Updated assessment of $\Delta^{14}CO_2$ measurement intercomparability using atmospheric records and standard materials

September 14, 2022

1 Introduction

This is currently going to be a running to do

- add summary table in methods summarizing all the institutes, their full names, types of analysis and other information. this could be part of the larger results table as well. ‘
- Address the flaring out of data in Heidelberg comparison around 1994 and 2006, this is due to the data already starting to drift in the direction of higher noise. This is the reason why I actually didn’t use 1995 and 2005 but actually extended the boundaries of the record by one year on each side.
- Try to explain the lower values in RRL NWT3 and NWT4 data around 2016-2017

2 Methods

The following institutions' radiocarbon measurements ($\Delta^{14}C$ and/or FM) were compared with the GNS Rafter Radiocarbon Laboratory (hereby shortened to "RRL") in turn, each elaborated upon in the following sections. Each intercomparison is tailored specifically to the type of data available between each institution.

2.1 RRL and Heidelberg University

2.1.1 Available Data

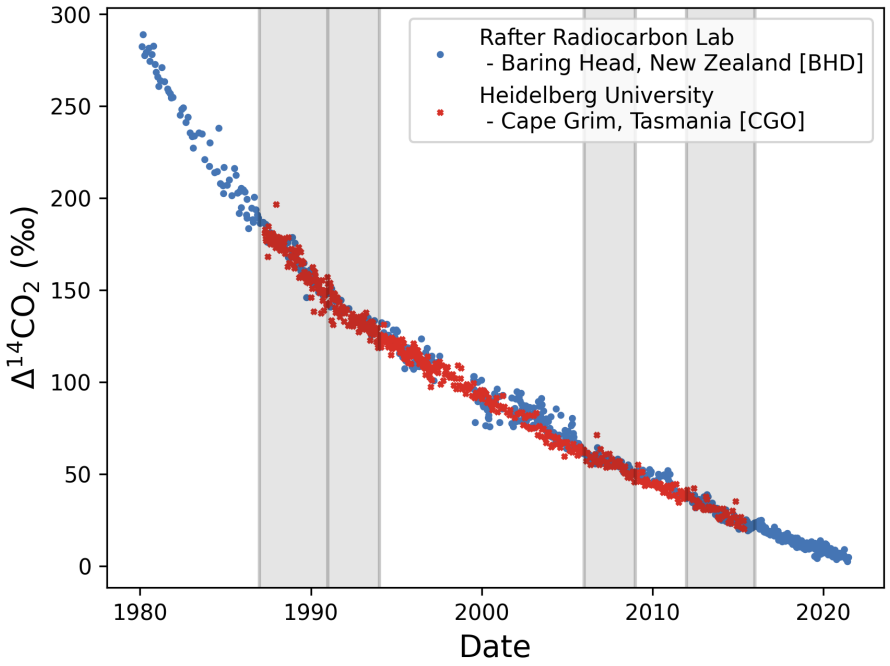

Figure 1: Visual summary of available data used for intercomparison between Heidelberg University and Rafter Radiocarbon Lab. The four time intervals of focus are highlighted with background grey: 1987 - 1991, 1991 - 1994, 2006 - 2009, 2012 - 2016

The RRL Wellington record is the earliest $\Delta^{14}CO_2$ dataset in the world, and the only one in the Southern Hemisphere to capture the "14C bomb-spike". Beginning in 1954 and still accumulating data, this record has seen changes in methods and sampling sites over the years, discussed in detail in [11]. Samples have been collected mainly via NaOH absorption; while Whole-air flask sampling was added as supplement since 1984. At RRL, sample CO_2 is extracted, graphitized [12] and $\Delta^{14}CO_2$ measured via accelerator mass spectrometry.[12, 14].

The Heidelberg University Institute of Environmental Physics in affiliation with the ICOS Central Radiocarbon Laboratory operates a network of time-series stations measuring $\Delta^{14}CO_2$ sampled using the NaOH absorption method [6, 7] and analyzed for $^{14}CO_2$ via decay counting [4]. One of these stations, Cape Grim, Tasmania (CGO; 40.68S, 144.68E, 94 m a.s.l [7]), is a reasonable candidate

through which to compare Heidelberg University to RRL $\Delta^{14}CO_2$ measurements through time.

