



# Two decades of flask observations of atmospheric $\delta(\text{O}_2/\text{N}_2)$ , $\text{CO}_2$ , and APO at stations Lutjewad (the Netherlands) and Mace Head (Ireland), and 3 years from Halley station (Antarctica)

Linh N. T. Nguyen<sup>1</sup>, Harro A. J. Meijer<sup>1</sup>, Charlotte van Leeuwen<sup>1</sup>, Bert A. M. Kers<sup>2</sup>, Hubertus A. Scheeren<sup>1</sup>, Anna E. Jones<sup>2</sup>, Neil Brough<sup>2,a</sup>, Thomas Barningham<sup>2</sup>, Penelope A. Pickers<sup>3</sup>, Andrew C. Manning<sup>3</sup>, and Ingrid T. Luijkx<sup>4</sup>

<sup>1</sup>Centre for Isotope Research, Energy and Sustainability Research Institute Groningen, University of Groningen, Groningen, the Netherlands

<sup>2</sup>British Antarctic Survey, Natural Environment Research Council, Cambridge, United Kingdom

<sup>3</sup>Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom

<sup>4</sup>Meteorology and Air Quality, Wageningen University and Research, Wageningen, the Netherlands

<sup>a</sup>now at: National Institute of Water and Atmospheric Research, Wellington, New Zealand

**Correspondence:** Linh N. T. Nguyen (n.t.l.nguyen@rug.nl)

Received: 25 June 2021 – Discussion started: 28 July 2021

Revised: 17 January 2022 – Accepted: 18 January 2022 – Published: 2 March 2022

**Abstract.** We present 20-year flask sample records of atmospheric  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and atmospheric potential oxygen (APO) from the stations Lutjewad (the Netherlands) and Mace Head (Ireland), and a 3-year record from Halley station (Antarctica). We include details of our calibration procedures and the stability of our calibration scale over time, which we estimate to be 3 per meg over the 11 years of calibration, and our compatibility with the international Scripps  $\text{O}_2$  scale. The measurement records from Lutjewad and Mace Head show similar long-term trends during the period 2002–2018 of  $2.31 \pm 0.07 \text{ ppm yr}^{-1}$  for  $\text{CO}_2$  and  $-21.2 \pm 0.8$  per meg  $\text{yr}^{-1}$  for  $\delta(\text{O}_2/\text{N}_2)$  at Lutjewad, and  $2.22 \pm 0.04 \text{ ppm yr}^{-1}$  for  $\text{CO}_2$  and  $-21.3 \pm 0.9$  per meg  $\text{yr}^{-1}$  for  $\delta(\text{O}_2/\text{N}_2)$  at Mace Head. They also show a similar  $\delta(\text{O}_2/\text{N}_2)$  seasonal cycle with an amplitude of  $54 \pm 4$  per meg at Lutjewad and  $61 \pm 5$  per meg at Mace Head, while the  $\text{CO}_2$  seasonal amplitude at Lutjewad ( $16.8 \pm 0.5 \text{ ppm}$ ) is slightly higher than that at Mace Head ( $14.8 \pm 0.3 \text{ ppm}$ ). We show that the observed long-term trends and seasonal cycles are in good agreement with the measurements from various other stations, especially the measurements from the Weybourne Atmospheric Observatory (United Kingdom). However, there are remarkable differences in the progression of annual trends between the Mace Head and Lutjewad records for  $\delta(\text{O}_2/\text{N}_2)$  and APO, which might in part be caused by sampling differences, but also by environmental effects, such as North Atlantic Ocean oxygen ventilation changes to which Mace Head is more sensitive. The Halley record shows clear trends and seasonality in  $\delta(\text{O}_2/\text{N}_2)$  and APO, the latter agreeing especially well with continuous measurements at the same location made by the University of East Anglia (UEA), while  $\text{CO}_2$  and  $\delta(\text{O}_2/\text{N}_2)$  present slight disagreements, most likely caused by small leakages during sampling. From our 2002–2018 records, we find a good agreement with Global Carbon Budget 2021 (Friedlingstein et al. (2021)) for the global ocean carbon sink:  $2.1 \pm 0.8 \text{ PgC yr}^{-1}$ , based on the Lutjewad record. The data presented in this work are available at <https://doi.org/10.18160/qq7d-t060> (Nguyen et al., 2021).

## 1 Introduction

The global carbon cycle is a dynamic system that comprises the exchanges of carbon between various reservoirs and is important for studying human-induced climate change and its impacts (Ciais et al., 2013). Accurate determination of anthropogenic  $\text{CO}_2$  emissions and their partitioning across different reservoirs plays a vital role in understanding the impact of the remaining atmospheric  $\text{CO}_2$  mole fraction on climate (Friedlingstein et al., 2020). High-precision atmospheric  $\text{O}_2$  measurements have been proven to be valuable in quantifying  $\text{CO}_2$  fluxes in the carbon cycle. By combining the decadal trends of atmospheric  $\text{CO}_2$  and  $\text{O}_2$ , we can quantify the global land and ocean carbon sinks (Bender et al., 1996; Keeling and Shertz, 1992; Manning and Keeling, 2006; Tohjima et al., 2019). This is because  $\text{CO}_2$  and  $\text{O}_2$  cycles are closely coupled – in most processes, there is an anti-correlation in the changes of their mole fraction, except for the oceanic uptake of  $\text{CO}_2$  (Manning and Keeling, 2006). To quantify the various components of the global carbon cycle, the changes in atmospheric mole fraction of the two species can be used in combination with their stoichiometric exchange ratio (ER), which is the ratio of  $\text{CO}_2$  and  $\text{O}_2$  exchanged (consumed / produced) in a process. The ER value varies depending on the process, and is close to 1.1 for photosynthesis / respiration (Severinghaus, 1995) and on average 1.38 for the global mix of fossil fuels (Keeling and Manning, 2014).

There are various techniques used to measure atmospheric  $\text{O}_2$  at high precision, such as interferometry (Keeling, 1988); mass spectrometry (Bender et al., 1994); paramagnetic analysis (Manning et al., 1999); gas chromatography (Tohjima, 2000); vacuum-UV absorption (Stephens et al., 2003, 2021); and fuel cell technology (Stephens et al., 2007). Despite many improvements to these techniques over the years, it is still very challenging to obtain  $\text{O}_2$  measurements with high accuracy and precision. This is mainly because the atmospheric background mole fraction of  $\text{O}_2$  is very high – around  $209\,392 \pm 3$  ppm (Tohjima et al., 2005) – while the observed variations are at the level of a few ppm. These challenges are magnified further for long-term measurements because of possible small biases, drifts, or other changes in the analysers or in the calibration scales. Thus, the sampling procedures and analysing (laboratory) conditions must be monitored and corrected for by a carefully designed use of calibration and reference gas cylinders over the years (Aoki et al., 2021). As a result, there are only a handful of programmes around the globe which are proficient in coupled  $\text{CO}_2$  and  $\text{O}_2$  measurements. These include the network of atmospheric stations maintained by the Scripps Institution of Oceanography (Manning and Keeling, 2006), the National Institute of Advanced Industrial Science and Technology (Aoki et al., 2021), the National Institute for Environmental Studies (Tohjima et al., 2008), Tohoku University (Goto et al., 2017), the University of East Anglia (UEA; Pickers et al., 2017),

the University of Groningen (van Der Laan-Luijkx et al., 2010), the National Center for Atmospheric Research, the Max Planck Institute for Biogeochemistry, and others. Our laboratory – the Centre for Isotope Research (CIO) of the University of Groningen (RUG) in the Netherlands – has been carrying out flask measurements of  $\text{CO}_2$  and  $\text{O}_2$  since the early 2000s at various locations (van Der Laan-Luijkx et al., 2010). Flask sampling for  $\text{CO}_2$  and  $\text{O}_2$  has been conducted at Lutjewad (the Netherlands), Mace Head (Ireland), Jungfrau-joch (Switzerland), and Halley (Antarctica).

In this paper, we present the  $\text{O}_2$  and  $\text{CO}_2$  measurements from flasks collected at Lutjewad (the Netherlands) and Mace Head (Ireland), both for the period 2000–2020, and Halley (Antarctica) for 2014–2017. From these measurements, a tracer called atmospheric potential oxygen (APO; the details of which are given in Sect. 2.5) is calculated. We first describe the measurement sites and the sampling procedure as well as the measurement methods, including the calibration procedure. Then we present the data and discuss the trends and seasonality, as well as the quality of the datasets. This paper builds on work previously presented in van Der Laan-Luijkx et al. (2010), Sirignano et al. (2010), and van Leeuwen (2015).

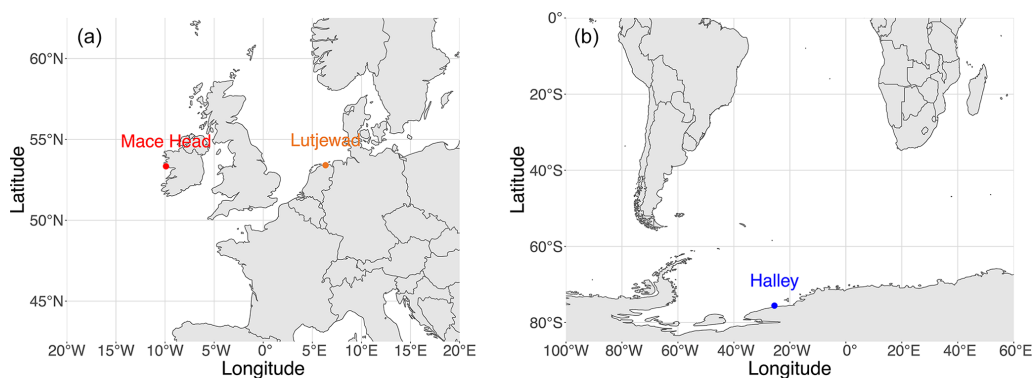
## 2 Methods

### 2.1 Site description

The stations from which our flasks were collected are Lutjewad Atmospheric Monitoring Station on the northern coast of the Netherlands ( $53^\circ 24' \text{ N}$ ,  $6^\circ 20' \text{ E}$ ) managed by the CIO (RUG); Mace Head Atmospheric Research Station on the western coast of Ireland ( $53^\circ 20' \text{ N}$ ,  $9^\circ 54' \text{ W}$ ) operated by the National University of Ireland's School of Physics and Ryan Institute Centre for Climate & Air Pollution studies; and Halley VI Research Station, at the time of the sampling situated on the Brunt Ice Shelf ( $75^\circ 34' \text{ S}$ ,  $25^\circ 30' \text{ W}$ ), operated by the British Antarctic Survey. Halley station was later relocated as that part of the ice shelf broke off. Figure 1 shows the locations of the three stations.

The Lutjewad station is a “class 2” station in the European Union's Integrated Carbon Observation System (ICOS) network. It comprises a 60 m tall tower, an additional platform of 10 m height, and a laboratory building containing analysers, flask sampling systems, measurement systems, and other equipment. The dominant wind direction in the Netherlands is south-west, meaning that the measurements acquired at the Lutjewad station often represent continental air masses influenced by anthropogenic and biogenic sources and sinks (van Der Laan et al., 2010). Otherwise, when the wind comes from the north, the station samples background air that comes from the North Sea and North Atlantic (van Der Laan-Luijkx et al., 2010).

The Mace Head station consists of field laboratories and a 20 m tower for sampling. The dominant wind arriving at



**Figure 1.** (a) Locations of the Mace Head (red) and Lutjewad (orange) stations. (b) Location of Halley station (blue).

the station is westerly, from the North Atlantic Ocean, carrying air masses that would not have been considerably affected by regional anthropogenic activities. Air masses from other directions carry contamination from local and continental sources (Derwent et al., 2002; Jennings et al., 1993).

Halley station is a “Global” station within the World Meteorological Organization’s Global Atmosphere Watch (WMO/GAW) programme that observes background atmospheric conditions at various locations around the globe. The main Halley station consists of eight modules that are atop ski-fitted hydraulic legs, within which the research facilities and living quarters are located. Air sampling for this project was carried out at the Clean Air Sector Laboratory, which is located 1.5 km from the main station at a location that receives minimal contamination from station activities (Jones et al., 2008). The predominant winds are from the east, bringing background air masses from the South Atlantic sector of the Southern Ocean (60 %) or from the continental plateaux (30 %). Westerly winds that have first passed over the Weddell Sea gyre occur 10 % of the time (Barningham, 2018; British Antarctic Survey, 2021).

## 2.2 Flask sampling procedure

At Lutjewad, we employ an automated flask sampling system, hereafter called the autosampler (Neubert et al., 2004). Air is pumped from the top of the 60 m tower via inlets connected to a series of tubing towards the laboratory building. The inlet is equipped with a Nafion drying tube (MD 110-72-S, Perma Pure, Toms River, New Jersey, USA) so that the incoming air is first partly dried. The flow in the outer side of the Nafion tube is the outlet of the same air sampling system, after the air is dried with the second stage cryogenic dryer in the laboratory to a dew point below  $-45^{\circ}\text{C}$  (Neubert et al., 2004). This ensures that, except for water, all constituents have a negligible gradient over the Nafion membrane. From the inlet, the sampled air is stored in glass flasks via a flask sampling system for further analyses in the CIO laboratories (Neubert et al., 2004). For storing air samples, we use 2.5 L glass flasks with dip tubes, capped with two

high-vacuum valves (Louwers, Hapert, NL) sealed with Viton o-rings (these flasks are also used at Mace Head and Halley). Our autosampler is designed to connect to and fill up to 20 flasks without requiring user intervention, and we can remotely control the opening/closing of the flask valves (via custom-made electric motor actuators) and the filling of samples (via a series including a small diaphragm pump, KNF N811; flow controllers; and magnetic solenoid valves). The autosampler schedule is controlled via custom-made software (written in the Delphi programming language), and carries out the sampling procedure automatically, but it can also be operated remotely using software such as VNC or TeamViewer when needed. A normal filling procedure starts with the air stream being cryogenically dried (to a dew point of  $-45^{\circ}\text{C}$ ) and flushed through a flask for at least 1 h at  $2.5\text{ L min}^{-1}$  before filling the flask slowly so that the sample remains at current atmospheric pressure (to prevent the sample from fractionation and differential permeation through the o-rings caused by a pressure gradient, Sturm et al., 2004) and moving to the next flask. Individual flasks can be preserved at any time. Samples at Lutjewad are collected under various conditions and time frequencies, but in this paper we present only the data from flasks collected under local background conditions, defined by van Der Laan-Luijkx et al. (2010) as flasks taken while the  $^{222}\text{Rn}$  activity monitored at the station was less than  $3\text{ Bq m}^{-3}$  and with a CO mole fraction of less than 200 ppb. This filtering procedure is applied to the dataset after the flasks are analysed.

We employ the same type of flasks, flow rates, and filling pressure (to current atmospheric pressure) at all stations. Due to the different set-up of the stations, the drying methods are different, and only Halley station has an aspirated inlet.

At Mace Head, flasks are collected once or twice per week via a manually operated system as described by Conway et al. (1994), at 35 m above sea level and mostly during restricted baseline conditions (Bousquet et al., 1996). A sampling sequence starts with the air being pumped from the inlet via a small diaphragm pump (KNF N86KT) into a drying tube packed with magnesium perchlorate; it is then flushed

through the flasks for about 30 min at  $2.5 \text{ L min}^{-1}$  at atmospheric pressure before each flask is manually closed. Also, for Mace Head, only flasks with a CO mole fraction of less than 200 ppb are retained.

At Halley, flasks are collected once per week depending on the meteorological conditions via a portable manual sampler. This consists of a diaphragm pump (KNF N86), flowmeter, drying agent (magnesium perchlorate),  $7 \mu\text{m}$  filter, and three sampling flasks connected in concession. The air is sampled about 6 m above the snow surface on the east side of the building via Synflex tubing connected to an aspirated inlet (the details of the aspirated inlet are as described by Blaine et al., 2006). The system is flushed for about 45 min at a flow rate of  $2.5 \text{ L min}^{-1}$  at atmospheric pressure before each flask is manually closed. The collected samples are stored in insulated aluminium boxes at room temperature until their annual return to the UK on a Antarctic supply ship.

After sample collection, flasks from the three stations are transferred back to our laboratory in Groningen for analysis. Typically, the mole fractions of  $\text{CH}_4$ , CO,  $\text{CO}_2$ , and  $\text{O}_2$  (reported as  $\delta(\text{O}_2/\text{N}_2)$ ; see next section) are measured (van Der Laan et al., 2009), and additional analyses such as those of stable isotopes (for example,  $^{13}\text{C}$  and  $^{18}\text{O}$  in  $\text{CO}_2$ ) and radiocarbon ( $^{14}\text{C}$  in  $\text{CO}_2$ ) are also conducted when required (van Der Laan et al., 2010).

