Ref 1:

This paper provides two decadal data of atmospheric CO_2 and O_2 observed at Lutjewad and Mace Head and 3year record at Halley. As is the case with the atmospheric CO_2 , the atmospheric O_2 data from a variety of laboratories are also expected to be synthetically analyzed by using atmospheric transport models,

- 5 biogeochemical models, and so on. However, compared with the atmospheric CO₂ measurements, the atmospheric oxygen measurements are still very challenging because we need much more efforts in the process of the air sampling, storing, analysis, and scale maintenance. Especially, it is crucially important to how to keep the O₂ scale stability. The authors describe the details of the calibration procedure and several efforts to check the O₂ scale stability. However, the authors should make much more effort to clarify the data quality and
- 10 quantitatively describe the uncertainties associated with the flask measurements in the manuscript. Although I found that the paper contains material that should be published in ESSD, I recommend the manuscript to be published after following minor revisions.

Thank you for these comments, we have implemented changes to better clarify the uncertainties following the suggestions throughout the manuscript.

15 General comments:

50

I understand authors' various effort to keep highly precise measurements of the atmospheric O_2/N_2 ratio of the flask samples. However, the $\delta(O_2/N_2)$ and APO values at Lutjewad and Mace Head plotted in Figs 6 and 7 show rather scattered plots, which don't seem to be real variations. So, I suspect that the uncertainty of the flask measurement is not so small to adequately detect the atmospheric variation. It is crucially important to clarify the total uncertainty associated with the O_2 data of the flask samples for the synthetic analyses together with the

20 the total uncertainty associated with the O₂ data of the flask samples for the synthetic analyses together with the data from other laboratories. Nevertheless, I cannot find any clear description of the analytical precision and the repeatability of the flask measurements in the manuscript. In addition, the authors described the contamination of the flask samples collected at Halley during the storing period. If the same type of the flasks were used for the air sampling at Lutjewad and Mace Head, there is a possibility that the contamination would cause the positive and negative biases of the CO₂ and O₂/N₂ values for the flask samples, respectively. These potential biases should be also evaluated in the manuscript.

We have added in the uncertainty from the flask measurements, and also the total uncertainty associated with the final long-term trends. As for the potential contamination, it is unlikely to cause significant biases on the flask samples from Lutjewad and Mace Head. We did a storability test on flasks going to Antarctica, where we
pre-filled a set of flasks that then went to Antarctica, stored there for ~2 years before coming back to our lab for re-measurements. We found a negligible drift of 0.4 per meg in dO2/N2 after 48 months; and a drift of -0.3 ppm in CO2 after 24 months, on a set of 20 flasks. These numbers would only amount to biases of 0.008 per meg /month in dO2/N2 and 0.013 ppm/month in CO2. We collected our flasks from Lutjewad weekly, and Mace Head monthly, therefore the systematic effects (if any) would only be negligible. Only leakages during each individual sampling session would give rise to outliers, and they should be discarded during our filtering

35 individual sampling session would give rise to outliers, and they should be discarded during our filtering process.

The O_2 scale stability is also very important as the authors also recognized. Although the evaluation of the stability is very difficult because there is no absolute scale at present, the authors should quantitatively evaluate the overall stability of the O_2 scale in this manuscript. In section 3.1, the authors described that the standard

40 deviations of the repeated measurements of the working tanks were less than 13.5 per meg. Did it mean that the uncertainty of the CIO scale stability was estimated to be about 1 per meg/yr (=13.5 per meg/14 years)? Probably, the results of the COLLUM cylinders would also give a clue of the quantitative evaluation of the O₂ scale stability. The evaluation of the uncertainty of the scale stability is directly related to the evaluation of the uncertainty of the carbon budget evaluation described in page 20.

45 Our O2 scale stability is determined by the stability of our long-term WTs (standard deviation of 13.5 per meg in 14 years) and that of our Scripps primary standard cylinders (8.6 per meg in 10 years). We have now added this information in our main text.

I think that the trend of APO at Mace Head is rather curious because the decreasing rate of APO trend gradually decrease from -15.15 per meg yr⁻¹ in 2002 to -5.83 per meg yr⁻¹ in 2018. The authors attributed to the O_2 emissions from North Atlantic associated with the gradual changes of the NAO. However, I cannot accept

the mechanism that the O_2 emissions from the North Atlantic only influenced the O_2/N_2 and APO at Mace Head. The APO decreasing rates are computed from the fitted quadratic functions. However, taking the data variability for both sites and sparse sampling frequency for Mace Head into account, I suspect that there are no significant differences in the trends between the two sites.

55 It is indeed, from the long-term calculation, that there is minimal differences in trends between Lutjewad and Mace Head. However, due to some unknown reasons – which we can only attribute speculatively to some possibilities – there are significant differences in the annual trends (i.e. the gradient of the fit curve).

Specific comments:

85

Page 2, line 55: "(Tohjima, 2005)" should be "(Tohjima et al., 2005)". And please add the coauthors to the reference (Page 29, line 827).

We have fixed the reference

Page 3, line 102-114: At Lutjewad, the air sample was dried by passing it through a Nafion drying tube. How about the Mace Head and Halley stations? The Nafion drying tubes were used at both sites?

We have added the drying agents at Mace Head (Mg(ClO₄)₂, the same as in Halley). Nafion is only used at Lutjewad.

Page 4, line 123: It would be better to clarify the temperature of the cryogenic drier.

We added in the temperature (although it is already described at line 94).

Page 4, line 132-136: Were the same 2.5-liter glass flasks as Lutjewad used at Mace Head and Halley stations?

We added the information (and yes, they are all of the same type of flask).

- 70 Page 8, line 252: Is it possible to describe the linear function to convert the O₂/N₂ value based on the CIO scale to that based on SIO scale? Is the conversion function fixed during the observation period of this study? In addition, I think it would be better to describe the uncertainties for the coefficients of the linear function. Such data would be useful to consider the propagation errors for the O₂/N₂ values of the flask samples and the standard cylinders.
- 75 We included the function now. The function is fixed, based on the measurements of the Scripps primary standard cylinder that have been corrected for drifts for the whole period.

Page 8, line 267-268: WT4845 show rather unstable O_2/N_2 values. It would be informative, if possible, to describe the reason of the instability.

Unfortunately, we don't know what was wrong with the WT4845, but it might be related to the fact that its value is rather low in comparison to the other tanks – which suggests that the cylinder might contain contaminated air or there could be leaks on the pressure reducer.

Page 9, line 298-299: I believe that the conversions of the CIO values to the SIO values are based on the fixed conversion function. If so, the discrepancies in the O_2/N_2 ratio between the assigned values and the measured values suggest that systematic change in the conversion function, which correspond to the change in the CIO scale, or change in the O_2/N_2 values in the Scripps primary cylinders.

The conversion of CIO to SIO is indeed a fixed linear function, based on all of the measurements of the SIO cylinders over time. The differences between the assigned and measured values are minimized in this function, and to our opinion there is no unambiguous indication to assume a change in this function over time.

Page 9, line 303-305: Do the authors mention that the linear conversion function is often calibrated based on the measurements of the Scripps primary standards as shown in Fig. 4?

We have added the fixed function that we use, so the function is not calibrated often, but the tank are measured and have been used in the fixed function spanning the whole period presented in the paper.

Page 10, line 317-318: "Manning et al., 2015" is not in the list of References.

It is now properly changed to Manning et al 2015, in the reference list.

95 Page 11, line 340: Why do the authors refer to the "WMO extended compatibility goal of 10 per meg"? The extended compatibility goal is set for the studies like urban observations that are strongly influenced by local fluxes. I believe that the authors aim to observe the background air through their three sites observation because they evaluated the global carbon budgets based on their observations in Section 5.1. Therefore, I think the authors should refer the "WMO Network compatibility goal of 2 per meg" here.

100 We have adjusted the comparison to the WMO network compatibility.

Page 11, line 349: All of the GOLLUM cylinders show the increasing drift (Fig. 5). Nevertheless, the authors described that the average overall drift rate significantly small (4 ± 6 per meg yr⁻¹). How did the authors calculate the uncertainty of 6 per meg yr⁻¹? In addition, I think that the scale drift rate of 4 per meg yr⁻¹ is not small because it corresponds to bias of 1.6 PgC yr⁻¹ for the carbon budget calculation.

105 The uncertainty of 6 per meg/yr is based on individual drifts of each GOLLUM cylinder. And indeed, the drift is not small, but significantly smaller than the uncertainty 11 ± 18 per meg of Cucumbers. The comparison between GOLLUM and Cucumbers is just to show that there is no clear indication of a significant drift in our scale. We have updated the text accordingly.

Page 11, line 349-350: The WMO compatibility is defined as "a measure of the persistent bias between 110 measurement records". Thus, it should not be compared with the scale drifting rate.

We have changed it, and removed the comparison.

Page 12, line 370-372: If the authors fit a combination quadratic function and three harmonics to the data by using a least square method and do not use a digital filtering method of Thoning et al. (1989), the authors don't need to refer to Thoning et al. (1989).

115 We have referred to Thoning et al. 1989 as the basis of the use of this function, but we added that we do not use the digital filtering

Page 13, line 392-394: I don't understand the reason the exclusion of the last 2 years data. For example, the data at Mace Head in 2017 are much sparser than the data in the last year (2019). Additionally, I cannot accept the authors' idea that the sparse data in the last two years introduce biases in the fits. Since there are enough data to determine the average seasonal cycle for both Lutjewad and Mace Head, the larger number of data, even if the sparse, can cause the better fitting results.

There was a significant problem with our DI-IRMS for the end of 2019 until all of 2020 that affected the quality of our measurements, so the best we can include is the first ³/₄ of 2019.

Page 13, line 403-405: Please see the comment for Page 13, line 392-394.

125 Please see above

120

Page 13, line 415: I think the longest period for the trend calculation is 17-year (from 2002 to 2018).

We have changed it to 17 years now.

Page 18, line 483-486: It should be better to describe the detail of the drying method at Mace Head because there is no description.

130 We have added the drying method at Mace Head

Page 19, line 534-537: For the comparison of the observed results between Lutjewad and Mace Head, it is enough to simply compare their trends. I think there is no need for the authors to examine much about carbon budget calculations.

We think it's still worthwhile to illustrate what information could our data convey, aside from just a long-term trend. Also, reviewer #2 highlights this in the general comments.

Page 19, line 536: The calculation method adopted in this study is not exactly same as that of Keeling and Manning (2014). In Keeling and Manning (2014), the NOAA's global mean CO_2 data was used to evaluate the accumulated CO_2 in the atmosphere and globally averaged annual mean APO estimated from the limited background observations was used to evaluate the change in the APO (not fitted trend line).

140 We have now fixed this to reflect the different method.

Page 20, Figure 10: There is no explanation about the red lines in the figures. In addition, the exact period for each annual average shown by the black dot is unclear. Is the period of the annual average for 2002 from January 2002 to December 2002 or from July 2001 to June 2002?

We have added in a description of the red line. The period of annual average is from January to December of each year.

Page 21, line 583: I cannot understand the meaning of "noisier seasonal amplitude".

We have changed this to seasonal cycles.

Page 21, line 596-597: Does the trend for Weybourne show much faster decrease than that for Lutjewad? It would be better to plot the trend for Weybourne in the figures.

150 WAO data is unfiltered, so there are a lot of non-background data points in this record. We did not manage to update the record to only background conditions as of yet, and are looking for possibilities with the co-author in charge. However, we would not like to postpone the re-submission due to this issue.