CGO and Wellington observe a similar mixture of air from the Southern Ocean and Australia [13], and have been found to be comparable in the past![11]. Differences in seasonality and shifting air-mass origins with various contribution of fossil-CO2 are discussed in [11], but will be ignored in this work which focuses on extracting longer-term trends (see section X).A short initial time-series indicates no measurable difference between the sites from 2017-2019 (See Figure **??**).

The available data used to compare RRL and Heidelberg University can be seen at a glance in Figure1). The Wellington record (1955-present) and CGO record (1987-2016) overlap for 30 years. Certain intervals of data will be ignored. It has been previously shown that data in the Wellington record sampled via NaOH absorption is anomalously different from CGO in the period between 1990 and 1993[11, 7]. In this time period, only Whole-air flask measurements are included for the Wellington Record. In 1995, RRL switched methods of $\Delta^{14}C$ measurement from gas-counting to AMS using an ENTandem system. In the period between 1995 and 2005, excess noise exists in the record, after which online ^{13}C measurement allowed for appropriate fractionation correction [11, 14], and significant decrease in noise. For this reason, the period between 1995 and 2005 is ignored. RRL data in the period between 2009 and 2012 also is significantly offset from the otherwise consistency of the data; likely due to intermittent changes in NaOH absorption sampling techniques during this period. This period is therefore ignored for further intercomparison.

2.1.2 Intercomparison Method

To extract long-term systematic offsets between institutions, and remove seasonality, the CCGCRV curve fitting procedure ([10]; www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/) is implemented similar to [11]. We employ the "smooth" and "trend" functions of the CCGCRV algorithm. "Smoothed" data includes the results of the polynomial and harmonic fits of the data, and a long-term low-pass filter of 667 days. "Trended" data is similar; but retains the polynomial fit to the function and ignores harmonic components. Since the BHD and CGO data were not sampled on the same dates (i.e., they have an unequal number and value of x-components which impairs direct comparison of fitted data), the CCGCRV algorithm is programmed to output each smoothed/trended curve in 348 equal steps from 1987 to 2016 (12 samples/year), slightly underestimating the average sampling resolution of each dataset (CGO: 17 samples/year; BHD: 12.25 samples/year). By controlling the x-values of smoothed/trended data output from CCGCRV, the datasets can be compared. Error-estimates from the curve smoothing processes are obtained via a Monte-Carlo simulation, run to 10,000 iterations (for further details see Supplementary Information).

2.2 RRL and Scripps Institute of Oceanography / Lawrence Livermore National Laboratory

RRL and Scripps Institute of Oceanography / Lawrence Livermore National Laboratory (SIO/LLNL) are compared through measurement of the same standard materials, NWT3 and NWT4. NWT3 is a cylinder of air collected at Niwot Ridge, Colorado, USA in 2009. NWT4 is the same air, but spiked with 14C-free CO2. NWT3 and NWT4 are collected, maintained, and measured by INSTAAR for long-term repeatability assessments [5]. Through measurement collaboration, these standard were shared with LLNL and RRL. CO2 was produced from the sample air, combusted, extracted and measured similar to described in section 2.1.1 [12].

At LLNL, NWT3 and NWT4 air was extracted according to techniques used for stable isotopic measurements at Scripps, independently tested for precision and accuracy [2], and measured on an

HVEC FN Tandem accelerator at LLNL [1].The measurements presented from LLNL were made between March and April of 2009, following the collaboration noted in [5], while the measurements at RRL were made between 2013 and 2019.

2.3 RRL and Australian Nuclear Science and Technology Organisation (ANSTO)

Intercomparability between RRL and ANSTO is assessed via measurement of the same Kauri tree-ring sampled from Eastbourne in Wellington, New Zealand. At RRL, the sample was pretreated to extract cellulose via solvent washes and subsequent oxidation according to [3, 9], and subsequently measured by AMS.

2.4 Statistical Analyses

For intercomparisons of non-stationary time-series, such as ANSTO and RRL tree-ring records and Heidelberg University and RRL atmospheric records, intercomparability is assessed using paired (relative) t-tests. Comparison with SIO/LLNL uses an independent t-test. In all cases, the null hypothesis is that no difference between the institutions exists. The hypothesis is rejected if p-values are ¡0.01 (these are displayed in Table X). Statistical calculations were made using the Python scipy.stats library (https://docs.scipy.org/doc/scipy/reference/stats.html).