### 2.3 $\text{CO}_2$ measurement

All flask samples are analysed on an Agilent HP6890N gas chromatograph (referred to as HPGC) equipped with a flame ionization detector to determine the mole fractions of  $\text{CO}_2$ , CO, and  $\text{CH}_4$ . The HPGC system has a set-up similar to the GC systems described by Worthy et al. (2003) and van Der Laan et al. (2009). All working standard mixtures (made from dried ambient air) that were used to calibrate the HPGC have been calibrated on the HPGC system at CIO against a suite of five primary standards linked to the World Meteorological Organization (WMO) X2007 scale, with  $\text{CO}_2$  ranging between 354 and  $426 \mu\text{mol mol}^{-1}$  (ppm). These primary standards were provided by the Earth System Research Laboratory (ESRL) of the National Oceanic and Atmospheric Administration (NOAA), USA. Since the summer of 2013, working standard gas cylinders were also calibrated for  $\text{CO}_2$ , CO, and  $\text{CH}_4$  mole fractions on a cavity ring-down spectrometer (CRDS), G2401-m from Picarro Inc., using the same suite of primary standards. We refer to Chen et al. (2010) for more details on the CRDS technique. The measurement precision and accuracy for flask measurements of  $\text{CO}_2$  on the HPGC are typically  $< 0.06$  ppm and  $< 0.07$  ppm respectively (van Leeuwen, 2015).

All  $\text{CO}_2$  measurements presented in this paper were originally calibrated against standards on the WMO X2007 scale, and are updated to the WMO X2019 scale (the new scale is explained in details by Hall et al., 2021).

### 2.4 $\text{O}_2$ measurement

Atmospheric  $\text{O}_2$  is typically reported as the  $\delta(\text{O}_2/\text{N}_2)$  value. The  $\delta(\text{O}_2/\text{N}_2)$  value of a sample is calculated as the difference between the  $\text{O}_2/\text{N}_2$  ratio of the sample and that of a reference gas (Keeling and Shertz, 1992):

$$\delta(\text{O}_2/\text{N}_2) = \frac{(\text{O}_2/\text{N}_2)_{\text{sample}} - (\text{O}_2/\text{N}_2)_{\text{reference}}}{(\text{O}_2/\text{N}_2)_{\text{reference}}}. \quad (1)$$

Since, for natural variations,  $\delta(\text{O}_2/\text{N}_2)$  values are very small, they are usually expressed in “per meg”, which is 1/1000 of a per mil, as typically used in the stable isotope community. Atmospheric  $\text{O}_2$  is reported as  $\text{O}_2/\text{N}_2$  ratio because it is not a trace gas, and its mole fraction is thus affected by changes in other atmospheric constituents such as  $\text{CO}_2$ . Atmospheric  $\text{N}_2$  is very stable (Keeling et al., 1998); therefore changes in the  $\text{O}_2/\text{N}_2$  ratio mostly reflect changes in atmospheric  $\text{O}_2$  (only in a detailed budget analysis are minor  $\text{N}_2$  variabilities still considered, as described in Keeling and Manning, 2014). For  $\delta(\text{O}_2/\text{N}_2)$  measurements, we use a Micromass Optima dual inlet isotope ratio mass spectrometer (DI-IRMS). The DI-IRMS analytical technique (which was first developed by Bender et al., 1994) follows the principles as explained by Keeling et al. (2004). Each measurement comprises 16 successive switches between sample and reference gases from the respective bellows. After every switch, the pressures of the two bellows are equalized, using a differential pressure meter (GA63, Effa France); subsequently there is an idle period of 120 s before the actual signal is measured for 30 s in order to account for the disturbances in the signals caused by the switching of the valves that affect measurement precision (Sirignano et al., 2010). Due to the sensitivity of the analyser, it is located inside a climate-controlled room in our CIO laboratory. However, it is inevitable that the measurements still drift over time. To correct for instrumental drifts, we perform frequent calibrations using a suite of reference gas cylinders. These cylinders are calibrated against the international Scripps scale using three primary standard cylinders purchased from the Scripps Institution of Oceanography (SIO), with  $\delta(\text{O}_2/\text{N}_2)$  values ranging from  $-792$  to  $-254$  per meg. Details of the extensive calibration procedure are thoroughly described by van Der Laan-Luijkx (2010) and van Der Laan-Luijkx et al. (2010), and are summarized in Sect. 3.

### 2.5 Atmospheric potential oxygen (APO)

Combining highly precise measurements of atmospheric  $\text{CO}_2$  and  $\text{O}_2$  can isolate the effects of the oceanic processes by removing the effects of the land biosphere (Stephens et al., 1998). This is achieved by deriving the tracer APO. The APO value of an air sample is determined by combining its  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$  measurements (Battle et al., 2006; Gru-

ber et al., 2001; Stephens et al., 1998):

$$\delta\text{APO} = \delta(\text{O}_2/\text{N}_2) + \frac{1.1 \times (\text{CO}_2 - 350)}{S_{\text{O}_2}}. \quad (2)$$

The value of 1.1 represents the mean  $\text{O}_2 : \text{CO}_2$  ER of terrestrial ecosystems (Severinghaus, 1995); for  $S_{\text{O}_2}$ , we take 0.2094, which is the standard atmospheric  $\text{O}_2$  mole fraction (Tohjima et al., 2005); and 350 is the consensus (arbitrary) reference value to be subtracted from the measured  $\text{CO}_2$  mole fraction, as defined in the SIO per meg scale conversion for APO (Manning and Keeling, 2006). Therefore, APO is not affected by land biosphere processes, and it mainly captures the seasonal and long-term air–sea exchange of  $\text{CO}_2$  and  $\text{O}_2$ , with an influence from fossil fuel combustion, caused by their higher average ER of  $\approx 1.4$  (Pickers et al., 2017; Sirignano et al., 2010).

### 3 Calibration of the DI-IRMS

In this section we present the calibration procedure and the stability achieved at our laboratory from 2006 to 2020. The calibration of the measurements taken in the 2000–2011 period and reported by van Der Laan-Luijkx et al. (2010) and van Der Laan-Luijkx et al. (2013) are kept intact, and the newly calibrated measurements from 2011 onwards are built on the principles of that work.

#### 3.1 The calibration procedure

The DI-IRMS compares the measurement of a sample gas with that of a reference gas (hereby called “machine reference” or “MREF”) in a sequence of several switches back and forth (“changeovers”). The result of this process is the  $\delta(\text{O}_2/\text{N}_2)$  value of the sample, as presented in Eq. (1). Each individual measurement is based on seven successive pairs of sample and reference measurements, which are used to calculate seven delta values (Eq. 1). The seven delta values then go through a filtering process. First, the mean and standard deviation of the seven delta values are calculated. Then, the delta value that is furthest from the mean is marked as a potential outlier. Next, a new mean and a new standard deviation are calculated for the remaining six delta values. If the excluded delta value is more than 2.7 times (equivalent to  $p = 0.01$ ) the new standard deviation away from the new mean, it is defined as an outlier and removed. This process is repeated to identify and remove a potential second outlier (at most two outliers are removed by this process, otherwise the reliability of the measurement is sacrificed). After removing possible outliers, the remaining delta values are averaged to produce one  $\delta(\text{O}_2/\text{N}_2)$  value per measurement. A flask is typically measured two to three times consecutively, for which we do not find any systematic biases. The final measurement for each flask (as presented in this paper) is the average of the filtered  $\delta(\text{O}_2/\text{N}_2)$  values of these repeated

measurements (van Der Laan-Luijkx et al., 2010). The precision of the DI-IRMS for flask measurements varies between 7 and 12 per meg, based on the averaged standard deviation of all flask measurements at Lutjewad and Mace Head flasks respectively.

To improve the stability of our measurements, we also measure local reference gas cylinders (hereafter called “working tank” or “WT”) on the sample side of the DI-IRMS. These WTs are also used to connect between periods of different MREF cylinders, where there may be shifts in the scales of the measurements and thus a scale conversion is required to keep all raw measurements on a comparable scale. A summary of different WTs and MREF cylinders used from 1998 to 2020 is shown in Fig. 2 and Table 1.

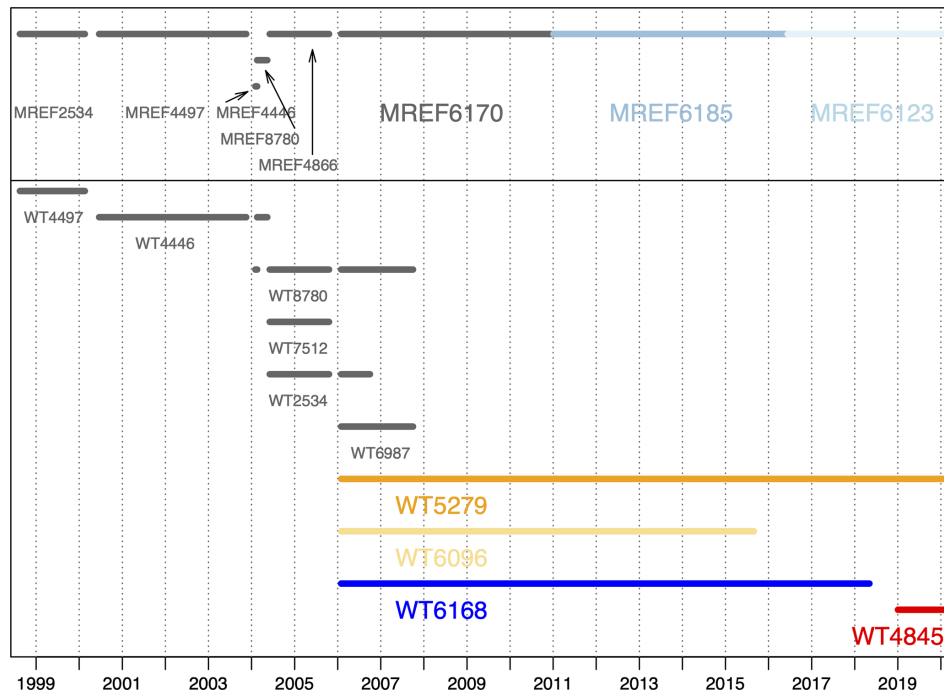
To connect the different MREF periods, we first convert all raw measurements (which are the ratios of the raw values to their respective MREF) to our internal 2534 CIO scale. Subsequently, they are converted to the SIO scale. Cylinder number 2534 has been chosen as the baseline for our internal reference scale, because it was the first MREF gas in 1998 and later on was measured as a WT against several other MREF cylinders (Fig. 2). When converting the measurements to the internal CIO scale, we need to take into account the “zero-enrichment” factor: measurements of a WT (on the sample side) against an MREF cylinder (on the reference side) do not produce the same value as when they are measured the other way around (van Der Laan-Luijkx et al., 2010).

In addition to the conversion to our internal CIO scale, the measurements are also affected by instrumental drifts over time. To correct for these drifts, we first divide our long measurement record into several periods, which are defined based on the timing of when the MREF cylinders are changed, and/or apparent fluctuations in the raw data related to, for example, repairs or modifications of the system. In this work, the calibration procedure is carried out for measurements from 2011 onwards, which were divided into seven periods (periods 9–15, Table 1).

These 7 periods were divided into 144 sub-periods (selected based on breaks in the records) which were then individually processed to derive the final corrections for all measurements in those sub-periods. The complete step in transforming the raw measurements of a sample ( $S$ ) against a current MREF ( $M$ ) into comparable data is to combine the drift correction with the shift to the CIO scale ( $R$ ), by using an equation described by van Der Laan-Luijkx (2010).

$$\delta_{S/R} = \left( (\delta_{M/R})_{\text{sub-period}} + \text{drift} \times \frac{\text{days}}{365} + 1 \right) \times (\delta_{S/M} + 1) - 1 \quad (3)$$

Here,  $\delta_{S/R}$  is the  $\delta(\text{O}_2/\text{N}_2)$  value of the sample against the CIO 2534 scale;  $(\delta_{M/R})_{\text{sub-period}}$  is the average  $\delta(\text{O}_2/\text{N}_2)$  value of the MREF cylinder against the CIO scale in a sub-period calculated based on the measurements of all WTs in that sub-period; drift is the average drift per day in a sub-



**Figure 2.** Summary of the different WTs and MREF cylinders in the 1998–2020 period. MREFs are shown along the top, with WTs below. In the case of WTs, there is typically overlap between more than one WT. Periods in grey are adapted from the work of van Der Laan-Luijkx (2010).

**Table 1.** Summary of the calibration periods defined in this paper and the corresponding MREF cylinder and WT cylinder numbers most recently used for the calibration of the DI-IRMS. The greyed-out rows are the cylinders used prior to this work, but are included here to demonstrate a complete record.

	Period	MREF	WTs
Previous	1 17 Aug 1998–18 Feb 2000	2534	4497
	2 19 Jun 2000–17 Nov 2003	4497	4446
	3 3–18 Feb 2004	4446	8780
	4 18 Feb–14 May 2004	8780	4446
	5 4 Jun 2004–19 Oct 2005	4866	2534   7512   8780
	6 30 Jan–30 Dec 2006	6170	2534   6987
	7 30 Jan–30 Dec 2007	6170	5279   6096   6987
	8 30 Jan 2008–15 Dec 2010	6170	5279   6096   6168
Current	9 3 Jan 2011–11 Mar 2014	6185	5279   6096   6168
	10 11 Mar 2014–29 Aug 2015		5279   6096   6168
	11 30 Aug 2015–10 Jun 2016		5279   6168
	12 11 Jun 2016–5 May 2018	6123	5279   6168
	13 6 May 2018–1 Jan 2019		5279
	14 2 Jan 2019–11 Mar 2020		5279   4845
	15 12 Mar 2020 to present		5279   4845

period (if any), calculated based on the WT values; and days is the number of days at the time of the sample since the start of the sub-period.

$\delta_{S/M}$  is the  $\delta(\text{O}_2/\text{N}_2)$  value of sample against the MREF cylinder (raw value).

The final step is to transform the  $\delta_{S/R}$  value of a sample onto the SIO scale via a linear conversion (shown in Sect. 3.2) using the values of the Scripps primary cylinders measured against the CIO scale. For an extensive and detailed explanation on how to calculate each component of Eq. (3), we refer to van Der Laan-Luijkx (2010). Figure 3

shows the results for the WTs of the new calibration procedure connected to the previously reported data by van Der Laan-Luijkx et al. (2010).

After these adjustments, the measurements of the three long-term WTs (5279, 6096, and 6168) show that all three were simultaneously stable over time. To verify this, we calculated the trends of all three WTs based on their annual averages; the weighted mean slope amounts to  $-0.4 \pm 0.7$  per meg  $\text{yr}^{-1}$ , thus not significantly different from zero (Fig. 4). In addition to this outcome, we calculated the year-to-year variability of the WTs, visible in Fig. 4 as the scatter of the points around the trend lines. We calculated the standard deviations of this scatter (the residuals) around these trend lines, and all three standard deviations were between 2.4 and 4.0 per meg. Therefore, we state our year-to-year stability as being 3 per meg over the 11 years of measurements.

WT 4845 was recently measured for a relatively short period only, and it appeared to be less stable and noisier than WT 5279 measured in the same period. It is not clear why this is the case, but it could be due to the fact that the value of this cylinder is very low, suggesting potential contamination when the cylinder was filled, or a leak in the pressure reducer when it was measured. Thus, WT4845 was not used for the calculations in the calibration procedure, and its measurements are only shown here for completeness.

In addition to their long-term stability, the three WTs also showed no systematic drifts across different MREF periods (Table 2). For WT 5279 and WT 6096, there were no significant changes (at least to  $\pm 0.3$  per meg) between the MREF periods 6170 and 6185, although there was a small decrease of 4.0 per meg in the mean measurement of WT 5279 in MREF 6123 period. For WT 6168, the mean value increased by 3.6 per meg from MREF 6170 to MREF 6185 period, and then dropped slightly (by 0.5 per meg) in MREF 6123 period. The stability demonstrated in both long-term measurements and for each MREF period confirms the quality of our calibration procedure.

### 3.2 Quality check of the Scripps primary cylinders

The final check on the quality of our scale is the regular measurement of the three Scripps primary standard cylinders that we purchased from SIO, numbered 7002, 7003, and 7008. These measurements were conducted at least once per year or when there was an additional need for recalibrating, e.g. after instrument failure or upgrade. Each measurement period took a different amount of time – some measurements were spread over a couple of days while others were repeated over (or after) a few weeks. From 2007 to 2018, 16 measurement periods were conducted (Fig. 5). The large gap between 2011 and 2014 was due to a lack of funding, and thus of personnel, leading to the situation that the laboratory was understaffed and we could not keep up measurement of the primary tanks.