Page 22, line 604: The authors described that the CO_2 discrepancies are shown in "the first half of 2016". But I think that the discrepancies are shown in July and, probably, August 2016. It should be clarified.

155 We have fixed this to be more specific.

Page 22, line 603-607: It would be better to show the correlation plot of the differences of the flask CO_2 and O_2/N_2 from the continuous observations. The slope of the scatter plot would give us the information about the origin of the contamination.

$\begin{array}{l} \mbox{There is no continuous } \delta O_2/N_2 \mbox{ measurements at Halley so it is not possible to plot the differences between flask} \\ \mbox{and continuous } \delta O_2/N_2. \end{array}$

Page 22, line 601: I cannot find "NOAA, 2021" in the list of References.

It's there but presented differently since it is a website. It is fixed now.

Page 22, line 618-621: The contamination of the flask samples collected at Halley is clearly shown from the comparison of CO₂ mole fractions of the flask samples with the in-situ continuous data.

165 Yes, we agree, and we changed the text to reflect this.

Ref 2:

- In this paper, the authors present 20 years of observational δ(O₂/N₂) and CO₂ data obtained at three ground-based stations. They also present a detailed description of the calibration procedures of their δ(O₂/N₂) scale over 15 years. The δ(O₂/N₂) scale was confirmed to be stable enough to estimate global ocean and land CO₂ sinks based on the long-term trends in the observed δ(O₂/N₂) and CO₂. It is important to validate the global CO₂ budget, reported by Global Carbon Project, using independent estimations such as those reported in this study. Therefore, the dataset is a valuable contribution to a better understanding of the global carbon cycle.
- 175 However, I have found some issues that need to be addressed before publication. These are listed below. In particular, some of the interpretations of the observational results are unwarranted. I understand that the ESSD is a data journal, but I think a substantial discussion is also recommended in the paper, particularly considering the high impact of the journal.

Thank you very much for your review, we have addressed the comments below.

180 1) Line 61: Tohoku University, Japan should be added as a research organization that continues to make long-term systematic observations of CO₂ and O₂. Goto et al (https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2017JG003845) and/or Ishidoya et al. (https://www.tandfonline.com/doi/full/10.3402/tellusb.v64i0.18964) need to be listed as suitable references.

We have added Goto et al. to the list.

185 2) Line 126: "(Sturm et al., 2004))" should be corrected to read "(Sturm et al., 2004)".

It is now fixed.

190

3) Lines 116–142: The descriptive detail of the flask sampling procedure at each of the sites need to be the same. Information about the models of the pump used, as well as about the flow rates, inner pressures of the flask, drying agents, and usage of an aspirated inlet need to be described for all the sites. If the size of the flask is different at each site, for example, then the size information needs to be given.

We have added additional information and made the description more uniform.

4) Lines 164-180: The measurement precision of $\delta(O_2/N_2)$ for flask measurements is not shown. Is it the same as the long-term standard deviation of 10.2 to 13.5 per meg for cylinder measurements? Please clarify.

Based on the flask data of LUT and MHD, the measurement precision is 7 to 13 per meg, and is now added to the main text.

5) Line 188 and references: "(Tohjima, 2005)" should be corrected to "(Tohjima et al., 2005)".

This is fixed.

6) Line 197: "(van der Laan-Luijkx et al., 2013)" should be corrected to "van der Laan-Luijkx et al. (2013)".

This is fixed.

7) Chapter 3: Examination of the long-term stability of the δ(O₂/N₂) scale presented in this chapter is highly detailed. It ensures reliability of the long-term trends in the observed δ(O₂/N₂). However, I was not able to follow how the authors evaluated the uncertainty in the observed long-term trends caused by the uncertainty of the δ(O₂/N₂) scale. The authors described "Bilbo and Frodo present a minor drift similarly to that observed by our SIO cylinder 7008 (while the other 2 SIO cylinders did not exhibit this behaviour as shown in Sect. 3.2);
and our internal WTs all show no overall drifts, we consider our calibration procedure as sufficient" (lines 355–358). Does this mean that the observed δ(O₂/N₂) values are determined against "the other 2 SIO cylinders" and no uncertainty is considered for the long-term trends in the observed δ(O₂/N₂) associated with the scale's uncertainty? In addition, quantitative information about the uncertainty in the δ(O₂/N₂) scale during the period prior to 2006 is not provided (line 387–394). Did the author consider the scale's uncertainty before 2006 to

210 determine the long-term trends of the observed $\delta(O_2/N_2)$?

The observations are determined against all 3 SIO cylinders. The conversion of the scales between each different period considers the uncertainties of the measurements in each, so therefore it's reflected in the final uncertainty of the flask measurements. The flask measurements before 2006 therefore had larger uncertainties due to the scale conversion and also affected the long-term trend's uncertainties. The scale uncertainties are included in the calculation of the final uncertainty.

8) Lines 379–381: If the larger fraction of discarded measurements at Lutjewad, compared to those at Mace Head, is related to the effects of local sources/sinks as the authors suggest, then not only $\delta(O_2/N_2)$ but also CO₂ would be observed to be more scattered at Lutjewad than Mace Head. Would the authors agree with this? If the scatter is seen only in $\delta(O_2/N_2)$, then it is highly likely that the scatter is due to an artificial fractionation of O₂ and N₂ rather than due to any of the local effects.

We think the referee has misunderstood the text, as the % discarded are for both $\delta O_2/N_2$ and CO_2 , not one of them.

9) Lines 495–499: What is the protective cap made of? If the authors confirmed that a permeation effect was reduced significantly by using the cap, then it is valuable to provide a fuller description. Anyway, I agree with the authors that the permeation effect and incomplete drying are not the causes of the significant difference in the long-term trends between Lutiewad and Mace Head

They are made of glass, we added in this information.

215

220

225

10) Lines 516–524: I think the discussion surrounding the interpretation of the difference in the long-term trends between Lutjewad and Mace Head from the viewpoint of changes in the North Atlantic oxygen ventilation is too speculative. Hamme and Keeling (2008) discussed differences in the interannual variations between the northerm and southern hemispheres in relation to the North Atlantic oxygen ventilation (the authors referred to Keeling & Manning (2014), but the original paper on this topic was published by Hamme and Keeling (2008)). However, since both Lutjewad and Mace Head are located on the European continent, the horizontal atmospheric transport is much faster than the meridional transport. Therefore, I expect the contribution of the North Atlantic oxygen ventilation to the interannual variations observed at the two sites would be similar. Do the authors have any

supporting information to clarify this issue, such as the simulated results using an atmospheric transport model?

It is indeed speculative since we lack the supporting data to confirm this. We did not check this with a model, so we added in the text that it is a potential cause for differences.

11) Line 552: The ER for globally averaged fossil fuel combustion should be calculated using the latest GlobalCarbon Budget data (Friedlingstein et al., 2020)

From the Global Carbon Budget data, the averaged ER of 1.434 is now used.

12) Lines 603–621: I think the argument to conclude that the CO₂ and δ(O₂/N₂) anomalies were most likely caused by a small inwards leak is weak. The CO₂ values at Halley observed by CIO appear to be higher by about 2 ppm than those obtained by UEA and NOAA, so that the corresponding APO decrease is about 10 per meg by assuming biospheric signal. On the other hand, short-term variabilities of δ(O₂/N₂) at Halley appear to be larger than 10 per meg, compared to the data from CIO and UEA, and I cannot distinguish systematic difference between them. Therefore, consistency between the APO from CIO and UEA does not provide enough evidence of the small inwards leak. I suspect the increase in CO₂ measured by CIO may be due to deterioration of CO₂ during the storing period, such as desorption of CO₂ from inner wall of the flask or some other effects. I would like to hear the authors' thoughts on this.

If we had he continuous $\delta O_2/N_2$ data at Halley, then this would be immediately clear. However, since we do not have that, we depend on the known CO_2 of both records (continuous and flasks) and the known $\delta O_2/N_2$ of the flasks to see the effects on APO. We do agree that the short-term variations may have masked the signals of the suspected leaks, however due to the lack of information, this conclusion appears to be the most probable cause. What we do know is, the effects of storing are not the cause, because we performed storage tests and they show

255 What we do know is, the effects of storing are not the cause, because we performed storage tests and they show remarkable quality over long period of time: we found a negligible drift of 0.4 per meg in $\delta O_2/N_2$ after 48 months; and a drift of -0.3 ppm in CO₂ after 24 months, on a set of 20 flasks. These numbers would only amount to biases of 0.008 per meg /month in dO2/N2 and 0.013 ppm/month in CO_2 so that is why leaks are the most probably cause.

260

Ref 3:

In this manuscript Nguyen et al. present ~20 years of observations of atmospheric oxygen and carbon dioxide from two Northern Hemisphere flask sampling stations and 3 years of observations of the same from an Antarctic station. The authors are presenting a substantial dataset of value to the greater O2 and CO2 communities. The data are openly accessible from the ICOS Carbon Portal, and the files are self-explanatory (with one minor exception). The manuscript is of good quality and I recommend publication after in ESSD after

Thank you for your review, we address the comments below.

addressing some of the minor points I raise below.

270

275

265

MAJOR COMMENTS

1. Missing details: Some of the important details pertaining to the measurements are not presented. The reader is instead pointed to some relevant citations. This forces the data user to read this paper, plus 2-3 more, including a PhD thesis. I think reproducing some of the key details here would be a service to the reader, particularly details on the flask design/shape, and more information about the mass spec analysis. I am also surprised there are no relevant changes to mention during 20 years of sampling in either sample collection, analytical approach, gas handling, storage, etc, or changes to the automatic flask sampler. Also useful would be details on the different

tanks used. Maybe I missed this, but I did not see anything about the valve type and seal, volume, interior, etc.
 Could the authors include some kind of change log, or table of notable events? If there is truly nothing to
 mention, I applaud the authors' consistency over 20 years of sampling!

We have added more information about the sampling procedure at the stations, and the measurements with the DI-IRMS. As for the plumbing diagram and design of the flasks, we think it is better to refer to the cited papers, since they provide detailed information about them.

Uncertainty and data quality: There is not much in the way of uncertainty analysis or constraint. I would say this is the biggest shortcoming of this paper. The reader is left with no real guidance as to how to assess the uncertainty in the individual measurements, or, perhaps more importantly, the trend. It does not seem like they have a good handle on the uncertainty due to primary tank drift (e.g. Keeling et al 2006 Tellus 59). A full uncertainty analysis may be out of scope, but they could at least put some constraints with the data at hand on the long-term trend and the reproducibility of a given flask measurement. As I see it, if someone wanted to use
 these flask records in some kind of analysis, this paper would be the main source of guidance. As such, I hope the authors can provide a bit more help in how an interested user could constrain the uncertainty of the

the authors can provide a bit more help in how an interested user could constrain the uncertainty of the measurements.

We have now added the uncertainties of the flask measurements and the effects on trends.

3. Supporting data: The authors are publishing sample time, sample height, and analyte concentration, but further data on the flasks is not included: analysis date, fill pressure, average flow rate, temperature data for sampling and analysis if it exists, etc. This supporting data would be helpful for anyone interested in further QA/QCing or using the data. The authors also do not include the other species used to filter/select the samples for background conditions. I suggest the authors should also seriously consider supplying the non-background data with flags, instead of only the background samples, and the CO/Radon data used to filter them. Or, at least provide DOIs as to where one could find it.

It is indeed a good idea to include the full raw data, we will do so in the coming time.

MINOR COMMENTS

Data files: It is not clear to me from the paper or the header what exactly the standard deviation column
 represents in the data files. I suggest the authors add this to the header, or put it in a subsection in the paper describing the files.