3 Results/Discussion

3.1 RRL vs. Heidelberg University

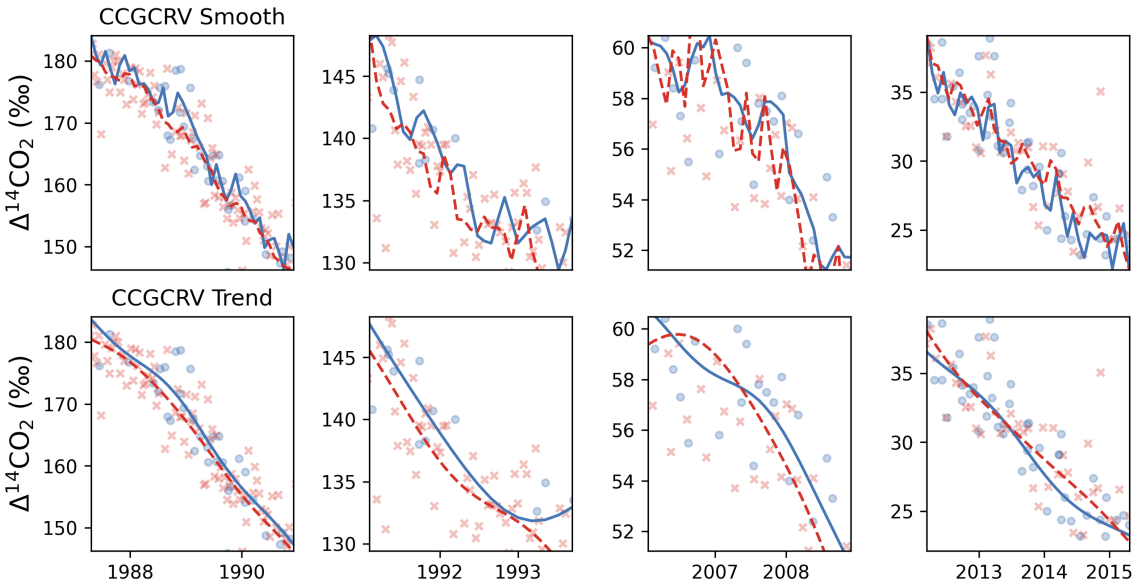

Figure 2: Means of Monte Carlo simulation using CCGCRV "smooth" function (top panel), and "trend" function (bottom panel) overlaid upon initial CGO and BHD data. Panel a-d represents the various time-intervals used in this intercomparison: 1987 - 1991, 1991 - 1994, 2006 - 2009, 2012 - 2016

This analysis focuses on investigating long-term systematic biases between the RRL Wellington record and the Heidelberg University CGO record, and ignores seasonality further explored in previous works [11]. The CCGCRV curve-fitting algorithm allows the data to be removed of noise and seasonality, using the "getSmooth" and "getTrend" functions. These functions fit the data with polynomial+harmonic, and polynomial terms only, respectively. The results of the CCGCRV curve fitting algorithm are shown in Figure 4 for time intervals 1987 - 1991, 1991 - 1994, 2006 - 2009, 2012 - 2016. Semi-transparent data is overlaid with the fits, which are the mean of output from the Monte Carlo simulation. The difference of each mean (14CO2 BHD - CGO) is recorded and the average for each interval deposited into Figure3.3. Uncertainties are the 1-σ error around the mean.

3.1.1 CCGCRV Smooth vs. Trend

$\Delta\Delta^{14}C$ values using the CCGCRV "smooth" and "trend" algorithm are the same within error for every case. Complete removal of seasonality results in a more clear visualization of an offset between Wellington and CGO records, further explored in section 3.1.2. In the first two time-intervals, paired t-tests show Wellington and CGO statistically different in all cases. In later years, when the observed offset decreases, the variability from seasonality in the "smoothed" data results in a "not different" result while the "trended" data is still different".

3.1.2 Wellington vs CGO

In the period between 1987 and 1994, RRL measurements are greater than 1.7‰higher than Heidelberg University (see Figure 4), and statistically different according to paired t-tests (p-value ¡0.01). There is a step-decrease in the offset in the following intervals: DeltaDelta14C is -0.54 ± 0.21 to 0.58 ± 0.26, within error of the intercomparability goals reestablished by the WMO and GGMT in 2020. Despite being within-error of the intercomparability goal, the data are still statistically different according to paired t-tests. The step change in offset between 1987-1991 and 2006-2009/2012-2016 happens in parallel to a significant shift in RRL workflow, namely the inclusion of a 12C cup in the EN Tandem facility at GNS Science, allowing online 13C correction. This dramatically decreased measurement noise [11]. As changes in RRL resulted in decreased measurement noise and better intercomparability, while systems at Heidelberg University remained constant; this points to the possibility that the Wellington record in the early period is offset too high leading to the systematic bias during that time (ugly sentence, fix later).