In Fig. 5, each data point is the mean value over each measurement period and the error bars are the standard deviations. The coloured lines are the overall linear fit of the measured values of the corresponding cylinders (and their associated  $2\sigma$  uncertainties), and the black horizontal lines are the assigned values of the cylinders (determined by the SIO, updated in 2020). The assigned and measured values of the primary standard cylinders over the whole period are compared in Table 3. The measured values are the weighted means of each cylinder, since each data point is calculated based on different numbers of separate measurements. It can be seen from Fig. 5 and Table 3 that cylinder 7008 exhibits a small upward drift over time of  $1.4 \pm 0.4$  per meg  $\text{yr}^{-1}$ , whereas the other two remain constant. The ensemble thus suggests that there is no clear systematic error in our scale conversion and calibration procedure. Overall, the SIO primary standards produce a weighted uncertainty of 8.6 per meg in 10 years. To improve the quality of our conversion into the SIO scale, and especially to check the behaviour of cylinder 7008, we are planning to purchase new primary standard cylinders in the future.

The conversion of the CIO scale to the Scripps scale is calibrated using the Scripps primary standards for measurements, and in such a way that the ensemble difference between the assigned values and weighted averages of our measurements of three Scripps cylinders is minimized (Mook, 2000).

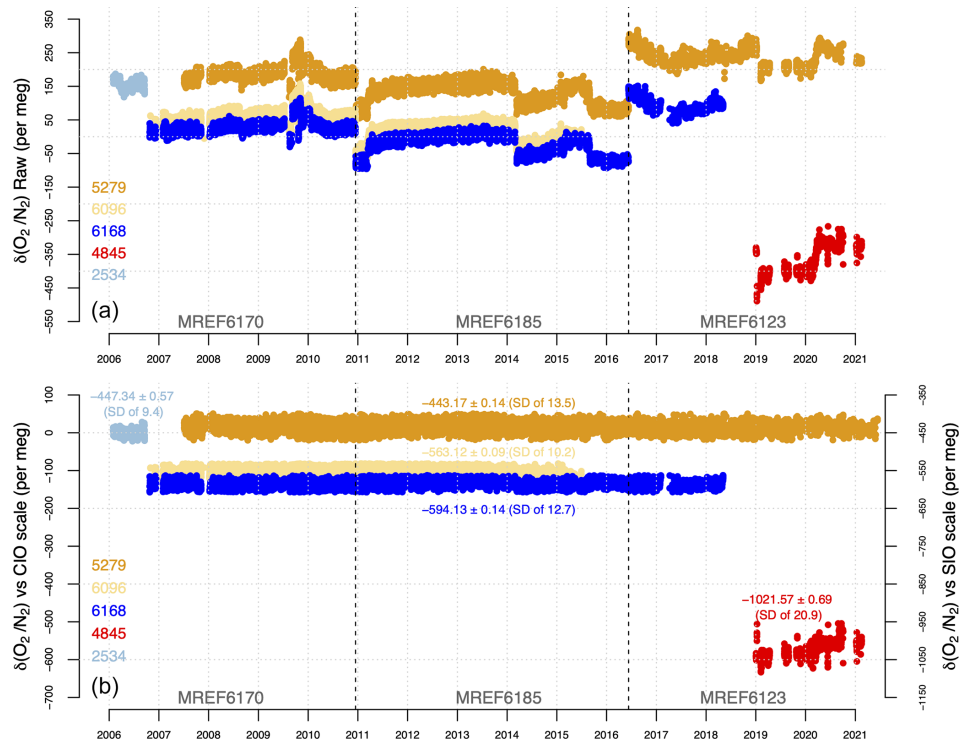
$$\delta(\text{O}_2/\text{N}_2)_{\text{SIO}} = \delta(\text{O}_2/\text{N}_2)_{\text{CIO}} \cdot 0.999544 + (0.999544 - 1) \times 10^6 + 1.4 \quad (4)$$

Here,  $\delta(\text{O}_2/\text{N}_2)_{\text{SIO}}$  and  $\delta(\text{O}_2/\text{N}_2)_{\text{CIO}}$  are the  $\delta(\text{O}_2/\text{N}_2)$  values of the SIO and CIO scales respectively; 0.999544 is the slope with an uncertainty of 0.000008; and 1.4 per meg is the weighted mean offset of the three Scripps primary standards with an uncertainty of 5 per meg (which is thus zero within its uncertainty, as it should be).

### 3.3 Inter-comparison programmes

In addition to measuring the primary standard cylinders, the CIO also took part in two inter-comparison programmes involving oxygen measurements: “Cucumber” Intercomparison, which was initialized in the European Union’s CarboEurope project and coordinated by the UEA (<http://cucumbers.uea.ac.uk/>, last access: 24 February 2021), and the Global Oxygen Laboratories Link Ultra-precise Measurements (GOLLUM) programme, also coordinated by UEA (Manning et al., 2015). These inter-comparison programmes provide an additional tool for checking the internal stability of our measurements, while also linking the oxygen measurements between global laboratories.

The Cucumber programme involves inter-comparison of nine atmospheric species (of which  $\delta(\text{O}_2/\text{N}_2)$  is one) between atmospheric research stations in Europe and a num-



**Figure 3.** Measurements of the three long-term WTs (5279, 6096, and 6168) for periods 7–15 (Table 1), across the final three MREF periods plus a recently added WT (4845). **(a)** Raw measurements of the WTs against different MREF cylinders. **(b)** Measurements of the WTs calibrated and converted to the CIO scale (left y axis) and against the SIO scale (right y axis). The values on the plot are the corresponding long-term means and 1 $\sigma$  standard errors of the WTs against the SIO scale, and the respective standard deviations are shown in parentheses. All numbers are in per meg. Visible gaps in the data are due to instrument issues, maintenance, or instrument relocation.

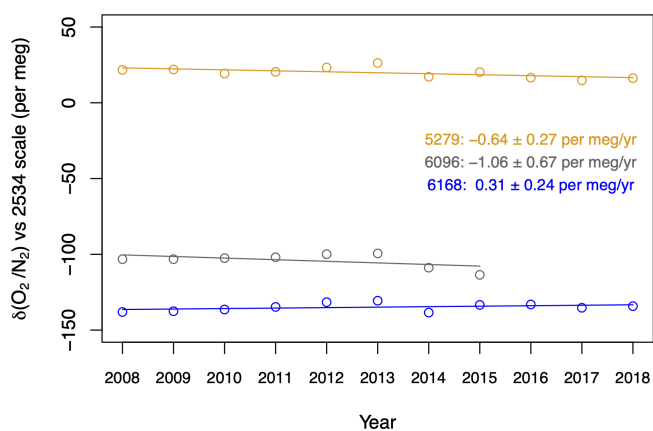
**Table 2.** Comparison of the WTs over three different MREF cylinder periods. The values (in per meg) are averaged over the corresponding period, accompanied by the standard errors. The n/a (not applicable) values in the MREF6123 period for WT6096 are due to its discontinuation in this period. The difference column values are calculated by subtracting the values of the old MREF periods from the new ones

		CIO scale	SIO scale	Difference (CIO scale)
WT 5279	MREF 6170	20.9 $\pm$ 0.2	-438.2 $\pm$ 0.2	
	MREF 6185	20.7 $\pm$ 0.3	-438.4 $\pm$ 0.2	-0.2
	MREF 6123	16.7 $\pm$ 0.2	-442.3 $\pm$ 0.2	-4.0
WT 6096	MREF 6170	-103.4 $\pm$ 0.1	-556.9 $\pm$ 0.1	
	MREF 6185	-103.1 $\pm$ 0.1	-556.7 $\pm$ 0.1	+0.3
	MREF 6123	n/a	n/a	
WT 6168	MREF 6170	-137.3 $\pm$ 0.2	-589.3 $\pm$ 0.2	
	MREF 6185	-133.7 $\pm$ 0.2	-585.9 $\pm$ 0.2	+3.6
	MREF 6123	-134.2 $\pm$ 0.4	-586.4 $\pm$ 0.3	-0.5

ber of laboratories in Europe, USA, Canada, Japan, and Australia. Within the programme, there are seven sets of three cylinders sent around in different rotations. The CIO participated in three rotations, with two involving oxygen measurements (called “Inter-1” and “Euro-3”; University of East Anglia, 2021).

The GOLLUM programme is specifically designed for the inter-comparison of oxygen measurements and involves 10 laboratories worldwide that carry out high-precision atmospheric oxygen measurements. Two sets (named “Bilbo” and “Frodo”) of three cylinders are rotated in opposite directions amongst participating laboratories (Manning et al., 2015).





**Figure 4.** Annual averages of WTs 5279, 6096, and 6168 (points) from 2008 to 2018, against the CIO scale. The fitted trends (lines) and the values of their slopes are also plotted.

**Table 3.** Comparison of the averaged measured values of the Scripps primary standards against their assigned values in per meg.

Cylinder ID	7008	7002	7003
Assigned by SIO	-254.3	-465.0	-791.6
Measured weighted mean	-247.3	-470.3	-799.0
Standard deviation	8.0	6.0	11.8
Standard error	1.9	2.0	2.8
Deviation from assigned	6.9	-5.3	-7.4

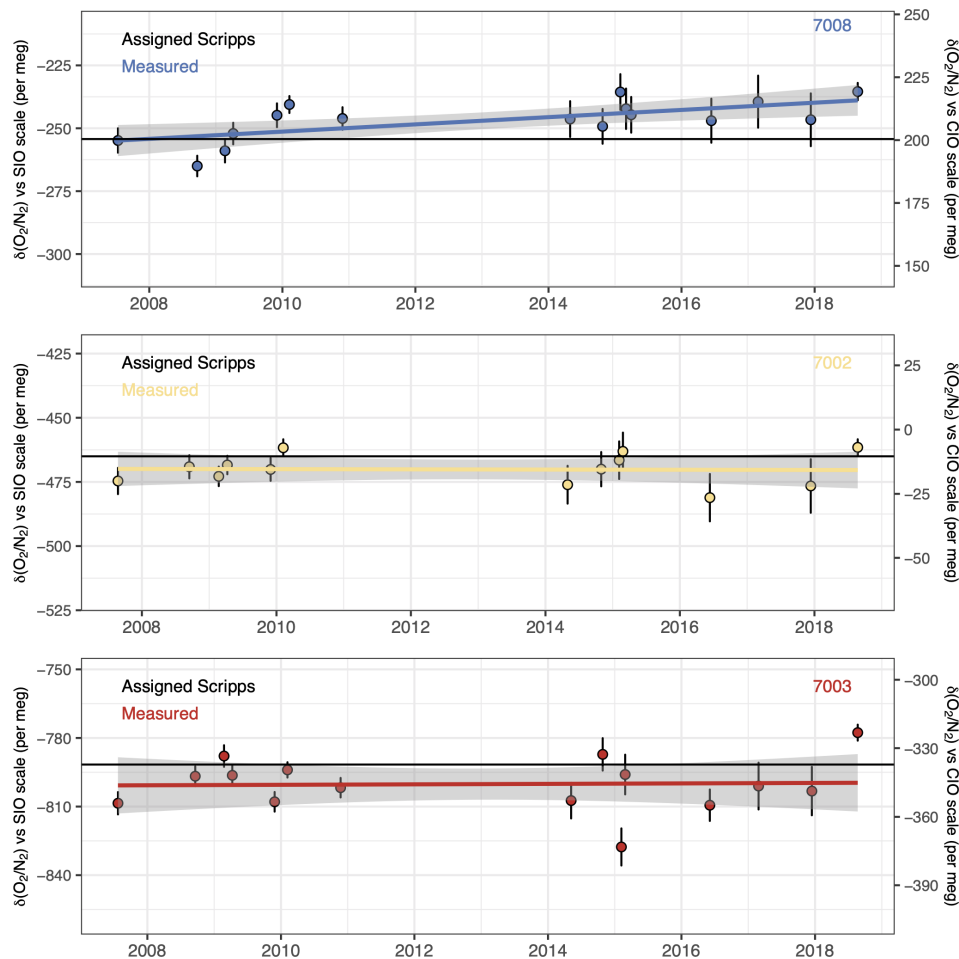
Figure 6 shows the measurements of the Cucumber cylinders (panels a and b), the cylinders in the Bilbo and Frodo rotations of GOLLUM (panels c and d respectively), and the measurements of three internal cylinders at CIO: the working tanks 5279, 6096, and 6168 along with the SIO primary standard cylinder 7008 (panel e). The measurements of the cylinders in the Inter-1 and Euro-3 rotations are plotted as the difference between the measured values of the cylinders against their own assigned values as originally measured at the Max Planck Institute for Biogeochemistry in Germany in January 2008. These results show that the cylinders in the Inter-1 and Euro-3 rotations were quite variable over time (varying within a range of less than 30 per meg) but in different directions and of different size, suggesting that there is no systematic scaling error but rather individual variations between cylinders and/or measurement periods. Due to the individual variations, the overall drifts for the Cucumber cylinders are  $11 \pm 18$  per meg  $\text{yr}^{-1}$ , significantly higher than the WMO network compatibility goal of 2 per meg (World Meteorological Organization, 2018). The lower quality of the measurements (not only in our laboratory) might well be connected to the fact that these cylinders are not part of a dedicated oxygen comparison programme, so the treatment of the cylinders (for example, vertical storage and unsuitable pressure reducers) are not of high enough standard for oxygen.

For GOLLUM cylinders, all measurements are also plotted as the difference between the measured values of the cylinders and their assigned values on the SIO scale. The assigned values for Bilbo, Frodo, and SIO cylinders are determined at the SIO, while those for the WTs are their averaged long-term value measured at CIO on the SIO scale. Compared to the Cucumber cylinders, GOLLUM cylinders show much less variation between the years (varying within a range of less than 20 per meg), and also significantly smaller overall drift over the duration of the measurements ( $4 \pm 6$  per meg  $\text{yr}^{-1}$ ). However, all six cylinders appear to drift in a similar direction, suggesting a significant drift in our scale rather than drifts in these cylinders. The SIO cylinder 7008 also shows similar stability and a general drift in the same direction as the GOLLUM cylinders, whereas the two other SIO cylinder do not (Fig. 5).

INTER-1 and EURO-3 do not show an apparent drift direction; Bilbo and Frodo present a minor drift similarly to that observed by our SIO cylinder 7008 (while the other two SIO cylinders did not exhibit this behaviour, as shown in Sect. 3.2); and our internal WTs all show no overall drifts. Since the cylinders show an inconclusive “drift”, we consider our calibration procedure as sufficient. Recalibration of the SIO cylinders might shed further light on these small discrepancies, mostly to see if cylinder 7008 has indeed drifted or not.

### 3.4 Treatment of analysed flask samples

After the calibration and conversion to the SIO scale, the individual flask sample measurements are scrutinized for outliers and background conditions. For this purpose, we perform several iterations of fitting a combination of quadratic and three-harmonic regression (following similar curve fitting methods applied to time series in NOAA without the use of a digital filtering method; Thoning et al., 1989) and filtering the outliers from the combined fit. This outlier filtering process uses the robust median absolute deviation (MAD) method (Rousseeuw and Verboven, 2002), in which the MAD value for a dataset is determined by first finding the median of the set, then subtracting the median from each individual value, and finally finding the median of the absolute differences. Measurements that are 3 times the MAD value away from the median of the measurement set are considered outliers and removed. The full principle of the procedure is described by van Der Laan-Luijkx (2010; though with a different filtering process that was described in Sect. 3.1). In total, after both filtering processes, we excluded from further analysis around 30 % of the flasks from the Lutjewad samples, 16 % from the Mace Head samples, and only 6 % from the Halley samples. The larger fraction of discarded measurements in the Lutjewad record is related to the sampling process, where we do not specifically only sample air at background conditions, which is the case at Mace Head. For Halley, since it is by design a background station, there



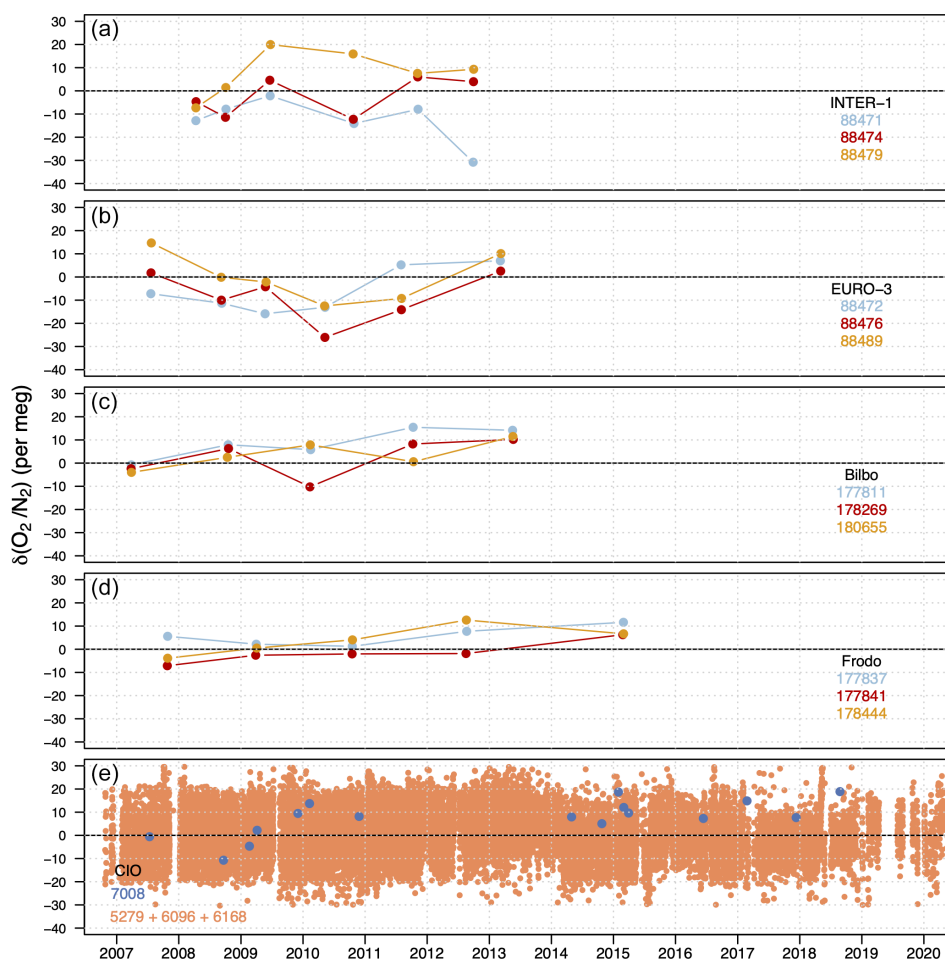
**Figure 5.** Scripps primary standard cylinder measurements over time. Each point is the averaged value over a measurement period. Error bars represent  $1\sigma$  standard deviations. Solid horizontal lines are the assigned values (black) and the linear least squares fit to the data (coloured) of each cylinder. The grey shading indicates the 95 % confidence interval uncertainties of the values.

are hardly any local sources and sinks, and the wind coming from the continental plateaux only accounts for 30 % of the total. The 6 % outlier fraction for Halley is a good indication of the fraction of actually failed sampling and/or analysis. The APO values of all stations are calculated from  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$  measurements (Eq. 2), when there is information on both species for each flask sample.