We have added the extra information in the updated data files.

L17-18: better to provide a metric here than to use the subjective "high-quality". Also, I am not sure if intercomparisons tell us anything about the quality of the calibration--all of the labs could be making the same mistake.

310 mista

That is true, but if at least we can show some consistent results, then it's better than all labs showing very different values.

L19: suggest striking the "internationally-recognised" for the sake of brevity.

It is changed to "international"

315 L25: Compatible can only be assessed if two measurements are made on the same air (tank or background), so I don't think it is correct to say that seasonal cycles are compatible if measured at different locations. Better to say they are in good agreement.

Indeed, it is changed to "in good agreement" now.

L40 - "a strong aide" -- Curious wording

320 Valuable is now used instead.

L50-63: suggest cutting this entire paragraph up to "Our Laboratory...", and combining with the next paragraph.

We will leave it as it is now, since it highlights the importance of O₂ measurements, and shows previous work on this topic for reference.

L81: What does the "(formerly)" mean? It used to be called this but the name has changed, or it's not operational now? Please clarify.

Indeed it had to cease operation in 2016/7 and moved to a new location due to a crack on the ice shelf.

L94: Nafion driers are not very common in O2/N2 measurements. Does Nafion fractionate O2/N2? If the authors have tested this, I would encourage them to include such results here (or provide a citation).

Except for water, all other species should have negligible gradient over the Nafion membrane, since we supply
 the outer side of the nafion drier with the exhaust of the system. It is therefore unlikely that anything can cause fractionation of the δO2/N2 values.

L115: Please provide a plumbing diagram(s) of the flask samplers.

Information is added, but as for the plumbing diagrams, we would like to refer readers to Neubert et al 2004 (cited) for much more detailed information.

335 L116: Could you include a drawing or picture of one of the flasks? Do they have dip tubes?

We have added the information - the flasks have dip tubes.

L123: Dried to what dewpoint? Please include specifications on the cryotraps.

This is already described in line 97, but I also added it in line 123.

L124: What is the flow rate during flask sampling?

340 This information has been added now.

L124: Atmospheric pressure varies, please give exact fill pressure with observed range.

We have now stated that flasks are filled to "current atmospheric pressure". The exact values vary slightly, but the flasks are always kept at the same pressure as the atmospheric pressure in the laboratory.

L131: Please give full details on sampling protocols for Mace Head.

345 This information has been added now.

L170: "Relatively very stable" -- ambiguously worded

Relative to most other gases (except noble gases), N2 is very stable. We have removed relatively.

L191: The influence of fossil fuel burning on APO is not small--that is why there is a large trend in APO.

"Small" is now omitted.

350 L210: Is there a systematic difference between first, second, and third analyses? Why sometimes 2 and sometimes 3?

There are no systematic differences between the duplicates. Usually there are 2, but sometimes when there is an obvious problem with one of the duplicates, we perform a third analysis. We added this to the text.

L215: I don't fully follow -- you are assigning the WT a value and then assigning flasks a value based on
 comparison with the flask? Or flasks are assigned values from the MREF and then corrected for long-term drift through the WT?

Yes, we measured flasks first as a difference against the MREF, then the MREF are used in combination with the WT to correct for the long-term drifts + changes in scales of the MREF (as can be seen in Fig.3 panel 1, showing the WTs through different MREFs).

360 L249: From Figure 3 it looks like some of the drift is not well-described by an average drift rate. Can the authors comment on this?

Yes, indeed there are still some small drifts not entirely corrected. We tried our best to eliminate as much drifts as possible but unfortunately some small periods are not as well-defined, which is not fully satisfactory.

L264: I am still a little confused about how values are assigned. The WTs are given a value based on the MREF,and then Equation 3 is applied to the flask samples? If so, wouldn't the WTs by definition have to be stable? Ordo you mean that they are stable relative to one another? Do the authors have a comment as to why 4845 is so variable?

Yes, both the WTs and the MREFs are stable, however due to many potentials that could cause drifts (analyser drifts mostly), the "measured" values are not, hence we first corrected for all these drifts presented in a sample
by relating all of these changes against a baseline that we chose as our internal baseline scale (i.e. the CIO scale), then from that we convert the measurements of the samples into the SIO by a direct connection that we established between CIO and SIO scale by calibrating with the Scripps cylinders. As for why 4845 was so variable, we think it might be related to the very low value of the cylinder, which suggests potentially contaminated air inside the cylinder or small leakages in the pressure reducer during measurements.

375 L275: I think this is a little misleading, since changing MREF cylinders leads to large offsets in the record. I agree that generally based on Figure 3 the scale looks stable after the correction, but as I understand what the authors describe they are blind to WT drift. Or?

The WT drifts are already shown in the raw data of the WTs vs MREFs, and we corrected for the drifts by individually separating the record into each individual MREF period, and dividing even smaller within those, to correct for the WT drifts.

L290: But the primaries look systematically low (7002 and 7003), and 7008 shows clear drift. I would strike this sentence ("The ensemble thus suggests...")

If there was a systematic error in our calibration, it would show in all cylinders, and we therefore decided to keep this sentence.

385 L305: One primary is clearly drifting relative to the other two, does it really make sense to include this tank in the ensemble? Also, one would expect cylinders to drift over time. If possible, these effects should be accounted for in an uncertainty analysis.

The drift, while noticeable when comparing to the other 2, only amounts to 1.4 per meg per year. For now, we decide to include it, but in the future, we may purchase new primary cylinders.

390 L350: It shows drift in your scale only if the GOLLUM cylinders are not drifting. It could be that 7008 is stable and the other two are not, or that all the cylinders are drifting together and 7008 is drifting slightly less or more than them. Without absolute constraints, it is unclear.

That is indeed true that we cannot be 100% sure, but they are the only standards that we have.

L404: "exact multitude of years" -- what does this mean?

395 It means complete calendar years, i.e. from Jan to Dec.

L465: I think this seasonal cycle section (and section 4.1) is perhaps out of scope for the journal. Suggest to cut.

We politely disagree with this, because the presentation of the data in more details including the seasonal cycles is worthwhile for this journal. We are not sure if the reviewer really means section 4.1, since that section with the main presentation of the data.

400 L486: "has been under much closer controlled thanks" -- should read e.g. "has been more closely controlled"

It is now fixed.

380

L492: I do not know what a valve cap is. Surely it is the o-ring which causes the seal? Why would permeation through the o-ring be impacted by an external cap?

The valve cap is just an additional cap (with O-ring) to lessen the potential permeation through the o-rings, as it forms a small buffer volume between flask and outside.

L496: This is great to see, could you include some actual figures or numbers here?

Answer here + The information will be added

L500: I find it hard to believe the trend could be impacted by sampling bias, particularly since Mace Head is sampled/filtered for background conditions.

410 Not biases in sampling condition per se, but the actual sampling procedure.

L505-510: The decadal trend in APO should be virtually identical between two European background sites, and can't be explained by country-level differences in fossil fuel use. The authors acknowledge this on L509-510. Suggest this be cut.

We agreed, and moved this point into the potential list as suggested below.

- 415 L516: I agree that North Atlantic ventilation signals are likely to be present, but there are numerous other possible causes contributing to the different trends visible in APO between the two sites. I think it would be better to simply state in a sentence or two that the authors do not know the cause of the difference between the two stations, and that the list of possible explanations include: changing continental fossil fuel signals, shifts in atmospheric transport, different degrees of sensitivity to North Atlantic ventilation, other possible causes, or an
- 420 artifact(s) in the data. I think it's important to acknowledge the last one here. I also wonder how much of the curvature at MHD is due to the fit itself--maybe the annual averages aren't actually that different? On this last point, I think calculating the terrestrial and oceanic sinks is out of scope for this paper. I suggest cutting this section and the figures.
- Yes, we agree that it is better to suggest a list of causes since there is no conclusive answer for the discrepancies. As for the terrestrial and oceanic sinks, we still want to give them, but then for Lutjewad only, again to illustrate the possible use of the data.

Section 5.2: I also think this is out of scope. The comparison of seasonal amplitudes between sites tells us really nothing about the quality of the data, since we expect there to be station to station differences. It also seems odd to me for an ESSD paper to briefly present data not pertaining to the dataset being published, such as the Weybourne or Halley measurements by other groups. Suggest cutting the entire section.

We would like to politely disagree, since the additional illustrations are still interesting to see. They show what potential information the data carry within them, and how inter-laboratory comparisons are for HAL.

Figure 1: Three more panels showing the local site (e.g. satellite, street view, or topography) would be helpful here.

435 The most important information about the stations are already included, so we would like to omit these from the figure.

Figure 6 and 7: It is difficult to see the actual data because the fitted curves are on top of the points. I would suggest removing the curves completely and let the data speak for itself. I would also suggest zooming in on the CO2 data a bit more.

440 We feel the fits really add value to the plots, so we would like to keep them as they are.-We chose the scale of CO₂ is to match the scale of δO₂/N₂, to a relative ratio of 1 ppm per 4.8 per meg, so that changes are comparable. We prefer to include the fits.

445

430

Two decades of flask observations of atmospheric $\delta(O_2/N_2)$, CO₂, and APO at stations Lutjewad (the Netherlands) and Mace Head (Ireland), and 3 years from Halley station (Antarctica)

Linh N.T. Nguyen¹, Harro A.J. Meijer¹, Charlotte van Leeuwen¹, Bert A.M. Kers¹, H.A. (Bert) 455 Scheeren¹, Anna E. Jones², Neil Brough^{2,*}, Thomas Barningham², Penelope A. Pickers³, Andrew C. Manning³, and Ingrid T. Luijkx⁴

¹Centre for Isotope Research, Energy and Sustainability Research Institute Groningen, University of Groningen, Groningen, the Netherlands

460 ²British Antarctic Survey, Natural Environment Research Council, Cambridge, United Kingdom ³Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, United Kingdom ⁴Meteorology and Air Quality, Wageningen University and Research, Wageningen, the Netherlands *Now at National Institute of Water and Atmospheric Research, Wellington, New Zealand

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

> Abstract. We present 20-year flask sample records of atmospheric CO₂, $\delta(O_2/N_2)$ and APO (Atmospheric Potential Oxygen) 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 less than 3 per meg over the 14 years of calibration, and our

- 470 compatibility with the international Scripps O2 scale, The measurement records from Lutjewad and Mace Head show similar long-term trends during the period 2002-2018 of 2.31 ± 0.07 ppm yr⁻¹ for CO₂ and -21.2 ± 0.8 per meg yr⁻¹ for $\delta(O_2/N_2)$ at Lutjewad, and 2.22 ± 0.04 ppm yr⁻¹ for CO_2 and -21.3 ± 0.9 per meg yr⁻¹ for $\delta(O_2/N_2)$ at Mace Head. They also show a similar $\delta(O_2/N_2)$ seasonal cycle with an amplitude of 54 ± 4 per meg at Lutjewad and 61 ± 5 per meg at Mace Head, while the CO₂ seasonal amplitude at Lutjewad (16.8 ± 0.5 ppm) is slightly
- 475 higher than that at Mace Head (14.8 ± 0.3 ppm). We show that the observed long-term trends and seasonal cycles are compatible with the measurements from various other stations, especially the measurements from Weybourne Atmospheric Observatory (United Kingdom). However, there are remarkable differences in the progression of annual trends between the Mace Head and Lutjewad records for $\delta(O_2/N_2)$ 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 $\delta(O_2/N_2)$ 480 and APO, where especially APO agrees well with continuous measurements at the same location made by the University of East Anglia, while CO_2 and $\delta(O_2/N_2)$ present slight disagreements, most likely caused by small leakages during sampling. From our 2002-2018 records, we find good agreement for the global ocean carbon sink: 2.0 ± 0.8 PgC yr⁻¹ and 2.2 ± 0.9 PgC yr⁻¹, based on Lutjewad and Mace Head, respectively. The data presented in 485 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 Deleted:

Deleted: The results of our inter-comparison involving gas cylinders from various research laboratories worldwide also show that our calibration is of high quality and compatible with the internationally-recognised Scripps O2 scale

530

determination of anthropogenic CO2 emissions and their partitioning across different reservoirs plays a vital role 495 in understanding the impact of the remaining atmospheric CO2 mole fraction on climate (Friedlingstein et al., 2020). High-precision atmospheric O2 measurements have been proven to be valuable in quantifying CO2 fluxes in the carbon cycle. By combining the decadal trends of atmospheric CO2 and O2, 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 CO₂ and O₂ cycles are closely coupled - in most processes, there is an anti-correlation 500 in the changes of their mole fraction, except for the oceanic uptake of CO2 (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 CO2 and O2 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 505 (Keeling and Manning, 2014).