3.2 SIO / LLNL

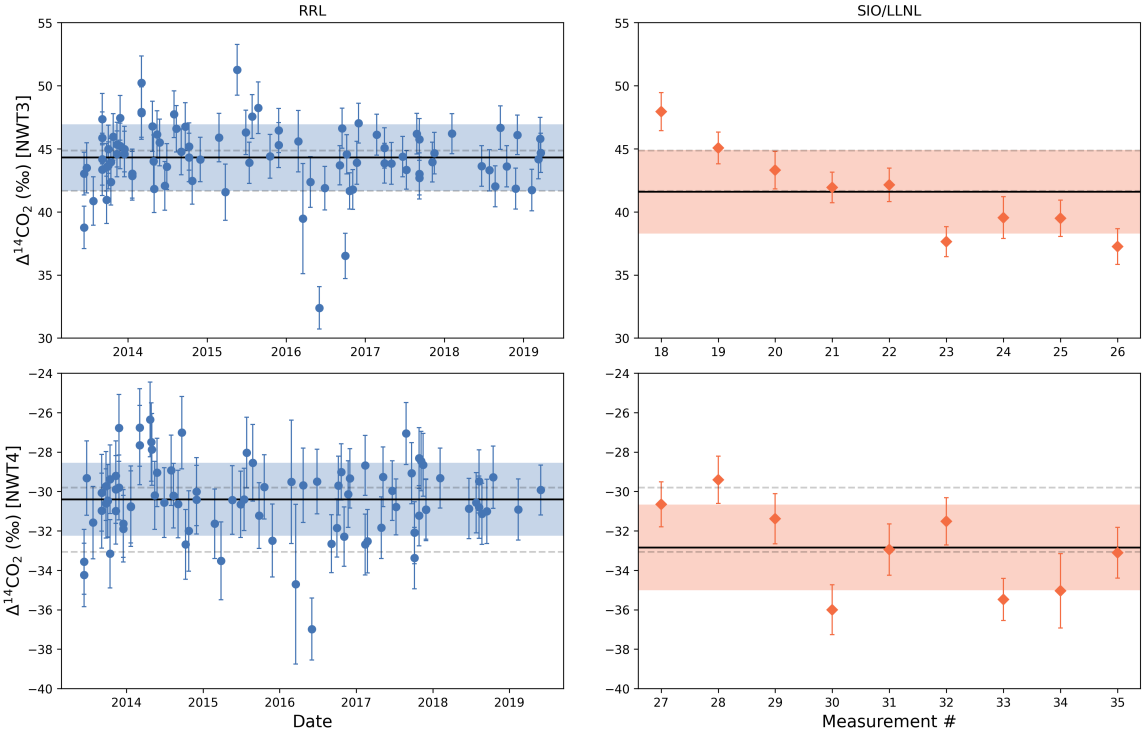

Figure 3: Top panel shows $\Delta^{14}CO_2$ measurements of NWT3 standard from RRL and SIO/LLNL, from left to right. Results for NWT4 $\Delta^{14}CO_2$ measurements are shown similarly on the bottom panel. The faint dotted lines indicate the 1-sigma error range for either standard measured at INSTAAR from 2009-2012

The Niwot Ridge standard materials include one cylinder of ambient air collected from Niwot Ridge, Colorado, USA in 2009 (NWT3) and a second cylinder spiked with some 14C-free CO2.

Figure 3 shows $\Delta^{14}CO_2$ measurements of each material made at RRL and LLNL, in the left and right columns, respectively. A summary of means and standard deviations of each institute's available data for NWT3 and NWT4 is provided in Figure 3.2. While RRL measurements are slightly higher than both INSTAAR [5] and SIO/LLNL, all three datasets are within $1-\sigma$ error of each other. Although they are within error, t-tests show that RRL and SIO/LLNL are statistically offset.

