to apply the corrections obtained from the water vapor test to the dry air values (Sect. 3.8.); however, it either had no effect (QC-TILDAS) or did not improve the comparison (ICOS-EP). Thus, for all of these instruments, the air comparison was performed with the dry values given by the instrument. The data were calibrated by doing an interpolation between the calibration before and after the comparison period.

Figure 6 presents the relative difference histograms for each instrument compared with the FTIR. Of the six instruments that were compared with the FTIR, the ICOS-SD, the ICOS-EP40 and the QC-TILDAS show an offset of the mean difference with the FTIR of more than 0.10 ppb, whereas the other three instruments show an offset smaller than 0.05 ppb. However, as observed previously, there is an offset between the FTIR and GC for the first and third periods (0.21 and 0.14 ppb). When conducting the comparison with the GC, the offset with the QC-TILDAS improved to 0.12 ppb, but compared with the ICOS-SD, CRDS and DFG, the offset increased to 0.25 to 0.38 ppb. As discussed in Sect. 3.5, increasing the frequency of the calibrations should decrease the offset for the DFG and the ICOS-EP38. For the QC-TILDAS, the calibration frequency was once every month, which should be increased to once every week or 2 weeks for the QC-TILDAS, as indicated by the small drift time (see Sect. 3.5).

As observed with the different standard deviations in Fig. 6, the CRDS and the two ICOS-EPs are the instruments that show the best correlation with the FTIR. For the



**Figure 5.** Air comparison between the FTIR and the GC (1 h averaged data) during the time when the air comparison tests were performed with the other instruments. The top panel presents the times series of the GC (red) and the FTIR (black) for the three periods (December 2012, May 2013 and January 2014). In grey is the difference between the instruments. The colored frames show the time periods chosen to conduct the air comparison between the FTIR and the other instruments (see Fig. 6). In the lower panel, three histograms give the difference for the two instruments for the (a) first test period, (b) second test period and (c) third test period.

comparison with the QC-TILDAS, the standard deviation of 0.16 ppb cannot be explained by the drift (only 0.05 ppb for 100 h). The lack of good pressure control is probably what caused this value. Finally, for the DFG and the ICOS-SD, the standard deviations are the highest of all instruments. This is easily explained by the high variability these instruments have shown, and a calibration every 8–9 days is clearly not sufficient. Once again, we see the importance of choosing a calibration frequency adapted to each instrument and its absolute necessity when trying to compare air measurements from different instruments or, on a larger scale, networks.

## 4 Summary

Here, we briefly summarize the most important findings of the different tests performed to specify the performance of  $\text{N}_2\text{O}$  analyzers for atmospheric measurements.

### 4.1 Continuous measurement repeatability

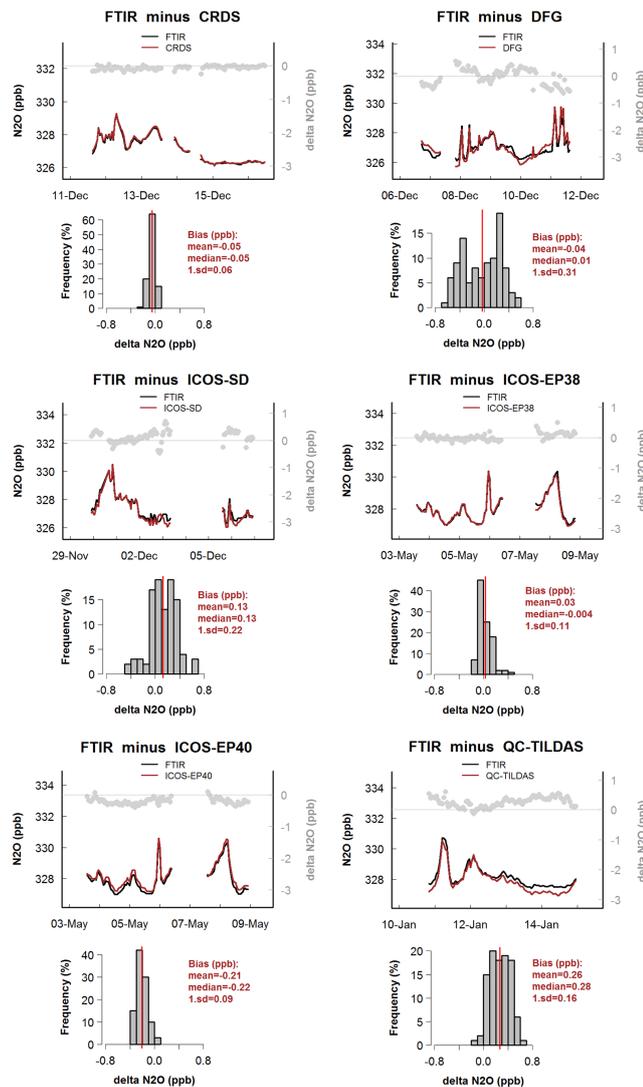
The raw data measurement interval varies between 1 s for the QC-TILDAS and 1 min for the FTIR. For atmospheric measurements at a tower, a typical averaging time between 1 to 10 min is used. The CRDS shows the best CMR for a 10 min average with a  $1\sigma$  standard deviation of 0.026 ppb. The CMR for the ICOS-EP and QC-TILDAS is approximately 0.07 ppb (10 min average), whereas that of the FTIR is 0.055 ppb. DFG and ICOS-SD are less appropriate for tower measurements, as the 10 min averages have a CMR greater than 0.1 ppb. For 1 min or less averaged data, the ICOS-EP and QC-TILDAS show the best CMR (0.1 ppb).