There are various techniques to measure atmospheric O_2 to 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; Stephens et al., 2021); and fuel cell technology 510 (Stephens et al., 2007). Despite many improvements to these techniques over the years, it is still very challenging to obtain O2 measurements with high accuracy and precision. This is mainly because the atmospheric background mole fraction of O₂ is very high – around $209,392 \pm 3$ ppm (Tohjima et al., $2005)_{v}$ – 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 515 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 CO2 and O2 measurements, for example, the

2006); National Institute of Advanced Industrial Science and Technology (Aoki et al., 2021); National Institute 520 for Environmental Studies (Tohjima et al., 2008); Tohoku University (Goto et al., 2017); University of East Anglia (UEA) (Pickers et al., 2017); and the University of Groningen (van der Laan-Luijkx et al., 2010). Our laboratory - the Centre for Isotope Research (CIO) of the University of Groningen (RUG) in the Netherlands - has been carrying out flask measurements of CO2 and O2 since the early 2000s from various locations (van der Laan-Luijkx et al., 2010). Flask sampling for CO2 and O2 has been conducted at Lutjewad (the Netherlands), Mace Head 525 (Ireland), Jungfraujoch (Switzerland) and Halley (Antarctica).

network of atmospheric stations maintained by the Scripps Institution of Oceanography (Manning and Keeling,

In this paper, we present the O2 and CO2 measurements from flasks collected at Lutjewad (the Netherlands), 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).

Deleted: a strong aide

Deleted: (Tohjima, 2005)

2 Methods

2.1 Site description

540

The stations from which our flasks were collected are: Lutjewad Atmospheric Monitoring Station on the northern coast of the Netherlands (53°24'N, 6°20'E) managed by the CIO (RUG); Mace Head Atmospheric Research Station on the western coast of Ireland (53°20'N, 9°54'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°34'S, 25°30'W) operated by the British Antarctic Survey. Halley station has been relocated later due to that part of the ice shelf breaking off. Figure 1 shows the locations of the three stations.

Deleted: and (formerly) Halley VI Research Station situated on the Brunt Ice Shelf (75°34'S, 25°30'W)





Figure 1: Left panel: Locations of the Mace Head (red) and Lutjewad (orange) stations. Right panel: Location of the Halley station (blue)

550

555

565

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 southwest, 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 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

560 and continental sources (Derwent et al., 2002; Jennings et al., 1993).

The Halley station is a "Global" station within the World Meteorological Organisation's Global Atmosphere Watch (WMO/GAW) programme, that observes background atmospheric conditions at various locations around the globe. The main Halley station consists of 8 modules that are atop ski-fitted hydraulic legs, within which are the research facilities and living quarters. Air sampling for this project was carried out at the Clean Air Sector

Moved down [1]: 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) 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 dewpoint below $45 \,^{\circ}$ 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 ClO laboratorics (Neubert et al., 2004)

Deleted:

585

Laboratory, which is located 1.5km from the main station in 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 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 590 laboratory building. The inlet is equipped with a Nafion drying tube (MD 110-72-S, Perma Pure, Toms River, New Jersey) 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 dewpoint below -45 °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 595 system for further analyses in the CIO laboratories (Neubert et al., 2004). For storing air samples, we use 2.5-litre glass flasks with dip tubes, capped with two high-vacuum valves (Louwers, Hapert, NL) sealed with Viton orings (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 600 diaphragm pump (KNF N811), flow controllers, and magnetic solenoid valves). The autosampler schedule is controlled via custom-made software (written in 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 dewpoint of -45°C) and flushed through a flask for at least an hour at 2.5 L min-1 before filling the flask slowly so that the sample 605 remains at current atmospheric pressure (to prevent the sample from fractionation and differential permeation through the o-rings cause 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 222Radon activity monitored at the station 610 was less than 3 Bq/m³ 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 different setup of the stations, the drying methods are different, and only Halley station has an aspirated inlet.

615

620

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, then flushed through the flasks for about 30

(Moved (insertion) [1]

Deleted: gas

minutes at 2.5 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.

625

At Halley, flasks are collected once a week depending on the meteorological conditions, via a portable manual sampler. This consists of a KNF diaphragm pump (KNF N86), flowmeter, drying agent (magnesium perchlorate), 7μ m filter and 3 sampling flasks connected in concession. The air is sampled about 6 metres 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 minutes at a flow rate of 2.5 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 the Antarctic supply ship.

635

630

After sample collection, flasks from the three stations are transferred back to our laboratory in Groningen for analysis. Typically, the mole fractions of CH₄, CO, CO₂ and O₂ (reported as δ(O₂/N₂), see next section) are measured (van der Laan et al., 2009), and additional analyses such as stable isotopes (for example ¹³C and ¹⁸O in CO₂) and radiocarbon (¹⁴C in CO₂) are also conducted when required (van der Laan et al., 2010).

2.3 CO₂ measurement

All flask samples are analysed on an Agilent HP6890N gas chromatograph equipped with a Flame Ionization
Detector (referred to as HPGC) to determine the mole fractions of CO₂, CO and CH₄. 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 5 primary standards linked to the World Meteorological Organization (WMO) X2007 scale with CO₂ ranging between 354 and 426 µmol 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 CO₂, CO and CH₄ mole fractions on a Cavity Ring-Down Spectrometer (CRDS) model G2401-m from Picarro Inc. using the same suite of primary standards. We refer to Chen et al. (2010) for more details on the CDRS technique. The measurement precision and accuracy for flask measurements of CO₂ on the HPGC are typically

650 <0.06 ppm and <0.07 ppm, respectively (van Leeuwen, 2015).

All CO₂ 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. (2020)).

2.4 O₂ measurements

Atmospheric O₂ is typically reported as the $\delta(O_2/N_2)$ value. The $\delta(O_2/N_2)$ value of a sample is calculated as the difference between the O₂/N₂ ratio of the sample and that of a reference gas (Keeling and Shertz, 1992):

$$\delta(O_2/N_2) = \frac{(O_2/N_2)_{sample} - (O_2/N_2)_{reference}}{(O_2/N_2)_{reference}}$$

(1)

Deleted: The filter to retain flasks with a CO mole fraction of less than 200 ppb is also applied for flask samples from Mace Head.

Since for natural variations, $\delta(O_2/N_2)$ 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 O2 is reported as O2/N2 ratio because it is not a trace gas, and its mole fraction is thus affected by changes in other atmospheric constituents 665 such as CO2. Atmospheric N2 is very stable (Keeling et al., 1998), therefore changes in the O2/N2 ratio would reflect mostly the changes in atmospheric O2 (only in a detailed budget analysis minor N2 variabilities are still considered, as described in Keeling and Manning (2014)). For $\delta(O_2/N_2)$ 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 670 measurement comprises sixteen 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, to account for the disturbances in the signals caused by the switching of the valves that affect the measurement precision (Sirignano et al., 2010). Due to the sensitivity of the analyser, it is located inside a climate-controlled

675 room in our CIO laboratory, However, it is inevitable that the measurements still drift over time. To correct for the 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 $\delta(O_2/N_2)$ 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-680 Luijkx et al. (2010), and are summarised in Sect. 3.

685

2.5 Atmospheric Potential Oxygen (APO)

Combining highly precise measurements of atmospheric CO2 and O2 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 Atmospheric Potential Oxygen (APO). The APO value of an air sample is determined by combining its δ(O₂/N₂) and CO₂ measurements (Battle et al., 2006; Gruber et al., 2001; Stephens et al., 1998):

$$\delta APO = \delta(O_2/N_2) + \frac{1.1 \times (CO_2 - 350)}{2}$$

(2)

690

The value of 1.1 represents the mean O2:CO2 ER of terrestrial ecosystems (Severinghaus, 1995); for the So2, wet take 0.2094, which is the standard atmospheric O2 mole fraction (Tohjima et al., 2005); and 350 is the consensus (arbitrary) reference value to be subtracted from the measured CO2 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 mainly captures the seasonal and long-term air-sea exchange of CO2 and O2, with an influence from fossil fuels combustion, caused by their higher average ER of ≈ 1.4 (Pickers et al., 2017; Sirignano et al., 2010),

3 Calibration of the DI-IRMS

695

Deleted: relatively

Deleted:

Deleted: however the measurements still inevitably

Formatted: Space Before: 0 pt, After: 0 pt, Line spacing: 1.5 lines

Deleted: small Deleted:

Deleted: 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 made 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.

705 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 ("change overs"). The result of this process is the $\delta(O_2/N_2)$ value of the sample, as presented in equation 1. Each individual measurement is based on seven successive pairs of sample and reference measurements, which are used to calculate seven delta values

- 710 (equation 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
- 715 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 $\delta(O_2/N_2)$ 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 $\delta(O_2/N_2)$ values of these repeated measurements (van der Laan-Luijkx et al., 2010). The precision of the 720 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

725

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. The summary of different WTs and MREF cylinders used from 1998 to 2020 is shown in Fig. 2 and Table 1.

Deleted:

Deleted: and t

Deleted:



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 colour are adapted from the work of van der Laan-Luijkx (2010).

735 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.
740 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).

745

750

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). 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 greved-out rows are the past cylinders used prior to this work, but included here to demonstrate a complete record.

		Period	MREF	WTs	
	1	17-08-1998 - 18-02-2000	2534	4497	
Previous	2	19-06-2000 - 17-11-2003	4497	4446	
	3	03-02-2004 - 18-02-2004	4446	8780	
	4	18-02-2004 - 14-05-2004	8780	4446	

	5	04-06-2004 - 19-10-2005	4866	2534 7512 8780
	6	30-01-2006 - 30-12-2006	6170	2534 6987
	7	30-01-2007 - 30-12-2007	6170	5279 6096
				6168 6987
	8	30-01-2008 - 15-12-2010	6170	5279 6096 6168
	9	03-01-2011 - 11-03-2014		5279 6096 6168
	10	11-03-2014 - 29-08-2015	6185	5279 6096 6168
	11	30-08-2015 - 10-06-2016		5279 6168
Current	12	11-06-2016 - 05-05-2018		5279 6168
	13	06-05-2018 - 01-01-2019	6123	5279
	14	02-01-2019 - 11-03-2020		5279 4845
	15	12-03-2020 to present		5279 4845

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):

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

Where:

755

760	-	$\delta_{S/R}$ is the $\delta(O_2/N_2)$ value of the sample against the CIO 2534 scale
100		ogra is the o(02/102) value of the sample against the CIO 2551 searce

 (δ_{M/R})_{sub-period} is the average δ(O₂/N₂) 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-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.
- 765 $\delta_{S/M}$ is the $\delta(O_2/N_2)$ value of sample against the MREF cylinder (raw value).

by van der Laan-Luijkx et al. (2010).

