Supplement of Atmos. Meas. Tech., 17, 4803–4823, 2024
https://doi.org/10.5194/amt-17-4803-2024-supplement

Atmospheric Measurement Techniques
Open Access
EGU

Supplement of

Atmospheric H_2 observations from the NOAA Cooperative Global Air Sampling Network

Gabrielle Pétron et al.

Correspondence to: Gabrielle Pétron (gabrielle.petron@noaa.gov)

S1. Limitations of NOAA GML 1988-2009 H_2 measurements on RGAs

Novelli et al. (1999) describe the NOAA H_2 flask air measurement procedure for 1988-1997. A few aspects of the program for the period 1988-2009 are summarized here to explain limitations in the older NOAA H_2 dataset and the decision to not convert older measurements to the current WMO recommended calibration scale. These limitations can broadly be categorized as 1) issues related to the non-linear response of the analyzers used for flask analysis, 2) instability in the underlying internal scale maintained by GML, and 3) lack of adequate electronic records to provide full transparency. These all impact the quality and internal consistency of the early data and the ability to retroactively convert the early data to the current WMO recommended H_2 in air calibration scale.

Insufficient instrument response characterization

Prior to 2009, NOAA GML used gas chromatography followed by hot mercuric oxide reduction (GC-HgO) and the UV absorption detection of the resulting elemental mercury for both standard air and flask air analyses of H_2. GML used commercial Reduction Gas Analyzer GC modules with HgO bed reduction gas detector from Trace Analytical Inc. (Menlo Park, California) and Peak Laboratories, LLC (Menlo Park, California). The NOAA RGA analyzers measured both H_2 and CO in the same chromatogram. Table S1 (further below) gives a list of the RGA instruments and working standards in service prior to the adoption of the GC-HePDD measurement technique.

The first instrument used, R2 (RGA3 GC module with RGD2 detector), was found to have a linear response for CO and H_2 over the range of mole fractions in the background atmosphere (Novelli et al., 1991, 1992). However, Novelli et al. (1992) cautioned that the instrument absolute response and linearity were HgO bed dependent and could change over time.

After 1990, all new HgO bed detectors had non-linear responses for both CO and H_2 (Novelli et al., 1998; Novelli et al., 2003). CSIRO and MPI-BGC H_2 measurement teams have reported similar results (Francey et al., 2003; Jordan and Steinberg, 2011).

In 1991, GML started using a suite of standards covering a range of CO mole fractions to create calibration curves during dedicated instrument response calibration episodes approximately bi-weekly (Novelli et al., 1998). This approach was not adopted for H_2, likely due to a lack of standards with stable H_2. Instead, for H_2 measurements, GML used a 1-point calibration strategy where the CO reference air tank, which brackets each sample aliquot, was value assigned for H_2 and used as the single H_2 working standard for calibrating flask air sample measurements. This strategy ignored the non-linear response of the detectors.

The non-linearity of the RGA3 response was assumed to be negligible over the narrow range of H_2 observed in background air samples from remote network sites. However, the impact of the non-linear response also depended on the H_2 working standards being themselves close to ambient H_2 mole fractions. In actuality, recorded H_2 assignments for the older working standards used for flask analysis ranged from 470 ppb to 644 ppb. This would give rise to persistent non-linearity induced biases on time scales of 6-18 months (the typical lifetime of the working standards) in the H_2 measurement records. GML did not characterize the non-linearity of the H_2 response of the RGAs so cannot retroactively correct for this

effect. The biases are expected to be significant for some time periods leading the authors to caution against using the NOAA early H_2 data records.

Instability in the NOAA H_2 X1996 calibration scale

NOAA H_2 mole fraction measurements from 1988 - 2009 are traceable to an internal calibration scale (NOAA H_2 X1996) maintained by GML. This scale was defined by five gravimetric standards made in 1995/1996 (CC73198, CC86013,CA01310, CC86208, CC86259) and covering the range 485 - 600 ppb H_2.

The X1996 scale was propagated to the five working standards (tanks ID with * in Table S1) used between 1988 and 1995 for flask air sample analyses by measurement against the gravimetric standards in 1996 (Novelli et al., 1999). However, these post deployment calibrations could not assess the stability of the working standards during usage prior to 1996 so any drift occurring in the working standards prior to 1996 would be unaccounted for leading to potential biases in the earliest records.

After 1996, the NOAA H_2 X1996 scale was maintained by bootstrapping secondary standards forward in time. In this method, each secondary standard was used to directly calibrate its successor. This method assumed no drift was occurring in either the initial secondary standard, nor in any subsequent secondary standard. While care was taken to use cylinders for secondary standards that did not display initial high drift of H_2, we now know that H_2 stability in air standards contained in aluminum cylinders is rare and growth of H_2 over time is much more likely. The bootstrap method is likely to have introduced long-term instability in the scale.

Incomplete record keeping early on

There is no electronic record of any calibration and no recorded assigned value for R7 working standard AAL-17259 used from February to August 1997. All R5 and R6 working standards have assignments on X1996 recorded back in June 2014, covering a wide range: 470-650 ppb. Only the later R5 standards (CC105928, CC71649) and R6 standards (CA06591, CC305198) have assignments with a linear drift correction. The other standards were assumed stable.