In the period prior to 2006, our internal calibration scale was not as well-established as in the later period, due to frequent changes in MREF and WT cylinders, especially in 2004, for which there is little information to connect the following period to the first period (as presented in Fig. 2 and Table 1). In addition, we also only obtained the SIO primary standards in late 2007, so all earlier measurements cannot be directly linked to the SIO scale and have to be converted via the internal CIO scale. The results of this quality check prompt us to exclude the first 2 years from the fits of Lutjewad and Mace Head data so that they are less affected by the problematic period. The last 2 years are also excluded, partly

because flask sampling was relatively sparse in those years and this could also introduce biases in the fits, and also because in the period of late 2019 through the entirety of 2020, our DI-IRMS experienced detrimental problems that affected the quality of the measurements. After several tests, we decided to establish our fits for Lutjewad and Mace Head based on the years 2002 to 2018.

In summary, in our 20 years of measurements, we have observed an uncertainty of flask measurements of 7 to 12 per meg (based on the averaged standard deviations of the individual flasks collected from Lutjewad and Mace Head), and we have maintained the stability of our internal scale (3 per meg over 11 years) and the Scripps primary standards (8.6 per meg over 10 years). Although some drift is observed in one of our Scripps cylinders, the other two have remained stable within the uncertainty. The same inconclusive picture emerges from our various sets of cylinders in the inter-comparison programmes. Therefore, we conclude that



**Figure 6.** Cylinders from the Cucumber programme (a, b) along with two sets of three cylinders in the GOLLUM programme (c, d) and three internal CIO cylinders (WT 5279, WT 6096, and WT 6168) and a primary standard cylinder at CIO (SIO 7008) (e). Each colour represents a different cylinder, and the legends show the corresponding cylinder IDs. The points are the measurements of the cylinders over time, plotted as the difference from their assigned values. For the Cucumber, GOLLUM, and SIO cylinders, the assigned values are determined at the SIO, and for the WTs, the assigned value is its long-term average measured at CIO on the SIO scale. y axis ranges are identical for all panels.

our calibration process is accurate within the uncertainties mentioned above.

## 4 Flask measurement results

### 4.1 $\text{CO}_2$ , $\delta(\text{O}_2/\text{N}_2)$ , and APO records

In this section, we present the long-term flask measurement records (from 2000 to 2020) of Lutjewad and Mace Head, along with a 3-year record from Halley. In general, Lutjewad and Mace Head show similar patterns for  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$ , with some differences in APO variations. Figures 6 to 8 show the  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO measurements for Lutjewad, Mace Head, and Halley respectively. The black points illustrate the final, filtered flask measurement values; the coloured lines are the total fit (combined quadratic trend and three-harmonic seasonal cycles) and the black lines are the trend parts of the total fit. The fit lines are shown for the whole

period, but for the fitting process, we left out the first and last 2 years to make sure that the fit period comprises complete calendar years (from January to December). Otherwise, the beginning and end of the curves can influence the trend part of the fit due to the irregular sampling frequency and other problems, as explained in Sect. 3.4. From the records, the total uncertainties associated with the trends are also calculated, based on a quadratic sum of the uncertainties of the flask measurements and other factors. For  $\text{CO}_2$ , the only other contributing factor is the uncertainty in the trend fit. For  $\delta(\text{O}_2/\text{N}_2)$ , and APO, the uncertainties associated with the measurements of the SIO primary standards, our internal scale, the long-term scale conversion between CIO and SIO scales, and the trend fits all contributed to the final uncertainty.

$\text{CO}_2$  measurements at Lutjewad and Mace Head show a positive and increasing trend over 17 years. Due to

the quadratic trend fit, the growth of the fitted increase is linear. The trend (given here in  $\text{ppm yr}^{-1}$  with 95 % confidence interval, CI, uncertainties) at Lutjewad grows from  $1.81 \pm 0.10 \text{ ppm yr}^{-1}$  in 2002 to  $2.27 \pm 0.03 \text{ ppm yr}^{-1}$  in 2010 and  $2.74 \pm 0.10 \text{ ppm yr}^{-1}$  in 2018. These values agree relatively well with the globally averaged values as measured by the NOAA Global Monitoring Laboratory:  $1.86 \pm 0.20 \text{ ppm yr}^{-1}$  in 2002,  $1.97 \pm 0.14 \text{ ppm yr}^{-1}$  in 2010, and  $2.57 \pm 0.19 \text{ ppm yr}^{-1}$  in 2018 (<https://gml.noaa.gov/ccgg/trends/global.html>, last access: 2 April 2021). The values from NOAA are calculated based on a 5-year average around the time points 2002, 2010, and 2018. In all three periods, the values at Mace Head are also in agreement with those of Lutjewad ( $1.86 \pm 0.06$  in 2002,  $2.24 \pm 0.02$  in 2010, and  $2.63 \pm 0.06 \text{ ppm yr}^{-1}$  in 2018 for Mace Head). When averaging the trends over the 17-year period, both stations show good agreement with each other and with the global average:  $2.31 \pm 0.07 \text{ ppm yr}^{-1}$  for Lutjewad,  $2.22 \pm 0.04 \text{ ppm yr}^{-1}$  for Mace Head, and  $2.1 \pm 0.3 \text{ ppm yr}^{-1}$  for the global average. The total uncertainty of the trend is  $0.07 \text{ ppm yr}^{-1}$  for Lutjewad and  $0.04 \text{ ppm yr}^{-1}$  for Mace Head. The largest contributing factor to the total  $\text{CO}_2$  long-term trend uncertainty is from the trend fits.

$\delta(\text{O}_2/\text{N}_2)$  measurements at Lutjewad also show a clear trend that becomes increasingly more negative throughout the 20 years. The trends (reported here in  $\text{per meg yr}^{-1}$  with 95 % CI uncertainties) in 2002, 2010, and 2018 are  $-18.01 \pm 1.17$ ,  $-20.99 \pm 0.29$ , and  $-23.98 \pm 1.17 \text{ per meg yr}^{-1}$  respectively. At Mace Head, we find an unexpected trend: while the trend in  $\text{CO}_2$  increases, that of  $\delta(\text{O}_2/\text{N}_2)$  becomes less negative ( $-22.4 \pm 1.3$ ,  $-21.2 \pm 0.3$ , and  $-20.0 \pm 1.3 \text{ per meg yr}^{-1}$  in 2002, 2010, and 2018 respectively), which is contrary to the expectations of an increasingly negative trend, based on increased fossil fuel consumption over the years, and also different from the measurements at Lutjewad. The lower number of flask samples from Mace Head between 2017 and 2019 makes it difficult to accurately interpret the cause of this change in the trend, and it also affects the determination of a proper fit through the period, potentially leading to inaccuracies in the long-term trend. When averaged over the entire period, however, both stations show almost identical trends:  $-21.2 \pm 0.8 \text{ per meg yr}^{-1}$  for Lutjewad and  $-21.3 \pm 0.9 \text{ per meg yr}^{-1}$  for Mace Head. The total uncertainty of the trend is  $1.3 \text{ per meg yr}^{-1}$  for Lutjewad and  $1.5 \text{ per meg yr}^{-1}$  for Mace Head. The largest contributing factor to the total  $\delta(\text{O}_2/\text{N}_2)$  long-term trend uncertainty for Lutjewad and Mace Head is the uncertainty in the trend fits, with a small effect from the scale stability (of 3 per meg in 11 years). However, at Mace Head the uncertainties in the flask measurements contributed more significantly than those at Lutjewad (12.5 compared to 7.4 per meg respectively).

The APO trend and seasonality can be determined either from fitting the APO values of the individual flasks themselves, or by combining the trend/seasonal

parameters of the  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$  fits. Both methods yield almost identical results. We present here the results from the first approach. Since APO is calculated from the combination of  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$  measurements, it shows a combination of the patterns as illustrated in the two species. The APO trend (also reported here in  $\text{per meg yr}^{-1}$ ) at Lutjewad does not differ significantly over time, varying from  $-9.4 \pm 0.8 \text{ per meg yr}^{-1}$  in 2002 to  $-9.31 \pm 0.20 \text{ per meg yr}^{-1}$  in 2010, and to  $-9.3 \pm 0.8 \text{ per meg yr}^{-1}$  in 2018. At Mace Head, however, the same pattern as  $\delta(\text{O}_2/\text{N}_2)$  is shown for APO: the trend becomes significantly less negative throughout the period ( $-13.15 \pm 1.20 \text{ per meg yr}^{-1}$  in 2002,  $-9.5 \pm 0.3 \text{ per meg yr}^{-1}$  in 2010, and  $-5.83 \pm 1.20 \text{ per meg yr}^{-1}$  in 2018). The total uncertainty of the trend is  $1.0 \text{ per meg yr}^{-1}$  for Lutjewad and  $1.3 \text{ per meg yr}^{-1}$  for Mace Head, and the largest contributing factors are the same as for  $\delta(\text{O}_2/\text{N}_2)$ .

Similarly to Lutjewad and Mace Head, at Halley station,  $\text{CO}_2$  increases over time while  $\delta(\text{O}_2/\text{N}_2)$  decreases, with much less variability in  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$  measurements, due to the absence of terrestrial biosphere influence. The averaged  $\text{CO}_2$  trend at Halley from 2014 to 2017 is  $2.60 \pm 0.20 \text{ ppm yr}^{-1}$ , similar to the trends at Lutjewad and Mace Head in the same period ( $2.62 \pm 0.08 \text{ ppm yr}^{-1}$  and  $2.53 \pm 0.05 \text{ ppm yr}^{-1}$  respectively). On the other hand,  $\delta(\text{O}_2/\text{N}_2)$  and APO trends at Halley are significantly smaller in size than those at Lutjewad and Mace Head. The  $\delta(\text{O}_2/\text{N}_2)$  trend at Halley over the 2014–2017 period is  $-15 \pm 3 \text{ per meg yr}^{-1}$ , while at Lutjewad and Mace Head, the trends are  $-23.2 \pm 0.9 \text{ per meg yr}^{-1}$  and  $-20.3 \pm 1.0 \text{ per meg yr}^{-1}$  respectively. For APO, the corresponding values are respectively  $-1.4 \pm 2.4$ ,  $-9.3 \pm 0.6$ , and  $-6.7 \pm 0.9 \text{ per meg yr}^{-1}$  for Halley, Lutjewad, and Mace Head.

## 4.2 Seasonal cycles

The seasonal cycles of  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO for all three stations are presented in Fig. 10. The seasonal components are extracted from the total fits (detrended) and presented as 1-year cycles. In general, the  $\text{CO}_2$  seasonal cycles at Lutjewad and Mace Head are similar in size and shape, although the average seasonal amplitude is higher at Lutjewad ( $16.8 \pm 0.5 \text{ ppm}$ ) than Mace Head ( $14.8 \pm 0.3 \text{ ppm}$ ). The  $\text{CO}_2$  seasonal cycle at Halley station, on the other hand, has a much smaller amplitude of  $3.0 \pm 0.3 \text{ ppm}$ , as is generally the case for the ocean-dominated Southern Hemisphere due to the absence of terrestrial biosphere influence. Lutjewad and Mace Head show very similar, and significantly higher,  $\delta(\text{O}_2/\text{N}_2)$  seasonal amplitudes ( $131 \pm 6 \text{ per meg}$  and  $130 \pm 6 \text{ per meg}$  respectively) than that at Halley ( $76 \pm 4 \text{ per meg}$ ), due to the influence of the terrestrial biosphere. In APO this influence is cancelled because APO is invariant to terrestrial biosphere processes, and the Halley

amplitude is even somewhat higher than that of Lutjewad and Mace Head ( $65 \pm 3$  per meg compared to  $54 \pm 4$  and  $61 \pm 5$  per meg respectively). All numerical seasonality parameters of the three stations are given in Table 4.

## 5 Discussion

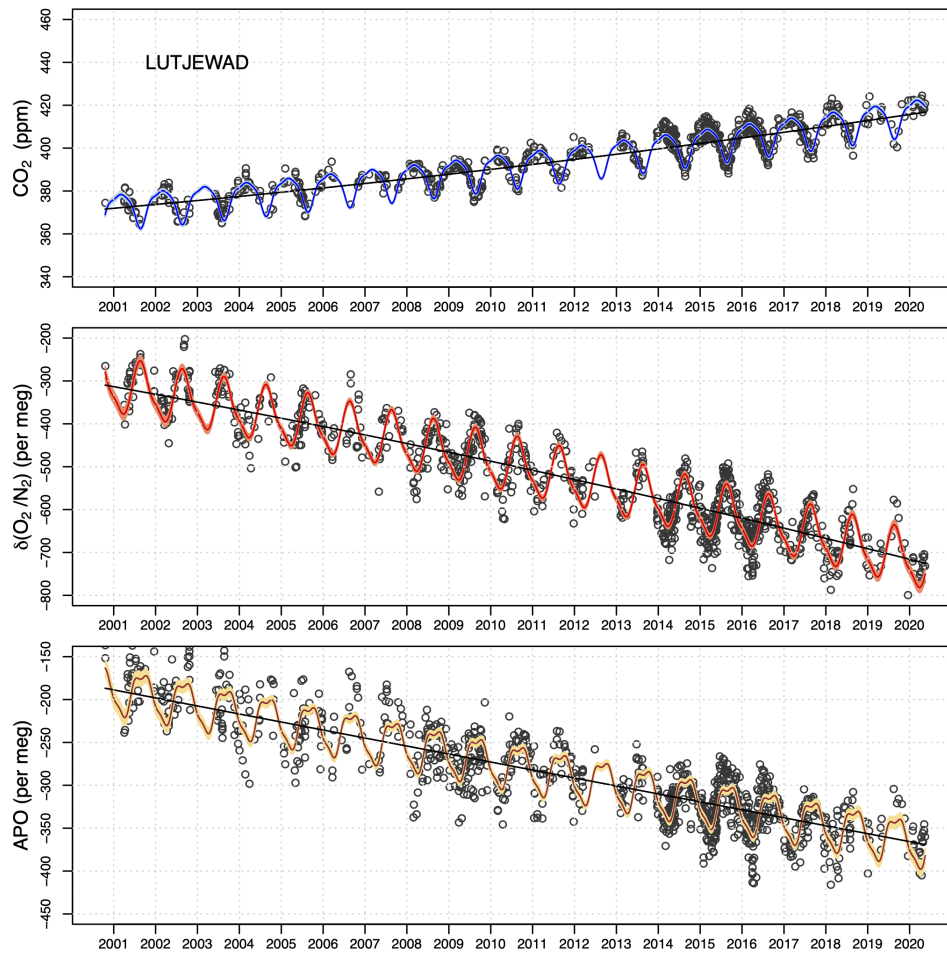
### 5.1 Measurements at Lutjewad, Mace Head, and Halley

Here, we discuss our measurement records in more detail. At first, the difference in the progression of trends in  $\delta(\text{O}_2/\text{N}_2)$  and APO between Lutjewad and Mace Head (Figs. 7 and 8) suggests that there could be an issue with the flask sampling procedure at Mace Head, such as the way the samples are dried. At Lutjewad, the sampling process has been more closely controlled thanks to the vicinity of our laboratory, enabling frequent visits, multiple tests, and other measurements taken from the same sample lines. Furthermore, a comparison of the Lutjewad data with data from the nearby Weybourne coastal station in the UK (presented in Sect. 5.2) showed very good agreement. As both Lutjewad and Mace Head samples share the same measurement procedure, measurement and calibration issues cannot explain their differences, so the differences must either be real or related to the flask sampling procedure. It takes longer to transport the flasks from Mace Head to Groningen than from Lutjewad and thus contamination of the samples through the valve caps might have occurred. For the samples from Halley station, the transport time is even longer, but here, additional protective caps (glass or aluminium) with Viton O-rings are used on the valve caps of the flasks to create small buffer volumes that slow down permeation effects. We tested the preservation of the samples using the protective caps by sending flasks to Halley station that were pre-filled with air of known composition, without actually using them. Back in Groningen, we could conclude the integrity of the samples by comparing the measurements before and after shipment, and we found no significant change in  $\delta(\text{O}_2/\text{N}_2)$  after 26 to 51 months. We found a small drift of 0.4 per meg in  $\delta(\text{O}_2/\text{N}_2)$  after 48 months and a drift of  $-0.3$  ppm in  $\text{CO}_2$  after 24 months for a set of 20 flasks. These numbers only amount to biases of 0.008 per meg per month in  $\delta(\text{O}_2/\text{N}_2)$  and 0.013 ppm per month in  $\text{CO}_2$ . Unfortunately, the protective caps were not applied to Mace Head samples. Still, it is hard to imagine how such permeation effects could cause a deviating long-term trend in the data given that the flasks were filled to ambient pressure. Furthermore, the time between taking the sample and analysis was a few months at most. If anything, one would expect more scatter in the record. The same holds for sampling problems, such as incomplete drying.