The final step is to transform the δ_{SR} 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 equation 2, 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

(Deleted: 3 and conversion to the SIO scale

20

⁷⁷⁰



785

790

795



Figure 3: Measurements of the 3 long-term WTs (5279, 6096, and 6168) for periods 7-15 (Table 1), across the final 3 MREF periods plus a recently added WT (4845). Top panel: raw measurements of the WTs against different MREF cylinders. Bottom panel: 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 in parentheses are the respective standard deviations. All numbers are in per meg. Visible gaps in the data are due to instrument issues, maintenance or instrument relocation.

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, and the weighted mean slope amounts to -0.4 ± 0.7 per meg yr⁻¹, so not significantly different from zero. To check the stability of our scale over time, we calculated the year-to-year variability between the WTs, and found a value of less than 3 per meg over the 14 years of measurements.

WT 4845 was recently measured for a relatively short period only, and appeared to be less stable and noisier compared to 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 a 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 3 WTs also showed no systematic drifts across different MREF periods (Table 2). For WT 5279 and WT 6096, there are no significant changes (at least to ±0.3 per meg) between the MREF periods 6170 and 6185, although there is 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, then dropped slightly (by 0.5 per meg) in MREF 6123 period. The stability demonstrated in both long-term measurements and per MREF periods consolidates the quality of our calibration procedure.

Table 2: Comparison of the WTs over 3 different MREF cylinder periods. The values (in per meg) are averaged over the corresponding period, accompanied by the standard errors. The N/A values in the MREF6123 period for WT6096 **Deleted:** After these adjustments, the measurements of the 3 long-term WTs (5279, 6096, and 6168) show that they were relatively stable over time, all showing standard deviations of 10.2 to 13.5 per meg (corresponding to standard errors of 0.09 to 0.14 per meg) over the 20-year period. WT 4845 was recently measured for a relatively short period only, and appears to be less stable and noisier compared to WT 5279 measured in the same period. Thus, WT4845 was not used for the calculations in the calibration procedure, and its measurements are only shown here for completeness.

Deleted: 1

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

815 are due to its discontinuation in this period. The Difference column is calculated by subtracting the values of the old MREF periods from the new ones

3.2 Quality check of the Scripps primary cylinders

820

The final check on the quality of our scale is the regular measurement of the 3 Scripps primary standard cylinders that we purchased from SIO, numbered 7002, 7003, and 7008. These measurements were conducted at least once a 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. 4). 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 the measurements of the primary cylinders.

825 <u>pri</u>

In Figure 4, 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 Figure 4 and Table 3 that cylinder 7008 exhibits a small upward drift over time of 1.4 ± 0.4 per meg 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.

Deleted:

The conversion of the CIO scale to the Scripps scale is <u>calibrated using these Scripps primary standards</u> <u>measurements</u> and in such a way that the ensemble difference between the assigned values and weighted averages of our measurements of three Scripps cylinders is minimised (Mook, 2000); $\delta(O_2/N_2)_{SIO} = \delta(O_2/N_2)_{CIO} * 0.999544 + (0.999544 - 1) * 10^6 + 1.4$

845

(4)

Where $\delta(O_2/N_2)_{SIO}$ and $\delta(O_2/N_2)_{CIO}$ are the $\delta(O_2/N_2)$ 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).



850

Figure 4: 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.

Table 3: Comparison of the averaged measured values of the Scripps primary standards against their assigned values855in per meg.

Cylinder ID	7008	7002	7003

Formatted: Font color: Text 1

Assigned by SIO	-254.3	-465.0	-791.6
Weighted mean measured	-245.9	-468.9	-797.6
Standard deviation	8.0	6.0	11.8
Standard Error	1.9	2.0	2.8
Deviation from assigned	8.4	-3.9	-6.0

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 initialised in the European Union's CarboEurope project and coordinated by the UEA (http://cucumbers.uea.ac.uk/); 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.

865

870

875

860

The Cucumber programme involves inter-comparison of nine atmospheric species (of which $\delta(O_2/N_2)$ is one) between atmospheric research stations in Europe and a number 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).

880

Figure 5 shows the measurements of the Cucumber cylinders (top two panels), the cylinders in the Bilbo and Frodo rotations of GOLLUM (third and fourth panels, 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 (bottom panel). 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 size, suggesting that there is not a systematic scaling error but rather individual variations 885 between cylinders and/or measurement periods. Due to the individual variations, the overall drifts for Cucumber cylinders is 11 ± 18 per meg 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)

Deleted: extended Deleted: 10

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.

895

900

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 variations between years (varying within a range of less than 20 per meg), and also significantly smaller overall drift over the duration of the measurements (4 ± 6 per meg yr⁻¹). However, all 6 cylinders appear to drift in similar direction, suggesting a drift (however small) 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 GOLLUM cylinders, whereas the two other SIO cylinder do not (Fig. 4).

Since the cylinders show an inconclusive "drift": 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 2 SIO cylinders did not exhibit this behaviour as shown in Sect. 3.2); and our internal WTs all show no overall drifts, 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.

Deleted:, which is within the extended WMO compatibility goal



Figure 5: Cylinders from the Cucumber programme (top 2 panels) along with two sets of three cylinders in the GOLLUM programme (middle 2 panels) and 3 internal CIO cylinders (WT 5279, WT 6096 and WT 6168) and a primary standard cylinder at CIO (SIO 7008) (bottom panel). 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 on all panels.

3.4 Treatment of analysed flask samples

915

- 920 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 3-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
- 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, around 30% of
the flasks were excluded from further analyses from Lutjewad samples, 16% from Mace Head samples, and only 6% from 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 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 δ(O₂/N₂) and CO₂ measurements (equation 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 when there is little information to connect the following period to the first period (as presented in Fig. 2 and Table 1). Next to this, 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 until the whole 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 on 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 (less than 3 per meg in 14 years) as well as the Scripps primary standards (8.6 per meg in 10 years). Although some drift is observed in one of our Scripps cylinders, the other two have remained stable within uncertainty. The same inconclusive picture emerges from our various sets of cylinders in the inter-comparison programmes. Therefore, we conclude that our calibration process is accurate within the uncertainties mentioned above.

4 Flask measurement results

4.1 The CO₂, δ(O₂/N₂) and APO records

960

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 $\delta(O_2/N_2)$ and CO_2 , with some differences in APO variations. Figures 6 to 8 show the CO_2 , $\delta(O_2/N_2)$, 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 3-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 the first and last two years out, to make sure that the fit period comprises <u>complete</u>

965

Deleted: an exact multitude of years

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 CO2, the only other contributing factor is the uncertainty in the trend fit. For $\delta(O_2/N_2)$, 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.

970

CO2 measurements at Lutjewad and Mace Head show a positive, and increasing trend over 20 years. Due to the 975 quadratic trend fit, the growth of the fitted increase is linear. The trend (given here in ppm yr¹ with their 95% confidence interval (CI) uncertainties) in Lutjewad grows from 1.81 ± 0.10 ppm yr⁻¹ in 2002 to 2.27 ± 0.03 ppm yr⁻¹ in 2010 and 2.74 ± 0.10 ppm yr⁻¹ in 2018. These values agree relatively well with the globally averaged values as measured by the NOAA's Global Monitoring Laboratory: 1.86 ± 0.20 ppm yr⁻¹ in 2002, 1.97 ± 0.14 ppm yr⁻¹ in 2010, and 2.57 ± 0.19 ppm yr⁻¹ in 2018 (<u>https://gml.noaa.gov/ccgg/trends/global.html</u>). The values from NOAA 980 are calculated based on a 5-year average around the time marks 2002, 2010 and 2018. In all three periods, the values at Mace Head are also in agreement with those of Lutjewad $(1.86 \pm 0.06 \text{ in } 2002, 2.24 \pm 0.02 \text{ in } 2010, \text{ and } 1.2010)$ 2.63 ± 0.06 ppm yr⁻¹ in 2018 for Mace Head). When averaging the trends over the 20-year period, both stations show good agreement with each other and with the global average: 2.31 ± 0.07 ppm yr⁻¹ for Lutjewad, 2.22 ± 0.04 ppm yr⁻¹ for Mace Head, and 2.1 ± 0.3 ppm yr⁻¹ for global. The total uncertainty of the trend is 0.07 ppm yr⁻¹ for 985 Lutjewad and 0.04 ppm yr⁻¹ for Mace Head. The largest contributing factor to the total CO₂ long-term trend uncertainty is from the trend fits.

 $\delta(O_2/N_2)$ measurements at Lutjewad also show a clear trend that becomes increasingly more negative throughout the 20 years. The trends (reported here in per meg yr-1 with their 95% CI uncertainties) in 2002, 2010, and 2018 990 are -18.01 ± 1.17 per meg yr⁻¹, -20.99 ± 0.29 per meg yr⁻¹, and -23.98 ± 1.17 per meg yr⁻¹, respectively. At Mace Head, we find an unexpected trend: while CO₂'s trend increases, that of $\delta(O_2/N_2)$ becomes less negative (-22.4 ± 1.3 per meg yr⁻¹, -21.2 \pm 0.3 per meg yr⁻¹, and -20.0 \pm 1.3 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 995 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 per meg yr⁻¹ for Lutjewad and -21.3 \pm 0.9 per meg yr⁻¹ for Mace Head. The total uncertainty of the trend is 1.3 per meg yr¹ for Lutjewad and 1.5 per meg yr¹ for Mace Head. The largest contributing factors to the total 1000 $\delta(O_2/N_2)$ long-term trend uncertainty for Lutjewad are equal between the CIO scale stability (13.5 per meg in 14 years) and the uncertainty in the trend fits. For Mace Head, the uncertainty in the trend fits is the most significant factor. 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 $\delta(O_2/N_2)$ and 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 $\delta(O_2/N_2)$ and CO_2 measurements, it shows a combination of the patterns as illustrated in the two species. The APO trend (reported here also in per meg yr-1) at Lutjewad does not differ significantly over time, varying from -9.4 ± 0.8 per meg yr¹ in 2002 to -9.31 ± 0.20 per meg yr¹ in 2010, and -9.3 ± 0.8 per meg yr¹ in 2018. In Mace Head, however, the same pattern as $\delta(O_2/N_2)$ is shown for APO: the trend gets significantly less negative throughout the period (-13.15 \pm 1.20 per meg yr⁻¹ in 2002, -9.5 \pm 0.3 per meg yr⁻¹ in 2010, and -5.83 \pm 1.20 per meg yr⁻¹ in 2018). The total uncertainty of the trend is 1.0 per meg yr⁻¹ for Lutjewad and 1.3 per meg yr⁻¹ 1 for Mace Head, and the largest contributing factors are the same as for $\delta(O_2/N_2).$

Measurements at Halley station show a similar trend as Lutjewad and Mace Head, where CO2 increases over time while $\delta(O_2/N_2)$ decreases, with much less variability in $\delta(O_2/N_2)$ and CO_2 measurements, due to the absence of a terrestrial biosphere influence. The averaged CO₂ trend at Halley from 2014 to 2017 is 2.60 ± 0.20 ppm yr⁻¹, 1020 similar to the trends at Lutjewad and Mace Head in the same period $(2.62 \pm 0.08 \text{ ppm yr}^{-1} \text{ and } 2.53 \pm 0.05 \text{ ppm})$ yr⁻¹, respectively). On the other hand, δ(O₂/N₂) and APO trends at Halley are significantly smaller in size than those at Lutjewad and Mace Head. The $\delta(O_2/N_2)$ trend at Halley over the 2014-2017 period is -15 ± 3 per meg yr⁻ ¹ while at Lutjewad and Mace Head, the trends are -23.2 ± 0.9 per meg yr⁻¹and -20.3 ± 1.0 per meg yr⁻¹, respectively. For APO, the corresponding values are -1.4 ± 2.4 per meg yr⁻¹, -9.3 ± 0.6 per meg yr⁻¹, and -6.7 ± 0.6 per 0.9 per meg yr⁻¹.