In addition to the other known limitations in the early implementation of the H_2 measurements, the lack of record keeping during the early years plays a role in the decision to not retroactively convert the early data to the current WMO recommended calibration scale. Documentation of decisions on standard value assignments, electronic records of raw data files for the instrument responses, and details of calibration hierarchy from the early records are often missing or lack sufficient detail. Unfortunately, this makes it impossible to recover the data, even within the larger uncertainties associated with the measurement issues discussed.

Examples of observed biases in the older NOAA H_2 measurements

Close in time analysis of CC119811on P2 in 2007 and 2008 against one of three SX standards (SX3540, SX-3523 or SX-3554) show a > 20 ppb spread in the derived H_2 (Figure S12), suggesting a strong non-linear response for P2. The response of the P2 instrument was never fully characterized. However, Novelli et al. (2009) (Table 4 herein) show results for eight tanks analyzed on P2 using one point or two

point calibration compared to their results on H9. The one point calibration results show the larger biases, especially for tanks with H_2 furthest from the H_2 in the reference/standard (525 ppb): underestimation for tanks with H_2 below 525 ppb reaching close to -20ppb at 420 ppb and overestimation for tanks with H_2 above 525 ppb reaching +12 ppb at 593 ppb.

The responses of the R5 and R6 instruments were never fully characterized. However NOAA started the regular analysis of target air tanks on the MAGICC-1 and MAGICC-2 systems in 2004. Results for target air tanks CC71583 (D) and CC1824 (H) are plotted in Figure S13 using different symbols and colors for different instruments and working standards. GC-HePDD measurements after 2008 show H_2 growing in both tanks. The earlier results on R5 and R6 are scattered and suggest inconsistent assignments between the working standards, also likely including incorrect drift estimates. It is not robust to extrapolate a tank H_2 assignment based on available measurements on H9 a few or several years back in time as the stability or growth of H_2 in high pressure aluminum cylinders can change over time.

S2. Same air comparison with CSIRO for NOAA historical H_2 data

CSIRO started measuring H_2 by gas chromatography with mercuric oxide detector (GC-HgO) using a Trace Analytical RGA instrument in 1991. Data are reported on the MPI X2009 H_2 calibration scale. CSIRO's implementation of the scale in recent years is defined by a suite of 5 H_2-in-air standards contained in electropolished 34L stainless steel cylinders (Essex Industries, St. Louis, MO) that were calibrated at MPI-BGC in 2016.

Implementation of the scale before 2016 is based on in-house calibration procedures involving repeat determinations of the non-linearity of instrument response and multi-decadal measurement histories of a large number of air standards.

From 1993, CSIRO started using "dilution experiments" of above ambient CH_4, H_2 and CO mole fraction air blended with varying proportions of purified zero air to periodically characterize the non-linearity of their GC-HgO instrument for CO and H_2. Dilution ratios were determined by measurements of CH_4 tied to a gravimetrically defined CH_4 calibration scale. They found the instrument response was "significantly nonlinear" and of similar shape for both H_2 and CO (of the form $y=ax^2+bx+cx^d$, where x = peak height and a,b,c,d are estimated parameters from the response function fit). A single response function was used for H_2 in the early 1991-1995 period due to insufficient well-behaved standards with mole fractions outside of the background atmospheric range to adequately monitor variations in the instrument response function (Francey et al., 2003).

Long term stability of CSIRO's H_2 records is constrained by observed relative stability of a large number of air standards stored in various stainless steel and aluminum cylinder types. A key constraint is relative stability to better than ± 0.2 ppb/yr among 59 individual standards in 34L or 35L stainless steel Essex cylinders as measured over time intervals of 7+ years. Of these, 29 were first analyzed between 1992 and 1999.

The early intercomparison of measurements by NOAA GML and CSIRO same air from the Kennaook/Cape Grim Observatory (1992-1998) showed significant (>2%) and trending biases (Masarie et al., 2001). The nonlinear response of the H_2 analytical system detector, the instability of H_2 standards stored in aluminum cylinders (commonly used for CO_2 and CH_4 standards) and the different calibration scales were presented as likely explanations for the observed time-dependent biases between the two labs.

S3. WMO/MPI-BGC X2009 H_2 calibration scale

To support advances in the understanding of the H_2 global budget, high quality and comparable observations are a non-negotiable requirement and should be anchored by a common stable calibration scale (WMO, 2007). The Max Planck Institute for Biogeochemistry (MPI-BGC) in Jena secured funding to support their laboratory work to investigate the stability of the H_2 mole fraction for reference air in various types of high pressure cylinders and to develop an accurate H_2 calibration scale. Jordan and Steinberg (2011) analyzed 100 air standards multiple times over a one to six year period on their GC-HgO instrument calibrated using multiple H_2 in real air standard gases to fully describe the detector nonlinear response. They concluded that the H_2 mole fraction for reference air in steel and stainless steel cylinders did not drift significantly (< 1.5 ppb/yr). For aluminum cylinders however, they found a wide range of H_2 mole fraction drift rates (< 1.5 ppb/yr to > 20 ppb/yr) and drift behaviors (short term, ie. drift over a few months, to continued growth in H_2). The MPI X2009 scale became the official WMO scale for H_2 in 2011 (Jordan and Steinberg, 2011). It is defined by thirteen standards (of which 12 are in stainless steel cylinders) with H_2 dry air mole fractions ranging from 139 ppb to 1226 ppb.