To summarize, the trends at Lutjewad are as expected while those at Mace Head are not, so if there are no systematic sampling errors, the differences in  $\delta(\text{O}_2/\text{N}_2)$  and APO at Mace Head compared to Lutjewad might be partially caused by the sparse and irregular sampling frequency at Mace Head

or technical issues that remain undiagnosed. However, it is also worthwhile to consider effects that may be caused by real environmental differences between the two stations. Two effects come to mind: the first is a difference in fossil fuel use (both in quantity and type), which would influence  $\delta(\text{O}_2/\text{N}_2)$  and to a lesser extent also APO. The average fossil fuel ER for the Netherlands, when accounting for all fossil fuel types, is  $1.60 \pm 0.02$  for the 2000–2020 period, much higher than that for Mace Head (1.49; see van Der Laan-Luijkx et al., 2010, and the  $\text{CO}_2$  release and Oxygen uptake from the Fossil Fuel Emission Estimate (COFFEE) database of Steinbach et al., 2011), and the global average value for all fossil fuel emissions (1.38), as also mentioned by Sirignano et al. (2010) and van Der Laan-Luijkx et al. (2010). However, it is unlikely that this is the main explanation for the difference between the two records. Firstly, at Lutjewad, sampling was selective so as to avoid continental (and thus local fossil fuel) influences as much as possible, and secondly, a difference in trends requires a gradual change in the ER. Data from Statistics Netherlands (CBS, 2021) show that the ER of the Netherlands changed by no more than 0.02 over the period 2000–2020, too small to have influenced the observed difference in the trends at Lutjewad and Mace Head. The second potential (though less likely) cause for the differences between Mace Head and Lutjewad is changes in North Atlantic oxygen ventilation (Keeling and Manning, 2014), to which the Mace Head observations are more sensitive. Such changes would influence  $\delta(\text{O}_2/\text{N}_2)$  and APO, but not  $\text{CO}_2$ . This is consistent with the fact that the  $\text{CO}_2$  trends of Mace Head and Lutjewad agree, whereas there are differences in  $\delta(\text{O}_2/\text{N}_2)$  and APO. Changes in the oxygen inventory of the North Atlantic have been reported by Stendardo and Gruber (2012) and Montes et al. (2016) and a relationship with the North Atlantic Oscillation (NAO) has been reported. Data obtained from the NOAA Climate Prediction Center (<https://www.cpc.ncep.noaa.gov/data/teledoc/nao.shtml>, last access: 13 June 2021) show that the NAO exhibited gradual changes over the period 2000–2020, from a noisy, more-or-less balanced positive–negative pattern in the first decade, through to a negative phase in the years 2010–2011, towards gradually mostly positive values for the period 2013–2019. Other potential explanations include a shift in atmospheric transport and data artefact(s). As our operation continues, the coming years might shine light on what are more or less likely causes.

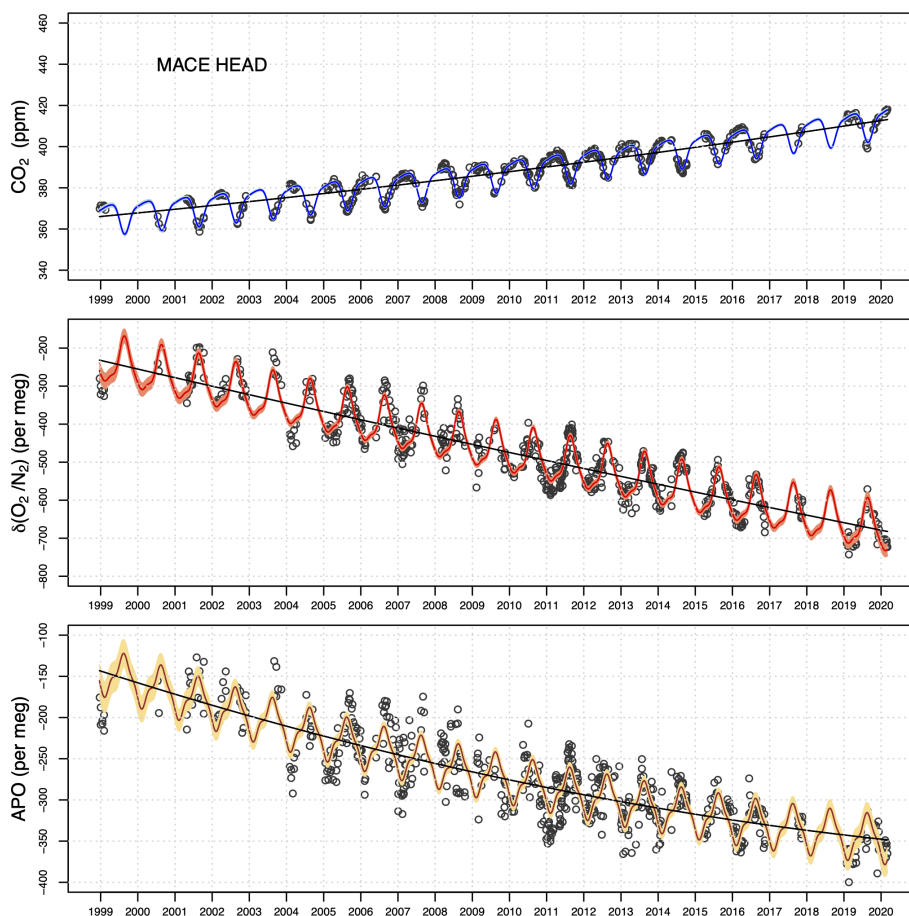
When comparing the seasonal cycles of the three stations, we can see that while  $\text{CO}_2$  and  $\delta(\text{O}_2/\text{N}_2)$  seasonal amplitudes at Halley are significantly smaller than those at Lutjewad and Mace Head, the APO seasonal amplitude is slightly higher, agreeing with the model simulation by Tohjima et al. (2012) that the APO seasonal variations in the Southern Hemispheric ocean are larger than those in the Northern Hemisphere due to larger air–sea  $\text{O}_2$  exchange. As mentioned in Sect. 2, APO values also contain a small influence from fossil fuels; however, by selecting for flasks based on



**Figure 7.** Flask record from Lutjewad station, showing  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO measurements from 2000 to 2020. The black points are the individual flask measurements, the black lines are the long-term trends, and the coloured lines indicate the trends with seasonal components derived from the combined quadratic and harmonic regression. The uncertainty ranges ( $2\sigma$ ) in the fits are indicated by lighter shades of the same colours. For comparability, the y-axes ranges are scaled to represent the 5 per meg : 1 ppm ratio.

**Table 4.** Trends and seasonality fit parameters of the measurement records from all three stations, as presented in Figs. 7–9.

		Lutjewad (2002–2018)	Mace Head (2002–2018)	Halley (2014–2017)
$\text{CO}_2$	Seasonal amplitude (ppm)	$16.8 \pm 0.5$	$14.8 \pm 0.3$	$3.0 \pm 0.3$
	Average trend ( $\text{ppm yr}^{-1}$ )	$2.31 \pm 0.07$	$2.22 \pm 0.04$	$2.60 \pm 0.20$
	Day of min. value	$236 (24 \text{ Aug}) \pm 13$	$238 (26 \text{ Aug}) \pm 11$	$11 (11 \text{ Jan}) \pm 12$
	Day of max. value	$62 (3 \text{ Mar}) \pm 26$	$105 (15 \text{ Apr}) \pm 30$	$216 (4 \text{ Aug}) \pm 14$
$\delta(\text{O}_2/\text{N}_2)$	Seasonal amplitude (per meg)	$131 \pm 6$	$131 \pm 6$	$76 \pm 4$
	Average trend (per meg $\text{yr}^{-1}$ )	$-21.2 \pm 0.8$	$-21.3 \pm 0.9$	$-15 \pm 3$
	Day of min. value	$85 (26 \text{ Mar}) \pm 23$	$42 (11 \text{ Feb}) \pm 33$	$239 (27 \text{ Aug}) \pm 18$
	Day of max. value	$234 (22 \text{ Aug}) \pm 19$	$234 (22 \text{ Aug}) \pm 13$	$59 (28 \text{ Feb}) \pm 21$
APO	Seasonal amplitude (per meg)	$54 \pm 4$	$61 \pm 5$	$65 \pm 3$
	Average trend (per meg $\text{yr}^{-1}$ )	$-9.3 \pm 0.5$	$-9.7 \pm 0.9$	$-1.4 \pm 2.4$
	Day of min. value	$96 (6 \text{ Apr}) \pm 21$	$38 (7 \text{ Feb}) + 30$	$250 (7 \text{ Sep}) \pm 12$
	Day of max. value	$284 (11 \text{ Oct}) \pm 29$	$229 (17 \text{ Aug}) \pm 29$	$66 (7 \text{ Mar}) \pm 17$



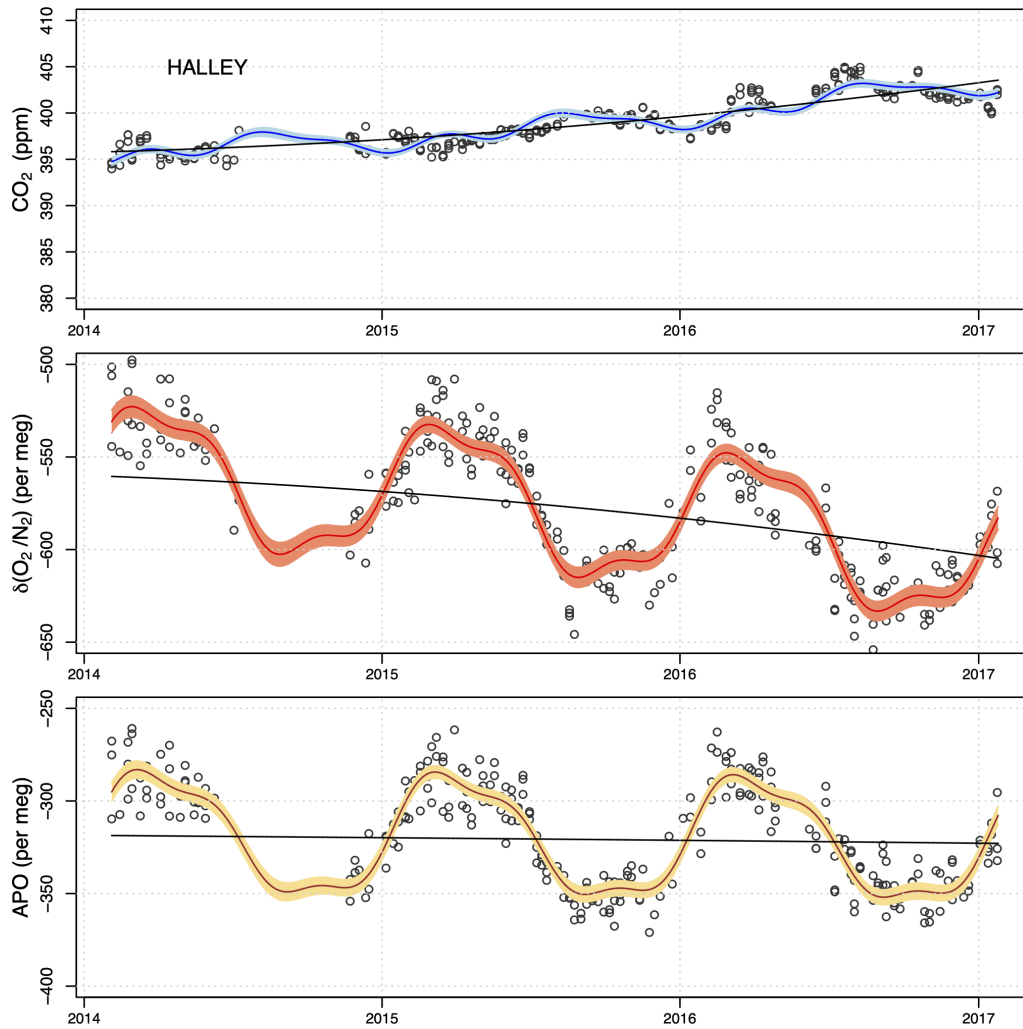
**Figure 8.** As for Fig. 7 but for Mace Head station.

the background conditions, we eliminate this influence as much as possible, especially for the Lutjewad record. As such, our APO values from these three stations mostly represent ocean influences.

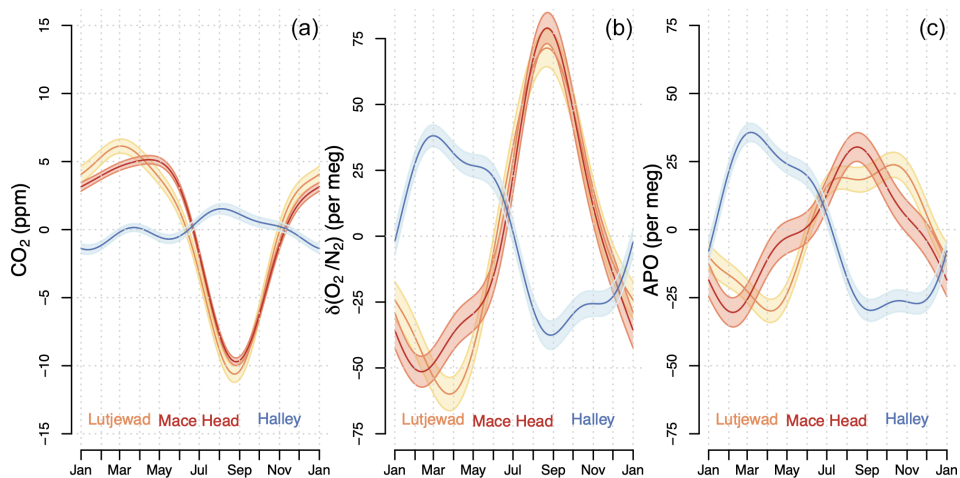
As an illustration of the usefulness of the  $\delta(\text{O}_2/\text{N}_2)$  measurement, we calculated the partitioning of  $\text{CO}_2$  uptake by the terrestrial biosphere and the ocean from the observations at Lutjewad, using the measurements of  $\text{CO}_2$  and APO concentrations from 2002 to 2018, following the method described by Keeling and Manning (2014), but using the fitted trend lines from Lutjewad instead of global averaged values. This partitioning is illustrated in Fig. 11.