1010

1015



Figure 6: Flask record from Lutjewad station, showing CO₂, δ(O₂/N₂), and APO measurements from 2000 to 2020. The black points are the individual flask measurements, while the black lines are the long-term trend and the coloured lines indicate the trend 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 yaxes ranges are scaled to represent the 5 per meg : 1 ppm ratio.



1035 Figure 7: As for Fig. 6 but for Mace Head station.



1040

Figure 8: As for Fig. 6 and 7, but for Halley station and from 2014 to 2017.

4.2 Seasonal cycles

1045

1050

The seasonal cycles of CO₂, δ (O₂/N₂), and APO for all three stations are presented in Fig. 9. The seasonal components are extracted from the total fits (detrended) and presented as 1-year cycles. In general, the CO₂ seasonal cycles at Lutjewad and Mace Head are similar in size and shape, although the average seasonal amplitude is higher at Lutjewad (16.8 ± 0.5 ppm) than Mace Head (14.8 ± 0.3 ppm). The CO₂ seasonal cycle at Halley station, on the other hand, has a much smaller amplitude of 3.0 ± 0.3 ppm, as is generally the case for the ocean-dominated Southern Hemisphere due to the absence of a terrestrial biosphere influence. Lutjewad and Mace Head show very similar, and significantly higher δ (O₂/N₂) seasonal amplitudes (131 ± 6 per meg and 130 ± 6 per meg, respectively) than that at Halley (76 ± 4 per meg), due to the influences 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 ± 3 per meg compared to 54 ± 4 and 61 ± 5 per meg, respectively). All numerical seasonality parameters of the three stations are given in Table 4 below.



1055 Figure 9: The detrended average seasonal cycles of CO₂ (left panel), δ(O₂/N₂) (middle panel), and APO (right panel) of stations Lutjewad (plotted in green), Mace Head (plotted in red), and Halley (plotted in blue). The uncertainty margins (2-sigma) in the fits have been indicated by lighter shades of the same colours.

Table 4: Trend and seasonality fit parameters of the measurement records from all three stations	s, as presented in
Fig. 6-8	

		Lutjewad (2002-2018)	Mace Head (2002-2018)	Halley (2014-2017)
	Seasonal amplitude (ppm)	16.8 ± 0.5	14.8 ± 0.3	3.0 ± 0.3
CO ₂	Average trend (ppm/year)	2.31 ± 0.07	2.22 ± 0.04	2.60 ± 0.20
	Day of min. value	$236 (Aug 24^{th}) \pm 13$	$238 (Aug 26^{th}) \pm 11$	$11 (Jan 11^{th}) \pm 12$
	Day of max. value	$62 (Mar 3^{rd}) \pm 26$	$105 (Apr 15^{th}) \pm 30$	$216 (Aug 4^{th}) \pm 14$
	Seasonal amplitude (per meg)	131 ± 6	131 ± 6	76 ± 4
$\delta(O_2$	Average trend (per meg/year)	-21.2 ± 0.8	-21.3 ± 0.9	-15 ± 3
/N ₂)	Day of min. value	85 (Mar 26^{th}) ± 23	42 (Feb 11 th) \pm 33	$239 (Aug \ 27^{th}) \pm 18$
	Day of max. value	$234 (Aug 22^{nd}) \pm 19$	$234 (Aug \ 22^{nd}) \pm 13$	59 (Feb 28 th) \pm 21
	Seasonal amplitude (per meg)	54 ± 4	61 ± 5	65 ± 3
APO	Average trend (per meg/year)	-9.3 ± 0.5	-9.7 ± 0.9	-1.4 ± 2.4
	Day of min. value	96 (Apr 6^{th}) ± 21	38 (Feb 7 th) + 30	$250 (\text{Sep } 7^{\text{th}}) \pm 12$
	Day of max. value	$284 (Oct 11^{th}) \pm 29$	$229 (Aug \ 17^{th}) \pm 29$	$66 (Mar 7^{th}) \pm 17$

1060 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 $\delta(O_2/N_2)$ and APO between Lutjewad and Mace Head (Fig. 6 and 7) 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 1065 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 Lutiewad 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 1070 procedure. It takes longer to transport the flasks from Mace Head to Groningen than from Lutjewad and thus contaminations of the samples through the valve caps might have occurred. For the samples from the Halley station, the transport time is even longer, but here, additional protective caps (glass or aluminium) with Viton orings 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 1075 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 $\delta(O_2/N_2)$ after 26 to 51 months. We found a small drift of 0.4 per meg in $\delta(O_2/N_2)$ after 48 months; and a drift of -0.3 ppm in CO2 after 24 months, on a set of 20 flasks. These numbers would only amount to biases of <u>0.008 per meg /month in $\delta(O_2/N_2)$ and 0.013 ppm/month in CO₂. Unfortunately, the protective caps were not</u> 1080 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 analysing 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.

1085 To summarise, 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 $\delta(O_2/N_2)$ and APO in 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 maybe 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 1090 quantity and type), which would influence $\delta(O_2/N_2)$ and to a lesser extent also APO. The average fossil fuel exchange ratio (ER) for the Netherlands, when accounting for all fossil fuel types, is 1.60 ± 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 CO2 release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) database by (Steinbach et al., 2011)), and the global average value for all fossil fuel emissions (of 1.38), as also mentioned by Sirignano et al. (2010) and van 1095 der Laan-Luijkx et al. (2010). However, it is unlikely that this is the main explanation of the difference between the two records. Firstly, because at Lutjewad, sampling was selective so as to avoid continental (and thus local fossil fuel) influences as much as possible, and second, because a difference in trends would need a gradual change in the ER. Data from Statistics Netherlands (CBS, 2021) show that the ER of the Netherlands has changed by no more than 0.02 over the period 2000-2020, too small to be of influence on the observed difference in the trends 1100 at Lutjewad and Mace Head. The next potential (though less likely) cause for differences between Mace Head and Lutjewad are changes in North Atlantic oxygen ventilation (Keeling and Manning, 2014) to which the Mace

Deleted: under much closer controlled

Deleted: counterac

Deleted: 1

Deleted: A more likely

Head observations are more sensitive. Such changes would influence $\delta(O_2/N_2)$ and APO, but not CO₂. This is consistent with the fact that the CO2 trends of Mace Head and Lutjewad agree, whereas there are differences in 1110 $\delta(O_2/N_2)$ 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 Prediction reported. Data obtained from the NOAA Climate Center (https://www.cpc.ncep.noaa.gov/data/teledoc/nao.shtml) 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 1115 negative phase in the years 2010-2011 towards gradually mostly positive values for the period 2013-2019. Other potential explanations could include a shift in atmospheric transport or also data artefact(s). As our operation continues, the coming years might shine light on what are likely or less likely causes.

When comparing the seasonal cycles of the three stations, we can see that while CO₂ and δ(O₂/N₂) 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 O₂ exchange. As mentioned in Sect. 2, APO values also contain a small influence from fossil fuels, however, by selecting for flasks based on the background conditions, we eliminate as much as possible this influence,
 especially for the Lutjewad record. As such, our APO values from these three stations represent mostly ocean

influences.

1130

As an illustration of the usefulness of the $\delta(O_2/N_2)$ measurement, we calculated the partitioning of CO₂ uptake by the terrestrial biosphere and the ocean from the observations at <u>Lutjewad</u> using the measurements of CO₂ and APO 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. 10.

-(Deleted: To conclude the comparison between Lutjewad and Mace Head
(Deleted: the two stations,
(Deleted: concentrations at Lutjewad and Mace Head from
4	Deleted: via





Deleted: Vector diagrams 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 CO₂ values at Lutjewad (left panel) and Mace Head (right panel). The black arrowed line represents the changes in the atmospheric APO and CO₂ values that would have occurred if all CO₂ 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/CO₂ molar ratio of 1.1 (that represents the removal of the biosphere signal in the definition of APO). The land biota uptake (green) is a horizontal line, as APO does not include a biosphere signal. The ocean O₂ outgassing effect is plotted in brown

\sum	is plotted in brown
Ć,	Deleted: concentrations
Q	Deleted: and Mace Head
	Deleted: , and clearly show the unexpected trend in APO for Mace Head, as discussed above
$\langle \rangle$	Deleted: 6
$\langle \rangle$	Deleted: 6
$\langle \rangle$	Deleted: 0
$\langle \rangle$	Deleted:
$\langle \rangle$	Deleted: Friedlingstein et al. (2020)
	Deleted: 38 ± 0.03
	Deleted: from Keeling and Manning (2014)
	Deleted: 1.7
	Deleted: 2.0
Ì	Deleted: 91

Deleted: The values calculated from Mace Head record are very similar to those from Lutjewad: at Mace Head, B is $1.7 \pm 1.1 \text{ PgC yr}^1$, O is $2.2 \pm 0.9 \text{ PgC yr}^1$ and A is $4.72 \pm 0.09 \text{ PgC yr}^1$. For both stations, the values for the B and O components agree well with those reported by Friedlingstein et al. (2020)

Figure 10: Vector diagrams 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 CO₂ values at Lutjewad, calculated from January to December of each calendar year. The black arrowed line represents the changes in the atmospheric APO and CO₂ values that would have occurred if all CO₂ 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/CO₂ molar ratio of 1.1 (that represents the removal of the biosphere signal in the definition of APO). The land biota uptake (green) is a horizontal line, as APO does not include a biosphere signal. The ocean O₂ outgassing effect is plotted in brown. The red line is a simple trend fitted through the period,
 The black points are the annual averages of the de-seasonalised measurements of APO and CO₂ mole fractions at

Lutjewad for the period 2002-2018. For calculating the partitioning of fossil fuel CO2, we use, from Keeling and Manning (2014), equations (2) to (10), and the ocean O₂ outgassing component (Z) of $0.44 \pm 0.45 \cdot 10^{14}$ mol yr⁻¹ (equivalent to an effect on the carbon sinks of 0.46 ± 0.48 PgC yr⁻¹). Furthermore, we use the total fossil fuel component for the years 2002-2018 of 8.9 ± 0.5 , PgC yr⁻¹ as derived from the Global Carbon Budget 2021, by 1150 (Friedlingstein et al., 2021), and the ER for globally averaged fossil fuel combustion of 1.43, from (Jones et al., 2021), From the Lutjewad record, the global land biotic sink (B) is 2.5 ± 1.1 PgC yr⁻¹, the oceanic sink (O) is 1.5. \pm 0.8 PgC yr⁻¹, and the CO₂ remaining in the atmosphere (A) amounts to 4.89 \pm 0.15 PgC yr⁻¹. However, these calculations did not take into account the riverine carbon fluxes, which are fluxes of carbon fixed on land and 155 transported into the oceans through the river systems (Jacobson et al., 2007). After correcting for the riverine fluxes of 0.6 PgC yr⁻¹(shifting from the land sink to the ocean sink) (Friedlingstein et al., 2021), the corresponding values are 1.9 PgC yr-1 for B and 2.1 PgC yr-1 for O. These values agree well with those reported by Friedlingstein et al. (2021): 1.6 ± 0.9 PgC yr⁻¹ for B (including emissions from land-use changes) and 2.5 ± 0.4 PgC yr⁻¹ for O. The value for component A at Lutjewad is slightly higher than the reported average value of 4.66 ± 0.02 PgC yr 160 ¹ for the 2002-2018 period, so therefore our sum of O and B is higher than that of Friedlingstein et al. (2021) by the same amount.