### 4.2 Stabilization/flushing time

Due to different cell/cavity volumes, geometry and flow rates, the flushing and stabilization time after a sample change (with contrasted level of  $\text{N}_2\text{O}$ ) is different for all analyzers. In our tests, we used the same inlet system for all analyzers and the flow rate recommended by the manufacturer for each instrument. These tests need to be performed at the field station prior to routine measurements because the flushing time also depends on the inlet system and the related dead volumes and flow rate used. To reach a stable  $\text{N}_2\text{O}$  value, which corresponds to  $\pm 2\sigma$  ppb of the final values, the CRDS and FTIR analyzers have a relatively long flushing time of more than 10 min. The ICOS and DFG analyzers varied from 2 to 3 min, whereas the QC-TILDAS, which is widely used for eddy covariance measurements, needs the shortest flushing time with less than 1 min after the change of a sample.

### 4.3 Temperature dependency

Temperature dependency of the instrument response is a major issue for stations and laboratories without air conditioning or with poor air conditioning. Daily room temperature changes can easily be  $5^\circ\text{C}$  or more if the measurements are performed in a container. Most of the tested instruments show temperature-dependent drifts. For most instruments, the dependency is linear, ranging from less than  $0.02 \text{ ppb } ^\circ\text{C}^{-1}$ , for the QC-TILDAS and ICOS-EP38, to  $0.07 \text{ ppb } ^\circ\text{C}^{-1}$ , for the ICOS-EP40, and could be corrected by the user. Only the ICOS-SD presents an important nonlinear temperature dependency and should be used in environments with fine control of the temperature.



**Figure 6.** Comparison between the FTIR and the other instruments: CRDS, DFG, ICOS-SD, ICOS-EP38, ICOS-EP40 and QC-TILDAS (1 h averaged data). The top panels present the times series of each instrument (red) and the FTIR (black). In grey is the difference between the instruments compared. In the lower panels, histograms give the difference for each comparison. All comparisons were conducted using 100 continuous hour-averaged air measurements. All data have been automatically corrected for water vapor using the manufacturer correction.

#### 4.4 Linearity and calibration strategy

All of the instruments showed response curves that can be fitted with linear curves. Using four calibration cylinders, the residuals differ between 0.06 and 0.10 ppb N<sub>2</sub>O for the different instruments. The calibration strategy chosen for the test with a 14-day frequency is only acceptable for the CRDS, FTIR and QC-TILDAS. For the other instruments, a more frequent calibration strategy needs to be developed. The results showed that for an ICOS-EP, a calibration frequency of

twice a day is necessary to reduce the LTR below the WMO recommendations.

#### 4.5 Water vapor

Wet ambient air measurements are influenced by dilution and interferences with atmospheric water vapor. The FTIR has an inbuilt drying system, whereas other manufacturers provide a H<sub>2</sub>O correction algorithm with H<sub>2</sub>O measurements. In this study, we tested the correction algorithms built in by the manufacturers for water vapor concentrations ranging from 0 to 3%. Nearly all of the correction algorithms showed large deviations of 0.5–1 ppb from the dry air value and are not suitable for our application. Only the QC-TILDAS instrument had a sufficient correction algorithm and showed differences smaller than 0.1 ppb.