The black points are the annual averages of the de-seasonalized measurements of APO and  $\text{CO}_2$  mole fractions at Lutjewad for the period 2002–2018. For calculating the partitioning of fossil fuel  $\text{CO}_2$ , we use Eqs. (2) to (10), and the ocean  $\text{O}_2$  outgassing component ( $Z$ ) of  $0.44 \pm 0.45 \times 10^{14} \text{ mol yr}^{-1}$  (equivalent to an effect on the carbon sinks of  $0.46 \pm 0.48 \text{ PgC yr}^{-1}$ ) from Keeling and Manning (2014). Furthermore, we use the total fossil fuel emissions for the years 2002–2018 of  $8.9 \pm 0.5 \text{ PgC yr}^{-1}$  as derived from the Global Carbon Budget 2021 by Friedlingstein et al. (2021),

and the ER for globally averaged fossil fuel combustion of 1.43 from Jones et al. (2021). To allow comparison of our  $\delta(\text{O}_2/\text{N}_2)$ -derived carbon budget with Friedlingstein et al. (2021), we need to adjust our estimate for the river flux of carbon of  $0.61 \text{ PgC yr}^{-1}$ , similar to their  $f\text{CO}_2$  estimates and inverse results, since all of these methods are based on contemporary observations (see also Hauck et al., 2020). Using the Lutjewad measurement of  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$ , we then derive for the period 2002–2018 a global land biotic sink ( $B$ ) of  $1.9 \pm 1.1 \text{ PgC yr}^{-1}$ , a global ocean sink ( $O$ ) of  $2.1 \pm 0.8 \text{ PgC yr}^{-1}$ , and the  $\text{CO}_2$  remaining in the atmosphere amounts to  $4.89 \pm 0.15 \text{ PgC yr}^{-1}$ . These values agree well with those reported by Friedlingstein et al. (2021) for the same period:  $1.6 \pm 0.9 \text{ PgC yr}^{-1}$  for  $B$  (including emissions from land-use changes) and  $2.5 \pm 0.4 \text{ PgC yr}^{-1}$  for  $O$ . The value for atmospheric component  $A$  at Lutjewad is slightly higher than the reported average value of  $4.66 \pm 0.02 \text{ PgC yr}^{-1}$  for the 2002–2018 period; therefore, our sum of  $O$  and  $B$  is lower than that of Friedlingstein et al. (2021) by the same amount. Additionally, we tested the sensitivity of the calculated sinks to different ER values. When applying an ER value of 1.38 (as was used by Keel-

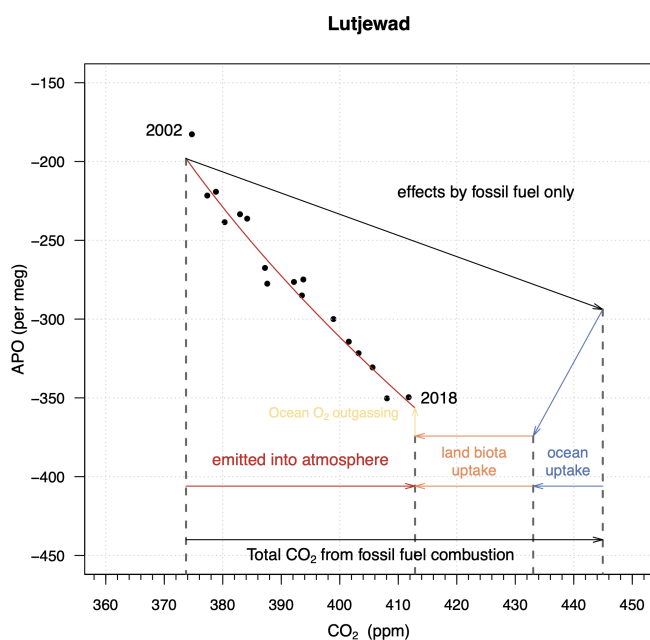


**Figure 9.** As for Figs. 7 and 8, but for Halley station and from 2014 to 2017.



**Figure 10.** The detrended average seasonal cycles of  $\text{CO}_2$  (a),  $\delta(\text{O}_2/\text{N}_2)$  (b), and APO (c) of stations Lutjewad (plotted in orange), Mace Head (plotted in red), and Halley (plotted in blue). The uncertainty margins ( $2\sigma$ ) in the fits are indicated by lighter shades of the same colours.





**Figure 11.** Vector diagram presenting the calculation of the global land biotic and oceanic carbon sinks for the 2002–2018 period. The black points are the annual averages of the measured APO and  $\text{CO}_2$  values at Lutjewad, calculated from January to December of each calendar year. The black arrowed line represents the changes in the atmospheric APO and  $\text{CO}_2$  values that would have occurred if all  $\text{CO}_2$  emitted from fossil fuel combustion remained in the atmosphere. The ocean uptake is presented by blue arrows and its slope is fixed to the APO/ $\text{CO}_2$  molar ratio of 1.1 (which represents the removal of the biosphere signal in the definition of APO). The land biota uptake (orange) is a horizontal line, as APO does not include a biosphere signal. The ocean  $\text{O}_2$  outgassing effect is plotted in brown. The red line is a simple trend fitted through the period.

ing and Manning, 2014), the  $B$  and  $O$  values from Lutjewad record change to 1.5 and  $2.5 \text{ PgC yr}^{-1}$  respectively, which is in even better agreement with Friedlingstein et al. (2021). This shows the importance of knowing the ER of the fossil fuel mix and its changes over time at a high level of detail in carbon budget calculations using atmospheric measurements of  $\delta(\text{O}_2/\text{N}_2)$  and  $\text{CO}_2$ .

The challenges we faced in taking  $\text{O}_2$  measurements have presented themselves clearly in this work: the sensitivity of the mass spectrometer, which requires intensive calibration; the quality maintenance of the internal calibration scale to make sure that our measurements can be reported with sufficient quality on the international scale; and the unexpected patterns (especially in APO for Mace Head) that could not be fully explained, partly due to the lack of consistent sampling frequency before 2004 (for both stations), during 2012 (for Lutjewad), and between 2017 and 2019 (for Mace Head). The trend and seasonality fitting procedure are also of great importance, as these are also highly sensitive to irregular sampling frequency and biases in the timing in which the

majority of the samples are collected. Nevertheless, our flask measurement records of Lutjewad, Mace Head, and Halley have proven to be informative and valuable in evaluating APO, and with future technical improvements (especially regarding the sampling frequency and the quality maintenance of our internal scale), they will be extended further. In the near future, in addition to more regular sampling frequency at Lutjewad and Mace Head, we aim to improve the frequency at which we perform the measurements on the SIO primary standard cylinders, and also to purchase new primary standard cylinders from them in order to produce higher precision conversion to the SIO scale. We also aim to employ more WTs as the current ones are either running out or experiencing considerable noise (see WT 4845 in Fig. 3). We have now added another cylinder to measure along with our last stable WT in order to ensure the continuation of our calibration scale quality. More protective measures for the flasks, such as using additional caps or switching to another type of valve, will also be considered in order to reduce the risks of potential leakages, permeations, and contamination during storage and transportation.

## 5.2 Comparison with other long-term records

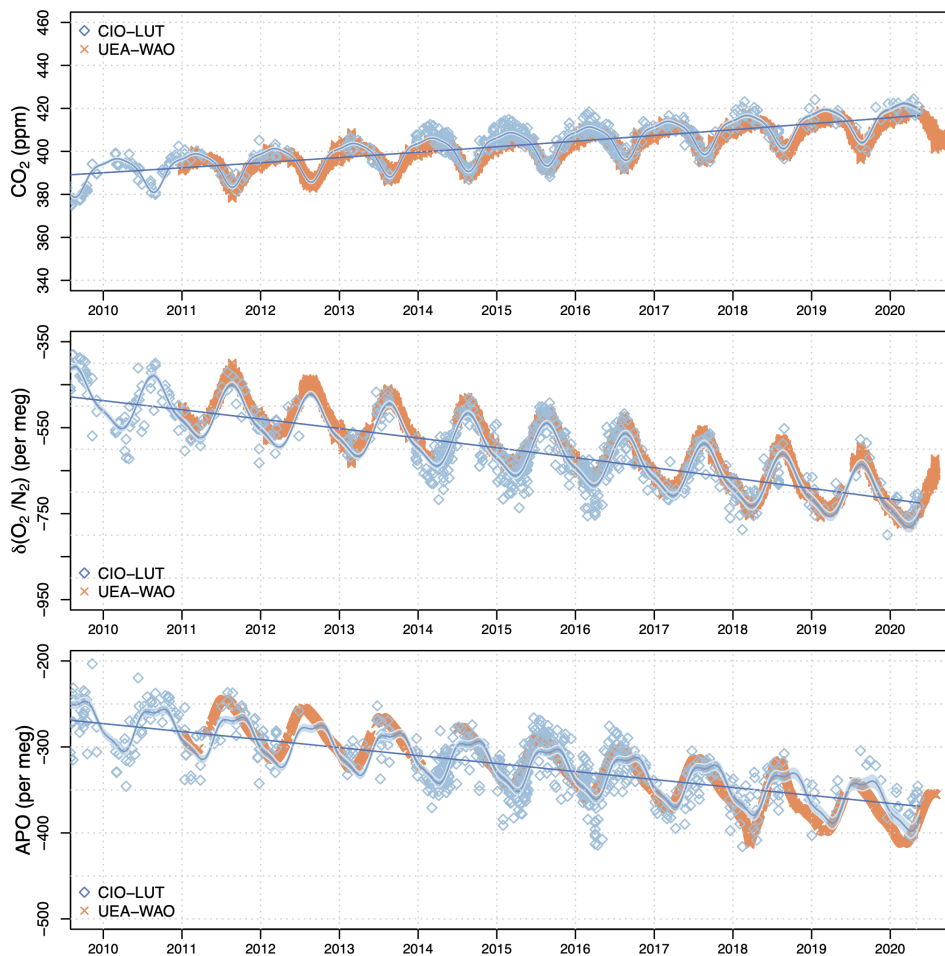
In Table 5, we compare the seasonal amplitudes of our  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO measurements with those of some other stations worldwide. As can be seen, the measurements for all three species at Lutjewad and Mace Head agree well with the measurements conducted at other Northern Hemisphere stations: Weybourne (UK), Sendai (Japan), and Ny Ålesund (Norway). In the Southern Hemisphere, our  $\delta(\text{O}_2/\text{N}_2)$  and APO measurements for Halley station show an excellent agreement with those at Syowa station. On the other hand, our  $\text{CO}_2$  measurements exhibit a much larger and noisier seasonal cycle, which is caused by small leaks during sampling (the details of which are given at the end of this section). Nonetheless, the general concurrence with these stations helps to confirm the quality of our measurements.

Additionally, we compare our long-term measurement record with an extended record of Weybourne station (Fig. 12), the first part of which has been published by Pickers (2016) and Barningham (2018). The figure shows the continuous Weybourne record as hourly averages. In general, the two records agree well, except for the period of late 2018 to the end of 2019, when the flask measurements (and the fit curves) of  $\text{CO}_2$  and APO at Lutjewad are slightly higher than those at Weybourne. This difference is due to the fact that the Weybourne hourly measurements make year-to-year variability (in trend and seasonal cycle) visible, whereas the Lutjewad record, due to its sparser sampling character, is fitted with a smooth trend and a seasonal cycle that is fixed over the years. Apparently, the 2018–2019 period deviated from the average trend and/or seasonal cycle. However, the overall agreement further confirms the quality of our measurements.

**Table 5.** Comparison of the seasonal amplitudes of  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO at various locations in the world.

Station	Time period	Latitude	$\text{CO}_2$ (ppm)	$\delta(\text{O}_2/\text{N}_2)$ (per meg)	APO (per meg)	Reference
Ny Ålesund, Spitsbergen	2001–2010	79° N	$15.2 \pm 0.4$	$129 \pm 4$	$52 \pm 3$	Ishidoya et al. (2012b)
Weybourne, UK	2008–2015	53° N	$15.2 \pm 1.1$	$130 \pm 8$	$51 \pm 6$	Barningham (2018)
Lutjewad, the Netherlands	2002–2018	53° N	$16.8 \pm 0.5$	$131 \pm 6$	$54 \pm 4$	This paper
Mace Head, Ireland	2002–2018	53° N	$14.8 \pm 0.3$	$130 \pm 6$	$61 \pm 5$	This paper
Sendai, Japan	1999–2012	38° N	$13.9 \pm 2.5$	$128 \pm 22$	$52 \pm 10$	Ishidoya et al. (2012a)
Syowa, Antarctica	2001–2010	69° S	$1.1 \pm 0.04$	$70 \pm 4$	$64 \pm 4$	Ishidoya et al. (2012b)
Halley, Antarctica	2014–2017	75° S	$3.0 \pm 0.3^*$	$76 \pm 4$	$65 \pm 3$	This paper

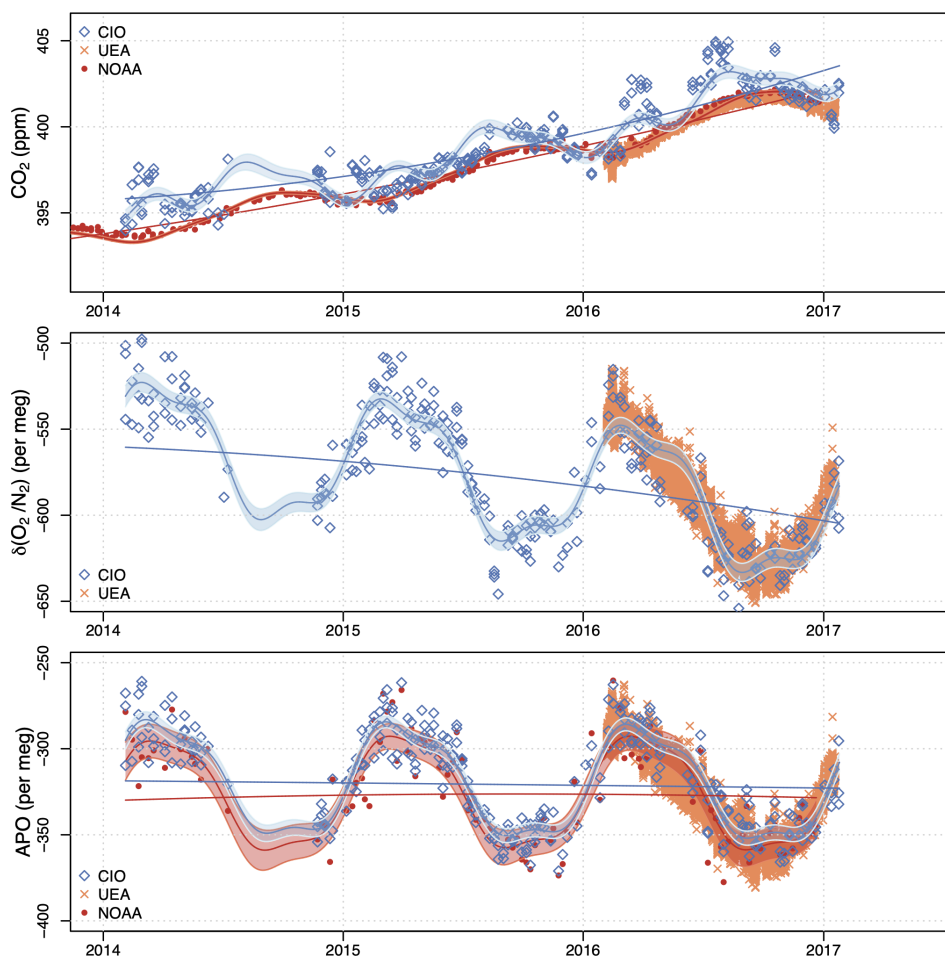
\* The  $\text{CO}_2$  seasonal amplitude at Halley is most likely incorrect; details are given at the end of this section.



**Figure 12.** Measurements of  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO at Lutjewad (black diamonds) and Weybourne (orange crosses) from 2010 to 2020. The black line and curve are the trend and the combined fit for Lutjewad respectively. The grey shadings are the 95 % CI associated with the total fit.