"	NWT3	NWT4
RRL	44.31±2.63	-30.40±1.85
SIO/LLNL	41.59±3.31	-32.84±2.16
INSTAAR [5]	43.25±1.6	-31.43±1.63

3.3 ANSTO and University of Magallanes

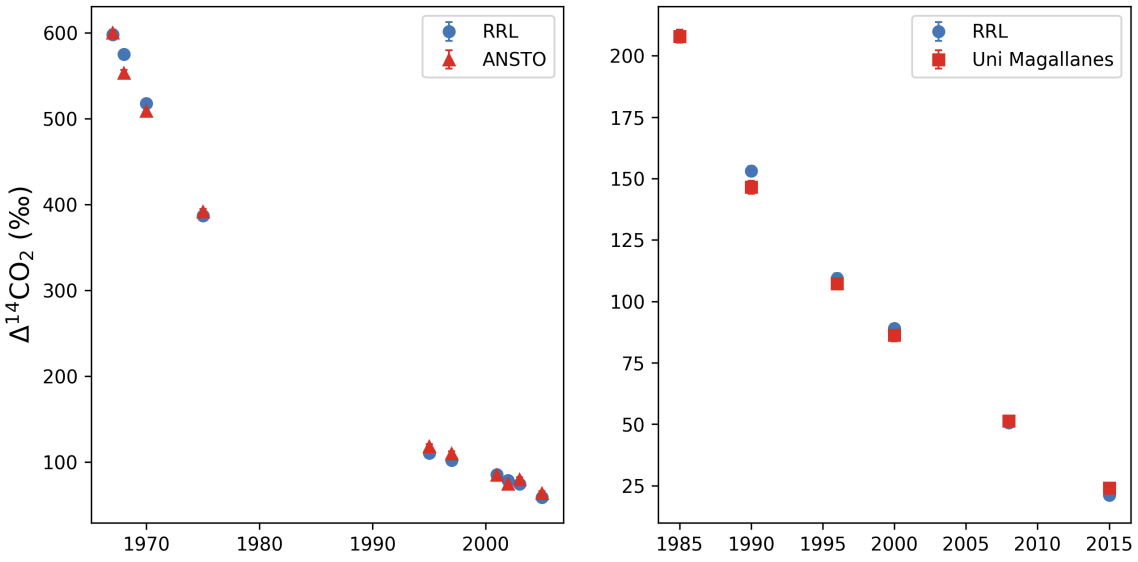

Figure 4: ADD CAPTION LATER

Institution	$\Delta\Delta^{14}C$	p-value	statistical result
(Heid. Uni) 1987-1991 (smooth/trend)	1.77±0.32/1.75 ± 0.10	2e-6 / 4.9e-21	Different/Different
(Heid. Uni) 1991-1994 (smooth/trend)	1.76±0.42/1.89 ± 0.20	3e-4 /1.3e-10	Different/Different
(Heid. Uni) 2006-2009 (smooth/trend)	0.58±0.26/0.55 ± 0.14	0.03 / 6.0e-4	Not Different/Different
(Heid. Uni) 2012-2016 (smooth/trend)	-0.54±0.21/ − 0.51 ± 0.10	0.01 / 2.15e-5	Not Different/Different
SIO/LLNL [NWT3]	2.72±1.14	0.005	Different
SIO/LLNL [NWT4]	2.44±0.75	4e-4	Different
ANSTO	0.46±2.76	0.88	Not Different

4 Conclusion

- The importance of this is: the dataset with the longest record reveals the potential variability in two institutions over time; wrt to instruments, users, and associated combustion/graphitization blanks that can all lead to small but persistent offsets that may propagate through the science. This Heidelberg intercomparison shows that in order to get meaningful understandings of the true intercomparability of institutions over time, it must be constantly repeated.
- if the data with the longest available record shows not only changes, but gradients of changes in time with instrument changes, it seeds doubt in the ability to use a short term intercomparison (such as SIO/LLNL) or ANSTO, for long-term offset corrections.
- Where do the offsets come from? Since normalizations to ox-1 should be the same, perhaps they come from combustion and graphitization? Perhaps they come from blank correction strategies? A good way to deal with this/ try to explore this is to have "nested" intercomparisons for example 1) we are all sent graphitze/zinc tubes and press them in house, therefore excluding combustion/graphitization blanks 2) send CO2 in breakseals (what we normally do), and 3) perhaps share blank-correction codes/excel sheets, python codes, software (Fudger, CALAMS), more widely so that users can test differences/intercomparability and edit if they see fit.
- this is something we will address in upcoming iterations of intercomparison, similar to [8].
- these data are hopefully going to be used to offset correct some more data for a new work...can they be trusted?
- how did the NWT3 and NWT4 data compare to the results from Miller 2013?
- Don't forget to mention the Instaar reults/comparison highlighted in Turnbull 2015b.
- For SIO/LLNL, the data are within error, but still statistically different. What do we do if, they are the same within error but the difference is still larger than the GGMT threshold? Need to reduce error, which is quite difficult

References

[1] Heather D Graven, Thomas P Guilderson, and Ralph F Keeling. Methods for high-precision 14c ams measurement of atmospheric co2 at llnl. *Radiocarbon*, 49(2):349–356, 2007.