The challenges in making O2 measurements have presented themselves clearly in this work: the sensitivity of the mass spectrometer that require intensive calibration; the quality maintenance of the internal calibration scale to 1210 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 1215 is 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 improvement (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 1220 new primary standard cylinders from them, 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, to ensure the continuation of our calibration scale quality. More protective measures to the flasks, such as using additional caps or switching to another type of valve, will also be considered, to reduce the risks of potential leakages, 1225 permeations, and contaminations during storage and transportation.

5.2 Comparison with other long-term records

1235

In Table 5, we compare the seasonal amplitudes of our CO₂, δ(O₂/N₂), 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 δ(O₂/N₂) and APO measurements for Halley station show an excellent agreement with those at the Syowa station. On the other hand, our CO₂ measurements exhibit a much larger and noisier seasonal amplitude, which is <u>caused by small leaks during sampling (the details of which are given at the end of this section)</u>. Nonetheless, the general concurrence with these stations helps to consolidate the quality of our measurements.

Deleted: most likely

Table 5: Comparison of the seasonal amplitudes of CO₂, δ(O₂/N₂), and APO at various locations in the world

Station	Time period	Latitude	CO ₂ (ppm)	$\delta(O_2/N_2)$	APO	Reference
	<u>^</u>			. ,		
				(ner meg)	(ner meg)	
				(per meg)	(per meg)	
No. Ålammed	2001 2010	70°N	15.2 + 0.4	120 + 4	52 + 2	Tabidana at al
Ny Alesund,	2001-2010	/9°N	15.2 ± 0.4	129 ± 4	52 ± 5	Ishidoya et al.
						(2012b)
Spitsbergen						
Wevbourne, UK	2008-2015	53°N	15.2 ± 1.1	130 ± 8	51 ± 6	(Barningham, 2018)
						(
Lutiewad	2002-2018	53°N	16.8 ± 0.5	131 ± 6	54 ± 4	This naper
Eugenau,	2002 2010	00 11	10.0 - 0.0	151 = 0	5	This puper
d · M · d · · l · · l						
the Netherlands						
						1

Mace Head,	2002-2018	53°N	14.8 ± 0.3	130 ± 6	61 ± 5	This paper
Ireland						
Sendai, Japan	1999-2012	38°N	13.9 ± 2.5	128 ± 22	52 ± 10	Ishidoya et al.
						(2012a)
Syowa station,	2001-2010	69°S	1.1 ± 0.04	70 ± 4	64 ± 4	Ishidoya et al.
						(2012b)
Antarctica						
Halley station,	2014-2017	75°S	$3.0 \pm 0.3*$	76 ± 4	65 ± 3	This paper
Antarctica						
1991 GO 1 11 1				1 0 1 1		

*The CO2 seasonal amplitude at Halley is most likely incorrect, details are given at the end of this section.

1240

1245

Additionally, we compare our long-term measurement record with an extended record of Weybourne station (Fig. 11), the first part of which has been published by Pickers (2016) and Barningham (2018). The figure shows the continuous Weybourne records as hourly averages. In general, the two records agree well, except for the period of late 2018 to the end of 2019, when flask measurements (and the fit curves) of CO₂ 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 consolidates the quality of our measurements.

Deleted:

Deleted: Because of differences between continuous and flask sampling frequency, there are large differences between the two records, obviously in the amount of data points, but also in the height of the peaks and troughs during the diurnal cycle, and also the temporal length of the peaks and troughs. This is mostly because our flask samples at Lutjewad are collected during background conditions (and further selected for even more strict conditions during analyses) and are sometimes irregular in time, whereas the measurements at Weybourne are continuous (and therefore also represent local events). However, when considering background conditions at Weybourne (the lower limits of O_2 and upper limits of $\delta(O_2/N_2)$), there is good agreement with the Lutjewad record.





Figure 11: Measurements of CO₂, δ(O₂/N₂), 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 CO₂, δ (O₂/N₂), and APO measurements with those conducted by UEA (Fig. 12) (Barningham, 2018). APO measurements between our laboratory and UEA show good agreement, while CO2 measurements show unexpected discrepancies in March, April, and June until August of 2016. $\delta(O_2/N_2)$ 1270 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 CO₂ and $\delta(O_2/N_2)$ anomalies were most likely caused by a small inwards leak when the flask samples were collected at the station. Laboratory air with higher CO2 mole fractions and lower $\delta(O_2/N_2)$ ratios due to human breathing, probably leaked in. An additional indication pointing to this is that the CH4 and CO mole fractions from the same flasks agree very well with long-term flask measurements 1275 made at Halley by NOAA (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 use the long-term flask measurements Halley NOAA made at from the website (https://gml.noaa.gov/dv/iadv/graph.php?code=HBA&program=ccgg&type=ts.), since the UEA's measurement 1280 period is too short to make a reliable comparison. For CO2, we perform the trend and seasonality fitting procedure, the same as for our own measurements. The measurements between NOAA flasks and UEA agree very well, showing the reliability of UEA's measurements. Thus, the disagreement of CO₂ and δ (O₂/N₂) measurements between our laboratory and UEA firmly indicate the presence of leakages during March-August 2016, possibly

Deleted: the first half of

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 CO₂ measurements as compared to NOAA's, but since there is no available information on δ (O₂/N₂), we combine NOAA's CO₂ measurements with our own δ (O₂/N₂) measurements to calculate APO. Plotted in blue are the results using the NOAA's CO₂ measurements. A clear bias in APO is visible coinciding with the CO₂ anomalies: the CO₂ anomalies are around 2 ppm, which would lead to corresponding changes of 10 per meg in APO (since the APO is constructed from "clean" CO₂ and "contaminated" δ (O₂/N₂)). The short-term variations in δ (O₂/N₂) and APO are greater than 10 per meg, masking the suspected leaks. However, the significant difference between the average values for our APO measurements and the ones calculated using the NOAA's CO₂ and our δ (O₂/N₂) (indicated by the black and blue lines in the APO plot, respectively), suggesting that our flasks must have been contaminated with inside air in the early 2014 period.



1300

1305

290

295

Figure 12: Measurements of CO_2 , $\delta(O_2/N_2)$, 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 blue points are the in-situ continuous measurements at Halley, taken by NOAA. The blue lines and curves are the trends and the combined fits for the continuous measurements, with the lighter blue shaded area the 95% CI associated with the total fit<u>t</u>. The black and blue lines in the APO plot are the average values for our APO measurements, and the ones calculated using the NOAA CO₂ and our $\delta(O_2/N_2)$, respectively. The latter is significantly lower, corroborating our conclusion that our CO₂ measurements must have been contaminated with inside air (human breathing). The CO₂ scale is zoomed in to show the anomalies in 2016 more clearly.

Deleted: For APO, since there is no available information on O_2 , we combine the NOAA's CO₂ measurements with our own $\delta(O_2/N_2)$ measurements to calculate APO, then proceeded with the fitting of the trend and seasonality. Plotted in blue are the results using the NOAA's CO₂ measurements. Leakages would lead to higher CO₂ mole fraction and lower $\delta(O_2/N_2)$ ratio than the background sample, but APO would largely cancel them. We now combine the background CO₂ measurements from NOAA and the $\delta(O_2/N_2)$ values from our (supposedly contaminated) flask measurements, and indeed a clear bias in the APO can be seen coinciding with the CO₂ anomalies. This confirms the presence of some contamination in a number of our flasks, especially during March-June 2016 and in early 2014.



Deleted: . From the combination of all data we conclude that our CO₂ measurements must have been contaminated with inside air (human breathing).

6 Conclusion

Head, along with 3-year records from Halley. We also presented results of the calibration procedures of our 1330 instruments. Due to the sensitive nature of oxygen measurements, we conducted an extensive and intensive calibration procedures, which demonstrated a long-term stability for $\delta(O_2/N_2)$ of less than 3 per meg in 14 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) consolidate the stability, quality, and comparability of our calibration 1335 procedure, although there are some indications that our calibration scale might not be entirely stable over the past 20 years. However, the results from those various programmes are not consistent, and therefore inconclusive. The long-term records from Lutjewad and Mace Head provided useful information on the two-decadal trends and seasonality of CO₂, δ (O₂/N₂), 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 1340 2.31 ± 0.07 ppm yr⁻¹ for CO₂ and -21.2 ± 0.8 per meg yr⁻¹ for $\delta(O_2/N_2)$ at Lutjewad, and 2.22 ± 0.04 ppm yr⁻¹ for CO_2 and -21.3 \pm 0.9 per meg yr⁻¹ for $\delta(O_2/N_2)$ at Mace Head. The notable differences in the year-to-year progression of $\delta(O_2/N_2)$ and APO trends between Lutjewad and Mace Head might in part be caused by the sparse sampling frequency at Mace Head, but also may potentially be indications of influences from the changes in continental fossil fuel use, different degrees of sensitivity to the North Atlantic O2 ventilation, a shift in 345 atmospheric transport, or an artefact in the data. Using the measurements at Lutjewad for 2002-2018, the partitioning of atmospheric CO₂ sinks into the global terrestrial biosphere and the oceans are 1.9+1.1 PgC yr¹ and 2. L+ 0.8 PgC yr⁻¹, respectively. These values agree well with the numbers reported in the most recent Global Carbon Budget. The Halley record shows that the APO seasonal variations in the Southern Ocean are slightly larger than those in the Northern Hemisphere due to larger air-sea O2 exchange there, and illustrates clearly the influences of oceanic processes on the variations in APO and atmospheric O2. With better maintenance of our

We have presented 20-year flask measurement records for $\delta(O_2/N_2)$, CO₂ and APO from Lutjewad and Mace

1350

Data availability

of our future flask measurements will be improved.

The accompanying database comprises three csv files. The files contain the information on the CO₂, δ (O₂/N₂), and APO measurements (measured values and associated uncertainties) of the three stations, and are named after the corresponding station and the measured parameter (9 files in total).

internal scale, more regular sampling frequency, and better quality-control of the sampling process, the reliability

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

1360

1355

The additional data presented in this paper are available upon request.

Author contributions

Deleted: long-term standard deviation for $\delta(O_2/N_2)$ of 13.5 per meg based on our own internal cylinders

Deleted: but also may be indications of real influences from the various types of fossil fuel used, or from changes in ocean O2 ventilation

Deleted: 7 Deleted: 0

Deleted: , and when calculated using the Mace Head record, these are 1.7 ± 1.1 PgC yr⁻¹ and 2.2 ± 0.9 PgC yr⁻¹, respectively.

1375

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 the technical work and prepared the flask samples at Lutjewad station, and carried out the CO₂ and δ (O₂/N₂) measurements from flasks collected at all three stations. HAS calibrated the CO₂ data at CIO. AEJ, NB, and TB performed the measurements at Halley station, prepared the flask samples, produced the data for comparison. PAP and ACM provided the data from Weybourne station. All co-authors contributed to the writing of the manuscript.

1380 Competing interests

The authors declare that they have no conflict of interest.