For Halley, we compare our  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO measurements with those conducted by UEA (Fig. 13; Barningham, 2018). The APO measurements of our laboratory and UEA show good agreement, while  $\text{CO}_2$  measurements show unexpected discrepancies in March, April, and June until August of 2016.  $\delta(\text{O}_2/\text{N}_2)$  measurements also show a

slight disagreement, but it is less visible due to a large seasonal cycle and higher scatter. Because APO agrees well, we conclude that the  $\text{CO}_2$  and  $\delta(\text{O}_2/\text{N}_2)$  anomalies were most likely caused by a small inwards leak when the flask samples were collected at the station. Laboratory air with higher  $\text{CO}_2$  mole fractions and lower  $\delta(\text{O}_2/\text{N}_2)$  ratios due to human



**Figure 13.** Measurements of  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO at Halley conducted by CIO (black diamonds) from 2014 to 2017, and continuous measurements conducted by UEA (orange crosses) in 2016. The black lines and curves are the trends and the combined fit for measurements by CIO respectively. The grey shaded area is the 95 % CI associated with the total fit. The red points are the in situ continuous measurements at Halley taken by NOAA. The red lines and curves are the trends and the combined fits for the continuous measurements, with the lighter red shaded area showing the 95 % CI associated with the total fit. The red and blue lines in the APO plot are the average values for our APO measurements, and the ones calculated using the NOAA  $\text{CO}_2$  and our  $\delta(\text{O}_2/\text{N}_2)$  respectively. The latter is significantly lower, corroborating our conclusion that our  $\text{CO}_2$  measurements must have been contaminated with inside air (human breathing). The  $\text{CO}_2$  scale is zoomed in to show the anomalies in 2016 more clearly.

breathing probably leaked in. An additional indication pointing to this is that the  $\text{CH}_4$  and CO mole fractions from the same flasks agree very well with long-term flask measurements taken at Halley by NOAA (2021; not shown here). Such leaks do not influence APO, as the ER from human breathing is close to the value of 1.1 used for the exclusion of the biosphere signal in APO. To better check how much these anomalies would have affected our measurements, we also use the long-term flask measurements taken at Halley from the NOAA website ([https://gml.noaa.gov/dv/iadv/graph.php?code=\\_HBA&program=\\_ccgg&type=\\_ts](https://gml.noaa.gov/dv/iadv/graph.php?code=_HBA&program=_ccgg&type=_ts), last access: 2 April 2021), since the UEA measurement period is too short to make a reliable comparison. For  $\text{CO}_2$ , we perform the same trend and seasonality fitting procedure as for our own measurements. The measurements of the NOAA and

UEA flasks agree very well, showing the reliability of UEA measurements. Thus, the disagreement of  $\text{CO}_2$  and  $\delta(\text{O}_2/\text{N}_2)$  measurements between our laboratory and UEA firmly indicate the presence of leakages during March–August 2016, possibly due to human breathing. As aforementioned, APO should be unaffected by these leakages, as can be seen in the agreement between our APO measurements. In the early 2014 period, there are also some anomalies in our  $\text{CO}_2$  measurements as compared to those of NOAA, but since there is no available information on  $\delta(\text{O}_2/\text{N}_2)$ , we combine the NOAA  $\text{CO}_2$  measurements with our own  $\delta(\text{O}_2/\text{N}_2)$  measurements to calculate APO. Plotted in red are the results using the NOAA  $\text{CO}_2$  measurements. A clear bias in APO is visible, coinciding with the  $\text{CO}_2$  anomalies: the  $\text{CO}_2$  anomalies are around 2 ppm, which leads to corresponding changes

of 10 per meg in APO (since the APO is constructed from “clean”  $\text{CO}_2$  and “contaminated”  $\delta(\text{O}_2/\text{N}_2)$ ). The short-term variations in  $\delta(\text{O}_2/\text{N}_2)$  and APO are greater than 10 per meg, masking the suspected leaks. However, the significant difference between the average values of our APO measurements and those calculated using the NOAA  $\text{CO}_2$  and our  $\delta(\text{O}_2/\text{N}_2)$  measurements (indicated by the red and blue lines in the APO plot respectively) suggest that our flasks must have been contaminated with inside air in the early 2014 period.

## 6 Data availability

The accompanying database comprises three csv files. The files contain the information on the  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO measurements (measured values and associated uncertainties) of the three stations, and are named after the corresponding station and the measured parameter (nine files in total).

All files are published by the ICOS Carbon Portal, and are available at <https://doi.org/10.18160/qq7d-t060> (Nguyen et al., 2021).

Other data presented in this paper are available upon request.

## 7 Conclusions

We have presented 20-year flask measurement records for  $\delta(\text{O}_2/\text{N}_2)$ ,  $\text{CO}_2$ , and APO from Lutjewad and Mace Head, along with 3-year records from Halley. We also presented results of the calibration procedures of our instruments. Due to the sensitive nature of oxygen measurements, we conducted an extensive and intensive calibration procedures, which demonstrated long-term stability for  $\delta(\text{O}_2/\text{N}_2)$  of 3 per meg in 11 years based on our own internal cylinders and 8.6 per meg in 10 years based on our Scripps primary standards. Measurements of the global primary standard cylinders (from SIO) and inter-comparison cylinders (from the Cucumber and GOLLUM programmes) confirm the stability, quality, and comparability of our calibration procedure, although there are some indications that our calibration scale might not be entirely stable over the past 20 years. However, the results from these various programmes are not consistent and are therefore inconclusive. The long-term records from Lutjewad and Mace Head provided useful information on the 20-year trends and seasonality of  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$ , and APO, showing good agreements with other stations around the world, especially the Weybourne Atmospheric Observatory in the UK. We found long-term trends during the period 2002–2018 of  $2.31 \pm 0.07 \text{ ppm yr}^{-1}$  for  $\text{CO}_2$  and  $-21.2 \pm 0.8 \text{ per meg yr}^{-1}$  for  $\delta(\text{O}_2/\text{N}_2)$  at Lutjewad, and  $2.22 \pm 0.04 \text{ ppm yr}^{-1}$  for  $\text{CO}_2$  and  $-21.3 \pm 0.9 \text{ per meg yr}^{-1}$  for  $\delta(\text{O}_2/\text{N}_2)$  at Mace Head. The notable differences in the year-to-year progression of  $\delta(\text{O}_2/\text{N}_2)$  and APO trends between Lutjewad and Mace Head might in part be caused by the sparse sampling frequency at Mace Head, but may also

potentially be indications of influence from the changes in continental fossil fuel use, different degrees of sensitivity to the North Atlantic  $\text{O}_2$  ventilation, a shift in atmospheric transport, or an artefact in the data. Using the measurements at Lutjewad for 2002–2018, the values from partitioning atmospheric  $\text{CO}_2$  sinks into the global terrestrial biosphere and the oceans are  $1.9 \pm 1.1$  and  $2.1 \pm 0.8 \text{ PgC yr}^{-1}$  respectively. These values agree well with the numbers reported in the most recent Global Carbon Budget (Friedlingstein et al., 2021). The Halley record shows that the APO seasonal variations in the Southern Ocean are slightly larger than those in the Northern Hemisphere due to the greater air–sea  $\text{O}_2$  exchange there, and it clearly illustrates the influence of oceanic processes on variations in APO and atmospheric  $\text{O}_2$ . With better maintenance of our internal scale, more regular sampling frequency, and better quality-control of the sampling process, the reliability of our future flask measurements will be improved.

**Author contributions.** LNTN, HAJM, and ITL conducted the data analyses, produced all figures and tables, and wrote the manuscript. ITL and HAJM designed the methodology and framework for the calibration procedure of the DI-IRMS. BAMK conducted technical work, prepared the flask samples at Lutjewad station, and carried out the  $\text{CO}_2$  and  $\delta(\text{O}_2/\text{N}_2)$  measurements from flasks collected at all three stations. HAS calibrated the  $\text{CO}_2$  data at CIO. AEJ, NB, and TB performed the measurements at Halley station, prepared the flask samples, and produced the data for comparison. PAP and ACM provided the data from Weybourne station. All co-authors contributed to the writing of the manuscript.

**Competing interests.** The contact author has declared that neither they nor their co-authors have any competing interests.

**Disclaimer.** Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Acknowledgements.** We would like to thank our colleagues and collaborators at CIO (Janette J. Spriensma for logistics, Henk Jansen for help with the Optima DI-IRMS, Marcel de Vries for the Halley sampler construction, and Ramon R. Richie for help with the measurement database). Furthermore, we thank the collaborators at the stations for collecting and transporting the flask samples, specifically Gerard Spain at Mace Head, and the overwintering teams at Halley station. We thank Eric J. Morgan (SIO) for updating us with new values for the primary standard cylinders. We are also grateful to UEA staff and students (Michael Pateki, Philip Wilson, Grant Forster, Anh Dieu Tran, and Leigh Fleming) and those at the Atmospheric Measurement and Observation Facility at the UK National Centre for Atmospheric Science (NCAS-AMOF) for kindly providing data from Weybourne. We also thank the three referees for taking their valuable time to provide us with

very detailed feedback, upon which we could further complete our work. The work at Lutjewad was partially financially supported over the years by the Dutch Research Council (NWO), the European Union Integrated Project Carbo-Ocean (511176), and the Dutch national CATO-2 programme. Measurement of atmospheric  $\text{O}_2$  and  $\text{CO}_2$  at Weybourne was supported by the UK Natural Environment Research Council (NERC) grants NE/F005733/1, NE/I013342/1, QUEST010005, and NE/S004521/1. Since December 2018, the measurement of atmospheric  $\text{O}_2$  and  $\text{CO}_2$  at Weybourne has also been supported by the National Centre for Atmospheric Science (NCAS), funding agreement R8/H12/83/037. Ingrid T. Luijckx received funding from the Dutch Research Council (NWO; grant 016.Veni.171.095). Anna E. Jones and Neil Brough were supported by the BAS core programme “Polar Science for Planet Earth”. Thomas Barningham and Penelope A. Pickers were supported by UK NERC studentships (NE/L50158X/1 and NE/K500896/1). Penelope A. Pickers and Andrew C. Manning have received support from the NERC-funded DARE-UK (Detection and Attribution of Regional greenhouse gas Emissions in the UK) project, grant agreement no. NE/S004211/1.

**Financial support.** This research has been supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, the Natural Environment Research Council (grant nos. NE/F005733/1, NE/I013342/1, QUEST010005, NE/S004521/1, NE/L50158X/1, NE/K500896/1, and NE/S004211/1), the National Centre for Atmospheric Science (grant no. R8/H12/83/037), the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (grant no. 016.Veni.171.095), and the Natural Environment Research Council (Polar Science for Planet Earth grant).

**Review statement.** This paper was edited by David Carlson and reviewed by three anonymous referees.

## References

- Aoki, N., Ishidoya, S., Tohjima, Y., Morimoto, S., Keeling, R. F., Cox, A., Takebayashi, S., and Murayama, S.: Intercomparison of  $\text{O}_2/\text{N}_2$  ratio scales among AIST, NIES, TU, and SIO based on a round-robin exercise using gravimetric standard mixtures, *Atmos. Meas. Tech.*, 14, 6181–6193, <https://doi.org/10.5194/amt-14-6181-2021>, 2021.
- Barningham, T.: Detection and Attribution of Carbon Cycle Processes from Atmospheric  $\text{O}_2$  and  $\text{CO}_2$  Measurements at Halley Research Station, Antarctica and Weybourne Atmospheric Observatory, U.K., PhD Thesis, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom, 2018.
- Battle, M., Fletcher, S. M., Bender, M. L., Keeling, R. F., Manning, A. C., Gruber, N., Tans, P. P., Hendricks, M. B., Ho, D. T., Simonds, C., Mika, R., and Paplawsky, B.: Atmospheric potential oxygen: New observations and their implications for some atmospheric and oceanic models, *Global Biogeochem. Cy.*, 20, GB1010, <https://doi.org/10.1029/2005gb002534>, 2006.
- Bender, M. L., Tans, P. P., Ellis, T., Orchardo, J., and Habfast, K.: A high precision isotope ratio mass spectrometry method for measuring the  $\text{O}_2/\text{N}_2$  ratio of air, *Geochim. Cosmochim. Ac.*, 58, 4751–4758, 1994.
- Bender, M. L., Ellis, T., Tans, P. P., Francey, R., and Lowe, D.: Variability in the  $\text{O}_2/\text{N}_2$  ratio of southern hemisphere air, 1991–1994: Implications for the carbon cycle, *Global Biogeochem. Cy.*, 10, 9–21, 1996.
- Blaine, T. W., Keeling, R. F., and Paplawsky, W. J.: An improved inlet for precisely measuring the atmospheric  $\text{Ar}/\text{N}_2$  ratio, *Atmos. Chem. Phys.*, 6, 1181–1184, <https://doi.org/10.5194/acp-6-1181-2006>, 2006.
- Bousquet, P., Gaudry, A., Ciais, P., Kazan, P., Monfray, P., Simmonds, P. G., Jennings, S. G., and O’Connor, T. C.: Atmospheric  $\text{CO}_2$  Concentration Variations Recorded at Mace Head, Ireland, From 1992 to 1994, *Phys. Chem. Earth*, 21, 477–481, 1996.
- British Antarctic Survey: Halley VI Research Station, <https://www.bas.ac.uk/polar-operations/sites-and-facilities/facility/halley/>, last access: 2 April 2021.
- CBS: The Netherlands in figures, <https://opendata.cbs.nl/statline/#/CBS/en/>, last access: 22 April 2021.
- 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 ( $\text{CO}_2$  and  $\text{CH}_4$ ) using the cavity ring-down spectroscopy (CRDS) technique, *Atmos. Meas. Tech.*, 3, 375–386, <https://doi.org/10.5194/amt-3-375-2010>, 2010.
- Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., C. J., C., L. Q., Myeni, R. B., Piao, S., and Thornton, P.: Carbon and Other Biogeochemical Cycles, in: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- Conway, T. J., Tans, P. P., and Waterman, L. S.: Atmospheric  $\text{CO}_2$  records from sites in the NOAA/CMDL air sampling network, in: *Trends ’93: a compendium of data on global change*, ORNL/CDIAC-65, Oak Ridge National Laboratory, USA, 1994.
- Derwent, R. G., Ryall, D. B., Manning, A. J., Simmonds, P. G., O’Doherty, S., Biraud, S., Ciais, P., Ramonet, M., and Jennings, S. G.: Continuous observations of carbon dioxide at Mace Head, Ireland from 1995 to 1999 and its net European ecosystem exchange, *Atmos. Environ.*, 36, 2799–2807, 2002.
- Friedlingstein, P., O’Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quééré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneeth, A., Arora, V., Bates, N. R., Becker, M., Benoit-Cattin, A., Bittig, H. C., Bopp, L., Bultan, S., Chandra, N., Chevallier, F., Chini, L. P., Evans, W., Florentie, L., Forster, P. M., Gasser, T., Gehlen, M., Gilfillan, D., Gkritzalis, T., Gregor, L., Gruber, N., Harris, I., Hartung, K., Haverd, V., Houghton, R. A., Ilyina, T., Jain, A. K., Joetzer, E., Kadono, K., Kato, E., Kitidis, V., Korsbakken, J. I., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Liu, Z., Lombardozzi, D., Marland, G., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., O’Brien, K., Ono, T., Palmer, P. I., Pierrot, D., Poulter, B., Resplandy, L., Robertson, E., Rödenbeck, C., Schwinger,