### 5 Conclusions

A new standardized protocol to evaluate the performances of trace gases analyzers was implemented at the ICOS/ATC metrological laboratory in Gif-sur-Yvette. Yver Kwok et al. (2015) described the different tests performed for each instrument and showed examples for 47 CO<sub>2</sub> analyzers. Because our study, which was dedicated to the evaluation of N<sub>2</sub>O analyzers for high-precision atmospheric measurement, was conducted between October 2012 and December 2014, the experimental protocols were not fully finalized and have since been continuously improved due to gains in experience. Though not all of the analyzers were tested in the exact same way, the tests performed do not differ sufficiently to make meaningful conclusions impossible.

Most of the analyzers showed a clear dependency to the room temperature. This needs further investigation and technical improvements by the manufacturers. As long as the room temperature is still an issue, the N<sub>2</sub>O analyzers should be used in an air-conditioned environment, and the room temperature should be monitored to assess its evolution and the validity of the measurements. All of the tests demonstrated that the water vapor correction functions provided by the manufacturer are not sufficient to analyze wet ambient air. Therefore, we recommend that for high-precision atmospheric measurements, ambient air should be dried prior to the analysis.

During our initial tests, the calibration strategy was driven too much by our experiences from CO<sub>2</sub> and CH<sub>4</sub> analyzers and the wish to have a similar performance for N<sub>2</sub>O. With a calibration performed only every 14–21 days (Yver Kwok et al., 2015), some of the tested N<sub>2</sub>O analyzers show a significant drift, which cannot be corrected. In the case of the ICOS-EP40, we tested for possible improvement when adding a fifth reference cylinder, which is used to correct for short-term drift. In our case, an injection frequency of 11 h for a reference gas led to an improvement of the short-term

repeatability of the target gas from 0.85 to 0.07 ppb. Thus, prior to the use of an analyzer, the calibration strategy should be studied and optimized for the instrument and station conditions.

This study of seven analyzers shows that new optical techniques have the potential to replace the gas chromatographic techniques, which were widely used over the past 20 years for atmospheric measurements of N<sub>2</sub>O. These new techniques require much less maintenance at the stations and have lower operational costs because they do not need consumables, such as carrier gas. It should be noted that, while we studied only whole N<sub>2</sub>O without consideration of possible variations in isotopic composition, all these optical techniques are sensitive to some degree to isotopic composition and this dependence has not been assessed. Independent analyses of individual isotopologues are also not assessed here. Users should be mindful of possible isotopic dependences until further studies have been made. Achieving the WMO recommendation for N<sub>2</sub>O network compatibility of 0.1 ppb is still challenging but is absolutely needed to characterize the small variability at continental or coastal stations. This can only be reached at the moment if the above described recommendations are followed.

*Acknowledgements.* This work has been funded by the InGOS EU project (284274). We acknowledge the financial support given by CEA, CNRS and UVSQ for ICOS France. Special thanks to David Griffith for his help during the installation of the FTIR (2011–2012) and for the help during later updates. We would like to thank Picarro Inc, especially Eric Crosson and Chris Rella, for providing us the prototype of the N<sub>2</sub>O G5101-i analyzer. We also want to thank Hans-Juerg Jost and Thermo Fisher Scientific for providing us with their N<sub>2</sub>O IRIS 4600 analyzer. We are grateful to Marcel van der Schoot and three anonymous referees for their detailed and constructive reviews.