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 1385 for help with the measurement database). Furthermore, we thank the collaborators at the stations in collecting and transporting the flask samples, specifically T.G. 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 the UEA staff and students (Michael Pateki, Philip Wilson, Grant Forster, And Dieu Tran and Leigh Fleming) and the Atmospheric Measurement and Observation Facility at the UK National Centre for 1390 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 received partial financial support over the years by the Dutch Research council (NWO), the European Union Integrated Project Carbo-Ocean (511176) and the Dutch national CATO-2 programme. Atmospheric O2 and CO2 measurements at Weybourne were supported by UK Natural Environment 1395 Research Council (NERC) grants NE/F005733/1, NE/I013342/1, QUEST010005 and NE/S004521/1. The Weybourne atmospheric O2 and CO2 measurements have been supported by the National Centre for Atmospheric Science (NCAS), funding agreement R8/H12/83/037 since December 2013 onwards. I.T.L. received funding from the Dutch Research Council (NWO) (Veni grant 016.Veni.171.095). A.E.J and N.B. were supported by the BAS core programme "Polar Science for Planet Earth". T.B. and P.A.P. were supported by UK NERC studentships. 1400 NE/L50158X/1 and NE/K500896/1, P.A.P. and A.C.M. 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.

References

1405

Aoki, N., Ishidoya, S., Tohjima, Y., Morimoto, S., Keeling, R. F., Cox, A., Takebayashi, S., and Murayama, S.:
Inter-comparison of O₂/N₂ Ratio Scales Among AIST, NIES, TU, and SIO Based on Round-Robin Using Gravimetric Standard Mixtures, Atmos. Meas. Tech. Discuss. [preprint], 2021, 1-26, 10.5194/amt-2020-481, 2021.

Deleted:

1410	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.
1415	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 Biogeochemical Cycles, 20, n/a-n/a, 10.1029/2005gb002534, 2006.
	Bender, M. L., Ellis, T., Tans, P. P., Francey, R., and Lowe, D.: Variability in the O ₂ /N ₂ ratio of southern hemisphere air, 1991–1994: Implications for the carbon cycle, Global Biogeochemical Cycles, 10, 9 - 21, 1996.
1420	Bender, M. L., Tans, P. P., Ellis, T., Orchardo, J., and Habfast, K.: A high precision isotope ratio mass spectrometry method for measuring the O ₂ /N ₂ ratio of air, Geochimica et Cosmochimica Acta,, 58, 4751 - 4758, 1994.
	Blaine, T. W., Keeling, R. F., and Paplawsky, W. J.: An improved inlet for precisely measuring the atmospheric Ar/N ₂ ratio, Atmos. Chem. Phys., 6, 1181-1184, 10.5194/acp-6-1181-2006, 2006.
1425	Bousquet, P., Gaudry, A., Ciais, P., Kazan, P., Monfray, P., Simmonds, P. G., Jennings, S. G., and O'Connor, T. C.: Atmospheric CO ₂ Concentration Variations Recorded at Mace Head, Ireland, From 1992 to 1994, Phys. Chem. Earth, 21, 477 - 481, 1996.
	access: 02/04/2021.
	The Netherlands in figures: https://opendata.cbs.nl/statline/#/CBS/en/, last access: 22/04/2021.
1430	Chen, H., Winderlich, J., Gerbig, C., Hoefer, A., Rella, C. W., Crosson, E. R., Van Pelt, A. D., Steinbach, J., Kolle, O., Beck, V., Daube, B. C., Gottlieb, E. W., Chow, V. Y., Santoni, G. W., and Wofsy, S. C.: High-accuracy continuous airborne measurements of greenhouse gases (CO ₂ and CH ₄) using the cavity ring-down spectroscopy (CRDS) technique, 3, 375 - 386, 2010.
1435	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., Myneni, 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., D. Qin, GK. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
1440	Conway, T. J., Tans, P. P., and Waterman, L. S.: Atmospheric 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

Barningham, T.: Detection and Attribution of Carbon Cycle Processes from Atmospheric O_2 and CO_2

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, Atmospheric Environment, 36, 2799 - 2807, 2002.

- 1445 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 Quéré, C., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S., Aragão, L. E. O. C., Arneth, 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.
- 1450 K., Joetzjer, 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 System Science Data, 12, 3269-3340, 10.5194/essd-12-3269-2020, 2020.

- 1460 Currie, K. I., Decharme, B., Djeutchouang, 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., Luijkx, 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., 2021, 1-191, 10.5194/essd-2021-386, 2021.
- 1470 Goto, D., Morimoto, S., Ishidoya, S., Aoki, S., and Nakazawa, T.: Terrestrial biospheric and oceanic CO2 uptakes estimated from long-term measurements of atmospheric CO2 mole fraction, δ13C, and δ(O2/N2) at Ny-Ålesund, Svalbard, Journal of Geophysical Research: Biogeosciences, 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:
 1475 Implications For the marine and atmospheric oxygen cycles, Global Biogeochemical Cycles, 15, 783-803, 10.1029/2000gb001302, 2001.

<sup>Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C. E., Hauck, J., Le Quéré, 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.,</sup>

Hall, B. D., Crotwell, A. M., Kitzis, D. R., Meffor, T., Miller, B. R., Schibig, M. F., and Tans, P. P.: Revision of the WMO/GAW CO₂ Calibration Scale, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2020-408, 2020.

Ishidoya, S., Aoki, S., Goto, D., Nakazawa, T., Taguchi, S., and Patra, P.: Time and space variations of the O₂/N₂ ratio in the troposphere over Japan and estimation of the global CO₂ budget for the period 2000–2010, Tellus B: Chemical and Physical Meteorology, 64, 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 CO₂ uptake estimated from atmospheric potential oxygen observed at Ny-Ålesund,
Svalbard, and Syowa, Antarctica, Tellus B: Chemical and Physical Meteorology, 64, 10.3402/tellusb.v64i0.18924, 2012b.

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., and Gloor, M.: A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, Global Biogeochemical Cycles, 21, https://doi.org/10.1029/2005GB002556, 2007.

1490 Jennings, S. G., McGovern, F. M., and Cooke, W. F.: Carbon Mass Concentration Measurements at Mace Head, on the West Coast of Ireland, Atmospheric Environment, 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.,
Clemitshaw, 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

1495 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, 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 Quéré, C.: Gridded fossil CO2 emissions and related O2 combustion consistent with national inventories 1959–2018, Scientific Data, 8, 2, 10.1038/s41597-020-00779-6, 2021.

1500 Keeling, R. F.: Development of an Interferometric Oxygen Analyzer for Precise Measurement of the Atmospheric O₂ 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 O₂ Content, in: Treatise on Geochemistry, 2nd ed., edited by: Holland, H. D., and Turekian, K. K., 10.1016/B978-0-08-095975-7.00420-4, 2014.

- 1505 4,
 - 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 O₂ concentration and their application in southern hemisphere air, Journal of Geophysical Research: Atmospheres, 103, 3381-3397, 10.1029/97jd02537, 1998.

1510

1525

1535

Keeling, R. F., Blaine, T., Paplawsky, B., Katz, L., Atwood, C., and Brockwell, T.: Measurement of changes in atmospheric Ar/N₂ ratio using a rapid-switching, single-capillary mass spectrometer system, Tellus B: Chemical and Physical Meteorology, 56, 322-338, 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 Biogeochemical Cycles, 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,

1520 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" O2 intercomparison programme: Latest results and next step, Second APO Workshop 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 underwater captured by sustained, long-term oceanographic time series observations, Global Biogeochemical Cycles, 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, 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, Journal of Atmospheric and Oceanic Technology, 21, 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 Luijkx, I. T.: Supplement data of the Two decades of flask observations of atmospheric δO₂/N₂, CO₂, and APO at stations Lutjewad (the Netherlands) and Mace Head (Ireland) plus 3 years from Halley station (Antarctica), ICOS-ERIC Carbon Portal, [to be added], 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: 02/04/2021. Pickers, P. A.: New applications of continuous atmospheric O₂ measurements: meridional transects across the
 Atlantic Ocean, and improved quantification of fossil fuel-derived CO₂, 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 O₂ and CO₂ reveal an unexpected O₂ signal over the tropical Atlantic Ocean, Global Biogeochemical Cycles, 31, 1289-1305, 10.1002/2017gb005631, 2017.

1545 Rousseeuw, P. J. and Verboven, S.: Robust estimation in very small samples, Computational Statistics & Data Analysis, 40, 741-758, 2002.

Severinghaus, J. P.: Studies of the terrestrial 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 CO₂ release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, Atmos. Chem. Phys., 11, 6855-6870, doi:10.5194/acp-11-6855-2011, 2011.

1555 Stendardo, I. and Gruber, N.: Oxygen trends over five decades in the North Atlantic, Journal of Geophysical Research: Oceans, 117, https://doi.org/10.1029/2012JC007909, 2012.

Stephens, B. B., Keeling, R. F., and Paplawsky, W. J.: Shipboard measurements of atmospheric oxygen using a vacuum-ultraviolet absorption technique, Tellus B: Chemical and Physical Meteorology, 55, 857-878, 10.3402/tellusb.v55i4.16386, 2003.

1560 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, Journal of Atmospheric and Oceanic Technology, 24, 82-94, 10.1175/jtech1959.1, 2007.

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 O₂ and CO₂ concentration, Global Biogeochemical
 Cycles, 12, 213-230, 10.1029/97gb03500, 1998.

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, 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,
 Journal of Geophysical Research: Atmospheres, 109, n/a-n/a, 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, Journal of Geophysical Research: Atmospheres, 94, 8549-8565, https://doi.org/10.1029/JD094iD06p08549, 1989.

1575 Tohjima, Y.: Method for measuring changes in the atmospheric O₂/N₂ ratio by a gas chromatograph equipped with a thermal conductivity detector, Journal of Geophysical Research: Atmospheres, 105, 14575-14584, 10.1029/2000jd900057, 2000.

Tohjima, Y., Mukai, H., Machida, T., Hoshina, Y., and Nakaoka, S. I.: Global carbon budgets estimated from atmospheric O2/N2 and CO2 observations in the western Pacific region over a 15-year period, Atmos. Chem.
Phys., 19, 9269-9285, 10.5194/acp-19-9269-2019, 2019.

Tohjima, Y., Mukai, H., Nojiri, Y., Yamagishi, H., and Machida, T.: Atmospheric O₂/N₂ 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.

1585 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, Journal of Geophysical Research, 110, 10.1029/2004jd005595, 2005.

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 Biogeochemical Cycles, 26, n/a-n/a, 10.1029/2011gb004110, 2012.

Cucumbers Intercomparison: http://cucumbers.uea.ac.uk/, last access: 24/02/2021.

1590

van der Laan, S., Neubert, R. E. M., and Meijer, H. A. J.: A single gas chromatograph for accurate atmospheric mixing ratio measurements of CO₂, CH₄, N₂O, SF₆ and CO, Atmos. Meas. Tech., 2, 549 - 559, 2009.

van der Laan, S., Karstens, U., Neubert, R. E. M., van der Laan-Luijkx, I. T., and Meijer, H. A. J.: Observation based estimates of fossil fuel-derived CO₂ emissions in the Netherlands using ¹⁴C, CO and ²²²Radon, 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.

1600 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.: CO₂, δO₂/N₂ and APO: observations from the Lutjewad, Mace Head and F3 platform flask sampling network, Atmospheric Chemistry and Physics, 10, 10691-10704, 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 CO₂, δ(O₂/N₂) and
 δ¹³CO₂ 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.

1610 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.

1615

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.

Page 36: [1] Deleted

I

v

Linh Nguyen

11/22/21 11:09:00 AM