- J., Séférian, R., Skjelvan, I., Smith, A. J. P., Sutton, A. J., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., van der Werf, G., Vuichard, N., Walker, A. P., Wanninkhof, R., Watson, A. J., Willis, D., Wiltshire, A. J., Yuan, W., Yue, X., and Zaehle, S.: Global Carbon Budget 2020, *Earth Syst. Sci. Data*, 12, 3269–3340, <https://doi.org/10.5194/essd-12-3269-2020>, 2020.
- Friedlingstein, P., Jones, M. W., O’Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quééré, C., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Bates, N. R., Becker, M., Bellouin, N., Bopp, L., Chau, T. T. T., Chevallier, F., Chini, L. P., Cronin, M., Currie, K. I., Decharme, B., Djéutchouang, L., Dou, X., Evans, W., Feely, R. A., Feng, L., Gasser, T., Gilfillan, D., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Lujikx, I. T., Jain, A. K., Jones, S. D., Kato, E., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lienert, S., Liu, J., Marland, G., McGuire, P. C., Melton, J. R., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Niwa, Y., Ono, T., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Schwingshackl, C., Séférian, R., Sutton, A. J., Sweeney, C., Tanhua, T., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F., van der Werf, G., Vuichard, N., Wada, C., Wanninkhof, R., Watson, A., Willis, D., Wiltshire, A. J., Yuan, W., Yue, C., Yue, X., Zaehle, S., and Zeng, J.: Global Carbon Budget 2021, *Earth Syst. Sci. Data Discuss.* [preprint], <https://doi.org/10.5194/essd-2021-386>, in review, 2021.
- Goto, D., Morimoto, S., Ishidoya, S., Aoki, S., and Nakazawa, T.: Terrestrial biospheric and oceanic  $\text{CO}_2$  uptakes estimated from long-term measurements of atmospheric  $\text{CO}_2$  mole fraction,  $\delta^{13}\text{C}$ , and  $\delta(\text{O}_2/\text{N}_2)$  at Ny-Ålesund, Svalbard, *J. Geophys. Res.-Biogeo.*, 122, 1192–1202, <https://doi.org/10.1002/2017JG003845>, 2017.
- Gruber, N., Gloor, M., Fan, S.-M., and Sarmiento, J. L.: Air-sea flux of oxygen estimated from bulk data: Implications For the marine and atmospheric oxygen cycles, *Global Biogeochem. Cy.*, 15, 783–803, <https://doi.org/10.1029/2000gb001302>, 2001.
- Hall, B. D., Crotwell, A. M., Kitzis, D. R., Mefford, T., Miller, B. R., Schibig, M. F., and Tans, P. P.: Revision of the World Meteorological Organization Global Atmosphere Watch (WMO/GAW)  $\text{CO}_2$  calibration scale, *Atmos. Meas. Tech.*, 14, 3015–3032, <https://doi.org/10.5194/amt-14-3015-2021>, 2021.
- Hauck, J., Zeising, M., Le Quééré, C., Gruber, N., Bakker, D. C. E., Bopp, L., Chau, T. T. T., Gürses, Ö., Ilyina, T., Landschützer, P., Lenton, A., Resplandy, L., Rödenbeck, C., Schwinger, J., and Séférian, R.: Consistency and Challenges in the Ocean Carbon Sink Estimate for the Global Carbon Budget, *Front. Mar. Sci.*, 7, 571720, <https://doi.org/10.3389/fmars.2020.571720>, 2020.
- Ishidoya, S., Aoki, S., Goto, D., Nakazawa, T., Taguchi, S., and Patra, P.: Time and space variations of the  $\text{O}_2/\text{N}_2$  ratio in the troposphere over Japan and estimation of the global  $\text{CO}_2$  budget for the period 2000–2010, *Tellus B*, 64, 18964, <https://doi.org/10.3402/tellusb.v64i0.18964>, 2012a.
- Ishidoya, S., Morimoto, S., Aoki, S., Taguchi, S., Goto, D., Murayama, S., and Nakazawa, T.: Oceanic and terrestrial biospheric  $\text{CO}_2$  uptake estimated from atmospheric potential oxygen observed at Ny-Ålesund, Svalbard, and Syowa, Antarctica, *Tellus B*, 64, 18924, <https://doi.org/10.3402/tellusb.v64i0.18924>, 2012b.
- Jennings, S. G., McGovern, F. M., and Cooke, W. F.: Carbon Mass Concentration Measurements at Mace Head, on the West Coast of Ireland, *Atmos. Environ.*, 27A, 1229–1239, 1993.
- Jones, A. E., Wolff, E. W., Salmon, R. A., Bauguitte, S. J.-B., Roscoe, H. K., Anderson, P. S., Ames, D., Clemmitshaw, K. C., Fleming, Z. L., Bloss, W. J., Heard, D. E., Lee, J. D., Read, K. A., Hamer, P., Shallcross, D. E., Jackson, A. V., Walker, S. L., Lewis, A. C., Mills, G. P., Plane, J. M. C., Saiz-Lopez, A., Sturges, W. T., and Worton, D. R.: Chemistry of the Antarctic Boundary Layer and the Interface with Snow: an overview of the CHABLIS campaign, *Atmos. Chem. Phys.*, 8, 3789–3803, <https://doi.org/10.5194/acp-8-3789-2008>, 2008.
- Jones, M. W., Andrew, R. M., Peters, G. P., Janssens-Maenhout, G., De-Gol, A. J., Ciais, P., Patra, P. K., Chevallier, F., and Le Quééré, C.: Gridded fossil  $\text{CO}_2$  emissions and related  $\text{O}_2$  combustion consistent with national inventories 1959–2018, *Scientific Data*, 8, 2, <https://doi.org/10.1038/s41597-020-00779-6>, 2021.
- Keeling, R. F.: Development of an Interferometric Oxygen Analyzer for Precise Measurement of the Atmospheric  $\text{O}_2$  Mole Fraction, PhD Thesis, Division of Applied Sciences, Harvard University, Cambridge, Massachusetts, 1988.
- Keeling, R. F. and Manning, A. C.: 5.15 – Studies of Recent Changes in Atmospheric  $\text{O}_2$  Content, in: *Treatise on Geochemistry*, 2nd Edn., edited by: Holland, H. D., and Turekian, K. K., <https://doi.org/10.1016/B978-0-08-095975-7.00420-4>, 2014.
- Keeling, R. F. and Shertz, S. R.: Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle, *Nature*, 358, 723–727, <https://doi.org/10.1038/358723a0>, 1992.
- Keeling, R. F., Manning, A. C., McEvoy, E. M., and Shertz, S. R.: Methods for measuring changes in atmospheric  $\text{O}_2$  concentration and their application in southern hemisphere air, *J. Geophys. Res.-Atmos.*, 103, 3381–3397, <https://doi.org/10.1029/97jd02537>, 1998.
- Keeling, R. F., Blaine, T., Paplawsky, B., Katz, L., Atwood, C., and Brockwell, T.: Measurement of changes in atmospheric  $\text{Ar}/\text{N}_2$  ratio using a rapid-switching, single-capillary mass spectrometer system, *Tellus B*, 56, 322–338, <https://doi.org/10.3402/tellusb.v56i4.16453>, 2004.
- Manning, A. C. and Keeling, R. F.: Global oceanic and land biotic carbon sinks from the Scripps atmospheric oxygen flask sampling network, *Tellus*, 58B, 95–116, 2006.
- Manning, A. C., Keeling, R. F., and Severinghaus, J.: Precise atmospheric oxygen measurements with a paramagnetic oxygen analyzer, *Global Biogeochem. Cy.*, 13, 1107–1115, 1999.
- Manning, A. C., Keeling, R. F., Etchells, A. J., Hewitt, M., Bender, M. L., Bracchi, K., Brailsford, G. W., Brand, W. A., Cassar, N., Cox, A. C., Leuenberger, M., Meijer, H. A. J., Morimoto, S., Nakazawa, T., Neubert, R. E. M., Paplawsky, W. J., Richter, J. M., Stephens, B. B., Tohjima, Y., van der Laan, S., van der Laan-Luijkx, I. T., Watt, A., and Wilson, P. A.: The “GOLLUM”  $\text{O}_2$  intercomparison programme: Latest results and next step, Second APO Workshop, [https://gollum.uea.ac.uk/docs/2015\\_APO\\_talks/Manning\\_et\\_al\\_GOLLUM\\_APO\\_Workshop\\_2015.pdf](https://gollum.uea.ac.uk/docs/2015_APO_talks/Manning_et_al_GOLLUM_APO_Workshop_2015.pdf) (last access: 18 February 2022), 2015.

- Montes, E., Muller-Karger, F. E., Cianca, A., Lomas, M. W., Lorenzoni, L., and Habtes, S.: Decadal variability in the oxygen inventory of North Atlantic subtropical under-water captured by sustained, long-term oceanographic time series observations, *Global Biogeochem. Cy.*, 30, 460–478, <https://doi.org/10.1002/2015GB005183>, 2016.
- Mook, W. G.: Volume I: Introduction Theory, Methods, Review, in: *Environmental isotopes in the hydrological cycle: principles and applications*, IAEA and UNESCO, [http://www-naweb.iaea.org/napc/ih/documents/global\\_cycle/EnvironmentalIsotopesintheHydrologicalCycleVol01.pdf](http://www-naweb.iaea.org/napc/ih/documents/global_cycle/EnvironmentalIsotopesintheHydrologicalCycleVol01.pdf) (last access: 18 February 2022), 2000.
- Neubert, R. E. M., Spijkervet, L. L., Schut, J. K., Been, H. A., and Meijer, H. A. J.: A Computer-Controlled Continuous Air Drying and Flask Sampling System, *J. Atmos. Ocean. Tech.*, 21, 651–659, [https://doi.org/10.1175/1520-0426\(2004\)021<0651:ACCADA>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<0651:ACCADA>2.0.CO;2), 2004.
- Nguyen, L. N. T., Meijer, H. A. J., van Leeuwen, C., Kers, B. A. M., Scheeren, B. A., Jones, A. E., Brough, N., Barningham, T., Pickers, P. A., Manning, A. C., and Luijckx, I. T.: Supplement data of the Two decades of flask observations of atmospheric  $\delta\text{O}_2/\text{N}_2$ ,  $\text{CO}_2$ , and APO at stations Lutjewad (the Netherlands) and Mace Head (Ireland) plus 3 years from Halley station (Antarctica), ICOS-ERIC Carbon Portal [data set], <https://doi.org/10.18160/qq7d-t060>, 2021.
- NOAA: Carbon Cycle Gasses Halley Station, Antarctica, United Kingdom, <https://www.esrl.noaa.gov/gmd/dv/iadv/graph.php?code=HBA&program=ccgg&type=ts>, last access: 2 April 2021.
- Pickers, P. A.: New applications of continuous atmospheric  $\text{O}_2$  measurements: meridional transects across the Atlantic Ocean, and improved quantification of fossil fuel-derived  $\text{CO}_2$ , PhD Thesis, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom, 2016.
- Pickers, P. A., Manning, A. C., Sturges, W. T., Le Quéré, C., Mikaloff Fletcher, S. E., Wilson, P. A., and Etchells, A. J.: In situ measurements of atmospheric  $\text{O}_2$  and  $\text{CO}_2$  reveal an unexpected  $\text{O}_2$  signal over the tropical Atlantic Ocean, *Global Biogeochem. Cy.*, 31, 1289–1305, <https://doi.org/10.1002/2017gb005631>, 2017.
- Rousseuw, P. J. and Verboven, S.: Robust estimation in very small samples, *Comput. Stat. Data Anal.*, 40, 741–758, 2002.
- Severinghaus, J. P.: Studies of the terrestrial  $\text{O}_2$  and carbon cycles in sand dune gases and in biosphere 2, PhD Thesis, United States, <https://doi.org/10.2172/477735>, 1995.
- Sirignano, C., Neubert, R. E. M., Rödenbeck, C., and Meijer, H. A. J.: Atmospheric oxygen and carbon dioxide observations from two European coastal stations 2000–2005: continental influence, trend changes and APO climatology, *Atmos. Chem. Phys.*, 10, 1599–1615, <https://doi.org/10.5194/acp-10-1599-2010>, 2010.
- Steinbach, J., Gerbig, C., Rödenbeck, C., Karstens, U., Minejima, C., and Mukai, H.: The  $\text{CO}_2$  release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, *Atmos. Chem. Phys.*, 11, 6855–6870, <https://doi.org/10.5194/acp-11-6855-2011>, 2011.
- Stendardo, I. and Gruber, N.: Oxygen trends over five decades in the North Atlantic, *J. Geophys. Res.-Oceans*, 117, C11004, <https://doi.org/10.1029/2012JC007909>, 2012.
- Stephens, B. B., Keeling, R. F., Heimann, M., Six, K. D., Murnane, R., and Caldeira, K.: Testing global ocean carbon cycle models using measurements of atmospheric  $\text{O}_2$  and  $\text{CO}_2$  concentration, *Global Biogeochem. Cy.*, 12, 213–230, <https://doi.org/10.1029/97gb03500>, 1998.
- Stephens, B. B., Keeling, R. F., and Paplawsky, W. J.: Shipboard measurements of atmospheric oxygen using a vacuum-ultraviolet absorption technique, *Tellus B*, 55, 857–878, <https://doi.org/10.3402/tellusb.v55i4.16386>, 2003.
- Stephens, B. B., Bakwin, P. S., Tans, P. P., Teclaw, R. M., and Baumann, D. D.: Application of a Differential Fuel-Cell Analyzer for Measuring Atmospheric Oxygen Variations, *J. Atmos. Ocean. Tech.*, 24, 82–94, <https://doi.org/10.1175/jtech1959.1>, 2007.
- Stephens, B. B., Morgan, E. J., Bent, J. D., Keeling, R. F., Watt, A. S., Shertz, S. R., and Daube, B. C.: Airborne measurements of oxygen concentration from the surface to the lower stratosphere and pole to pole, *Atmos. Meas. Tech.*, 14, 2543–2574, <https://doi.org/10.5194/amt-14-2543-2021>, 2021.
- Sturm, P., Leuenberger, M., Sirignano, C., Neubert, R. E. M., Meijer, H. A. J., Langenfelds, R., Brand, W. A., and Tohjima, Y.: Permeation of atmospheric gases through polymer O-rings used in flasks for air sampling, *J. Geophys. Res.-Atmos.*, 109, D04309, <https://doi.org/10.1029/2003jd004073>, 2004.
- Thoning, K. W., Tans, P. P., and Komhyr, W. D.: Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–1985, *J. Geophys. Res.-Atmos.*, 94, 8549–8565, <https://doi.org/10.1029/JD094iD06p08549>, 1989.
- Tohjima, Y.: Method for measuring changes in the atmospheric  $\text{O}_2/\text{N}_2$  ratio by a gas chromatograph equipped with a thermal conductivity detector, *J. Geophys. Res.-Atmos.*, 105, 14575–14584, <https://doi.org/10.1029/2000jd900057>, 2000.
- Tohjima, Y., Machida, T., Watai, T., Akama, I., Amari, T., and Moriwaki, Y.: Preparation of gravimetric standards for measurements of atmospheric oxygen and reevaluation of atmospheric oxygen concentration, *J. Geophys. Res.*, 110, D11302, <https://doi.org/10.1029/2004jd005595>, 2005.
- Tohjima, Y., Mukai, H., Nojiri, Y., Yamagishi, H., and Machida, T.: Atmospheric  $\text{O}_2/\text{N}_2$  measurements at two Japanese sites: estimation of global oceanic and land biotic carbon sinks and analysis of the variations in atmospheric potential oxygen (APO), *Tellus B*, 60, 213–225, <https://doi.org/10.1111/j.1600-0889.2007.00334.x>, 2008.
- Tohjima, Y., Minejima, C., Mukai, H., Machida, T., Yamagishi, H., and Nojiri, Y.: Analysis of seasonality and annual mean distribution of atmospheric potential oxygen (APO) in the Pacific region, *Global Biogeochem. Cy.*, 26, GB4008, <https://doi.org/10.1029/2011gb004110>, 2012.
- Tohjima, Y., Mukai, H., Machida, T., Hoshina, Y., and Nakaoka, S.-I.: Global carbon budgets estimated from atmospheric  $\text{O}_2/\text{N}_2$  and  $\text{CO}_2$  observations in the western Pacific region over a 15-year period, *Atmos. Chem. Phys.*, 19, 9269–9285, <https://doi.org/10.5194/acp-19-9269-2019>, 2019.
- University of East Anglia: Cucumbers Intercomparison, <http://cucumbers.uea.ac.uk/>, last access: 24 February 2021.
- van der Laan, S., Neubert, R. E. M., and Meijer, H. A. J.: A single gas chromatograph for accurate atmospheric mixing ratio measurements of  $\text{CO}_2$ ,  $\text{CH}_4$ ,  $\text{N}_2\text{O}$ ,  $\text{SF}_6$  and  $\text{CO}$ , *Atmos. Meas. Tech.*, 2, 549–559, <https://doi.org/10.5194/amt-2-549-2009>, 2009.
- van der Laan, S., Karstens, U., Neubert, R. E. M., van der Laan-Luijckx, I. T., and Meijer, H. A. J.: Observation-based estimates

- of fossil fuel-derived  $\text{CO}_2$  emissions in the Netherlands using  $^{14}\text{C}$ ,  $\text{CO}$  and  $^{222}\text{Rn}$ , *Tellus*, 62B, 389–402, 2010.
- van der Laan-Luijkx, I. T.: Atmospheric oxygen and the global carbon cycle. Observations from the new F3 North Sea platform monitoring station and 6 additional locations in Europe and Siberia, PhD Thesis, Groningen, The Netherlands, 2010.
- van der Laan-Luijkx, I. T., Karstens, U., Steinbach, J., Gerbig, C., Sirignano, C., Neubert, R. E. M., van der Laan, S., and Meijer, H. A. J.:  $\text{CO}_2$ ,  $\delta\text{O}_2/\text{N}_2$  and APO: observations from the Lutjewad, Mace Head and F3 platform flask sampling network, *Atmos. Chem. Phys.*, 10, 10691–10704, <https://doi.org/10.5194/acp-10-10691-2010>, 2010.
- van der Laan-Luijkx, I. T., van der Laan, S., Uglietti, C., Schibig, M. F., Neubert, R. E. M., Meijer, H. A. J., Brand, W. A., Jordan, A., Richter, J. M., Rother, M., and Leuenberger, M.: Atmospheric  $\text{CO}_2$ ,  $\delta(\text{O}_2/\text{N}_2)$  and  $\delta^{13}\text{CO}_2$  measurements at Jungfraujoch, Switzerland: results from a flask sampling intercomparison program, *Atmos. Meas. Tech.*, 6, <https://doi.org/10.5194/amt-6-1805-2013>, 2013.
- van Leeuwen, C.: Highly precise atmospheric oxygen measurements as a tool to detect leaks of carbon dioxide from Carbon Capture and Storage sites, University of Groningen, Groningen, The Netherlands, 2015.
- World Meteorological Organization: 19th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Measurement Techniques (GGMT-2017), World Meteorological Organization (WMO), Switzerland, 2018.
- Worthy, D. E. J., Platt, A., Kessler, R., Ernst, M., and Racki, S.: The greenhouse gases measurement program, measurement procedures and data quality, *Meteorol. Serv. of Can., Environ. Can., Downsview, Ont., Canada*, 97–120, 2003.