[2] P.R. Guenther, A.F Bollenbacher, C.D. Keeling, E.F. Stewart, and M. Wahlen. Calibration methodology for the scripps 13c/12c and 18o/16o stable isotope program. 2001.

[3] Q Hua, M Barbetti, GE Jacobsen, U Zoppi, and EM Lawson. Bomb radiocarbon in annual tree rings from thailand and australia. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 172(1-4):359–365, 2000.

[4] Bernd Kromer and Karl Otto Münnich. Co2 gas proportional counting in radiocarbon dating — review and perspective. In R. E. Taylor, Austin Long, and Renee S. Kra, editors, *Radiocarbon After Four Decades*, pages 184–197, New York, NY, 1992. Springer New York.

[5] Scott J Lehman, John B Miller, Chad Wolak, John Southon, Pieter P Tans, Stephen A Montzka, Colm Sweeney, Arlyn Andrews, Brian LaFranchi, Thomas P Guilderson, et al. Allocation of terrestrial carbon sources using 14co2: Methods, measurement, and modeling. *Radiocarbon*, 55(3):1484–1495, 2013.

[6] Ingeborg Levin, KO Münnich, and Wolfgang Weiss. The effect of anthropogenic co2 and 14c sources on the distribution of 14c in the atmosphere. *Radiocarbon*, 22(2):379–391, 1980.

[7] Ingeborg Levin, Tobias Naegler, Bernd Kromer, Moritz Diehl, Roger Francey, Angel Gomez-Pelaez, Paul Steele, Dietmar Wagenbach, Rolf Weller, and Douglas Worthy. Observations and modelling of the global distribution and long-term trend of atmospheric 14co2. *Tellus B: Chemical and Physical Meteorology*, 62(1):26–46, 2010.

[8] John Miller, Scott Lehman, Chad Wolak, Jocelyn Turnbull, Gregory Dunn, Heather Graven, Ralph Keeling, Harro A J Meijer, Anita Th Aerts-Bijma, Sanne W L Palstra, and et al. Initial results of an intercomparison of ams-based atmospheric 14co2 measurements. *Radiocarbon*, 55(3):1475–1483, 2013.

[9] Margaret Norris. Reconstruction of historic fossil co emissions using radiocarbon measurements from tree rings. 2015.

[10] Kirk W. Thoning, Pieter P. Tans, and Walter D. Komhyr. Atmospheric carbon dioxide at mauna loa observatory: 2. analysis of the noaa gmcc data, 1974–1985. *Journal of Geophysical Research: Atmospheres*, 94(D6):8549–8565, 1989.

[11] J. C. Turnbull, S. E. Mikaloff Fletcher, I. Ansell, G. W. Brailsford, R. C. Moss, M. W. Norris, and K. Steinkamp. Sixty years of radiocarbon dioxide measurements at wellington, new zealand: 1954–2014. *Atmospheric Chemistry and Physics*, 17(23):14771–14784, 2017.

[12] Jocelyn C Turnbull, Albert Zondervan, Johannes Kaiser, Margaret Norris, Jenny Dahl, Troy Baisden, and Scott Lehman. High-precision atmospheric 14co2 measurement at the rafter radiocarbon laboratory. *Radiocarbon*, 57(3):377–388, 2015.

[13] T. Ziehn, A. Nickless, P. J. Rayner, R. M. Law, G. Roff, and P. Fraser. Greenhouse gas network design using backward lagrangian particle dispersion modelling minus; part 1: Methodology and australian test case. *Atmospheric Chemistry and Physics*, 14(17):9363–9378, 2014.

[14] A. Zondervan, T.M. Hauser, J. Kaiser, R.L. Kitchen, J.C. Turnbull, and J.G. West. Xcams: The compact 14c accelerator mass spectrometer extended for 10be and 26al at gns science, new zealand. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 361:25–33, 2015. The Thirteenth Accelerator Mass Spectrometry Conference.