Edited by: D. Griffith

## References

- Bergamaschi, P., Corazza, M., Karstens, U., Athanassiadou, M., Thompson, R. L., Pison, I., Manning, A. J., Bousquet, P., Segers, A., Vermeulen, A. T., Janssens-Maenhout, G., Schmidt, M., Ramonet, M., Meinhardt, F., Aalto, T., Haszpra, L., Moncrieff, J., Popa, M. E., Lowry, D., Steinbacher, M., Jordan, A., O'Doherty, S., Piacentino, S., and Dlugokencky, E.: Top-down estimates of European CH<sub>4</sub> and N<sub>2</sub>O emissions based on four different inverse models, *Atmos. Chem. Phys.*, 15, 715–736, doi:10.5194/acp-15-715-2015, 2015.
- 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-Lasers O.*, 92, 403–408, 2008.
- 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.
- Hall, B. D., Dutton, G. S., and Elkins, J. W.: The NOAA nitrous oxide standard scale for atmospheric observations, *J. Geophys. Res.-Atmos.*, 112, D09305, doi:10.1029/2006JD007954, 2007.
- Hammer, S., Griffith, D. W. T., Konrad, G., Vardag, S., Caldow, C., and Levin, I.: Assessment of a multi-species in situ FTIR for precise atmospheric greenhouse gas observations, *Atmos. Meas. Tech.*, 6, 1153–1170, doi:10.5194/amt-6-1153-2013, 2013.
- Lopez, M., Schmidt, M., Yver, C., Messenger, C., Worthy, D., Kazan, V., Ramonet, M., Bousquet, P., and Ciais, P.: Seasonal variation of N<sub>2</sub>O emissions in France inferred from atmospheric N<sub>2</sub>O and Rn-222 measurements, *J. Geophys. Res.-Atmos.*, 117, D14103, doi:10.1029/2012JD017703, 2012.
- Nevison, C. D., Mahowald, N. M., Weiss, R. F., and Prinn, R. G.: Interannual and seasonal variability in atmospheric N<sub>2</sub>O, *Global Biogeochem. Cy.*, 21, GB3017, doi:10.1029/2006GB002755, 2007.
- Nevison, C. D., Dlugokencky, E., Dutton, G., Elkins, J. W., Fraser, P., Hall, B., Krummel, P. B., Langenfelds, R. L., O'Doherty, S., Prinn, R. G., Steele, L. P., and Weiss, R. F.: Exploring causes of interannual variability in the seasonal cycles of tropospheric nitrous oxide, *Atmos. Chem. Phys.*, 11, 3713–3730, doi:10.5194/acp-11-3713-2011, 2011.
- Popa, M. E., Gloor, M., Manning, A. C., Jordan, A., Schultz, U., Haensel, F., Seifert, T., and Heimann, M.: Measurements of greenhouse gases and related tracers at Bialystok tall tower station in Poland, *Atmos. Meas. Tech.*, 3, 407–427, doi:10.5194/amt-3-407-2010, 2010.
- Prather, M. J., Holmes, C. D., and Hsu, J.: Reactive greenhouse gas scenarios: Systematic exploration of uncertainties and the role of atmospheric chemistry, *Geophys. Res. Lett.*, 39, L09803, doi:10.1029/2012GL051440, 2012.
- Ravishankara, A. R., Daniel, J. S., and Portmann, R. W.: Nitrous Oxide (N<sub>2</sub>O): The Dominant Ozone-Depleting Substance Emitted in the 21st Century, *Science*, 326, 123–125, 2009.
- Scherer, J. J., Paul, J. B., Jost, H. J., and Fischer, M. L.: Mid-IR difference frequency laser-based sensors for ambient CH<sub>4</sub>, CO, and N<sub>2</sub>O monitoring, *Appl. Phys. B-Lasers O.*, 110, 271–277, 2013.
- Schmidt, M., Glatzel-Mattheier, H., Sartorius, H., Worthy, D. E., and Levin, I.: Western European N<sub>2</sub>O emissions: A top-down approach based on atmospheric observations, *J. Geophys. Res.-Atmos.*, 106, 5507–5516, 2001.
- 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.
- Thompson, R. L., Dlugokencky, E., Chevallier, F., Ciais, P., Dutton, G., Elkins, J. W., Langenfelds, R. L., Prinn, R. G., Weiss, R. F., Tohjima, Y., O'Doherty, S., Krummel, P. B., Fraser, P., and

- Steele, L. P.: Interannual variability in tropospheric nitrous oxide, *Geophys. Res. Lett.*, 40, 4426–4431, 2013.
- Wuebbles, D. J.: Atmosphere, Nitrous oxide: no laughing matter, *Science*, 326, 56–57, 2009.
- 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<sub>2</sub>O, CO<sub>2</sub>, CH<sub>4</sub> and CO, *Atmos. Meas. Tech.*, 8, 3867–3892, doi:10.5194/amt-8-3867-2015, 2015.