Once a CCL was established for H_2, experts from the WMO Global Atmospheric Watch recommended measurement laboratories adopt the WMO/MPI 2009 scale and develop procedures to track drifts in their standards and to appropriately characterize their instrument responses (WMO/GAW, 2014).

In 2007-2008, GML prepared 6 H_2 gravimetric standards ranging from 230 to 790 ppb in electropolished stainless steel cylinders (Essex Cryogenics, with tank IDs SX-#). Early results in WMO GAW measurement laboratories suggested H_2 was likely more stable in these cylinders than in aluminum cylinders. However, the new gravimetric mixtures differed by about +20 ppb compared to two H_2 secondary standards in aluminum cylinders GML used for the calibration of tertiary standards on the X1996 scale (Novelli, personal communication). In following years, GML continued using the 1996 gravimetric primary standards to define its internal H_2 calibration scale and also regularly measured the H_2 secondary standards against the stainless steel standards.

S4. MAGICC-3 reference air CA04145

To evaluate the stability of the reference air H_2 and the validity of the H_2 instrument response curve fit coefficients between MAGICC-3 instrument response calibration dates, we derive an H_2 assignment for the reference air cylinder for each instrument response calibration date (ratio of peak heights =1). For each MAGICC-3 reference air cylinder, we calculate its mean H_2 for the time period for which it was in use. The mean H_2 values for the 6 reference air cylinders used between July 2019 and January 2023 range

between 542 and 583 ppb.

In Figure S3 we plot the deviation of each reference air cylinder assignment from its mean value as a function of the MAGICC-3 instrument response calibration date. The very first reference CA04145 air cylinder had the largest growth in its H_2 mole fraction: + 7.5 ppb in 5 months (~ 18 ppb/yr). The incremental increase between calibration dates is larger when the calibration becomes less frequent in late 2019. We apply a correction of 18 *(Δt) to flask analysis results on MAGICC-3 between 11/6/2019 and 1/16/2020 with Δt being the difference between the flask analysis decimal date and the preceding response calibration decimal date (corresponding to calendar dates 11/6/2019, 12/4/2019 or 1/7/2020). For the period 3/26 to 8/1 2020 with the second reference air cylinder, MAGICC-3/H8 was more noisy and the increments in the reference air H_2 between response calibration dates jumped from -1 ppb to 1 ppb twice.

S5. References

Francey, R. J., Steele, L. P., Spencer, D. A., Langenfelds, R. L., Law, R. M.; Krummel, P. B., Fraser, P. J., Etheridge, D. M., Derek, N., Coram, S. A., Cooper, L. N., Allison, C. E., Porter, L., Baly, S.: The CSIRO (Australia) measurement of greenhouse gases in the global atmosphere. In: Tindale, N. W.; Derek, N.; Fraser, P. J. editors. Baseline Atmospheric Program Australia. Melbourne: Bureau of Meteorology and CSIRO Atmospheric Research; pp. 42-53. http://hdl.handle.net/102.100.100/191835?index=1, 2003.

Jordan, A., Damak, F.: ICOS Central Analytical Laboratories Flask and Calibration Laboratory, Quality Control Report 2022, 85p, available at: https://www.icos-cal.eu/fcl/qc-report, (Last Accessed December 13, 2023), 2022.

Masarie, K. A., Langenfelds, R. L., Allison, C. E., Conway, T. J., Dlugokencky, E. J., Francey, R. J., Novelli, P. C., Steele, L. P., Tans, P. P, Vaughn, B., White, J. W. C.: NOAA/CSIRO Flask Air Intercomparison Experiment: A strategy for directly assessing consistency among atmospheric measurements made by independent laboratories, J. Geophys. Res., 106(D17), 20445–20464, doi:10.1029/2000JD000023, 2001.

Novelli, P. C., Elkins, J. W., Steele, L. P.: The Development and Evaluation of a Gravimetric Reference Scale For Measurements of Atmospheric Carbon Monoxide, J. Geophys. Res., 96, 13,109-13,121, doi: 10.1029/91JD01108, 1991.

Novelli, P. C., Steele, L. P., and Tans, P. P.: Mixing ratios of carbon monoxide in the troposphere, J. Geophys. Res., 97, 20,731-20,750, doi: 10.1029/92JD02010, 1992.

Novelli, P. C., Lang, P. M., Masarie, K. A., Hurst, D. F., Myers, R., Elkins, J. W.: Molecular hydrogen in the troposphere: Global distribution and budget, J. Geophys. Res., 104, 30427–30444, doi: 10.1029/1999JD900788, 1999.

Novelli, P. C., Crotwell, A. M., and Hall, B. D., Application of Gas Chromatography with a Pulsed Discharged Helium Ionization Detector for Measurements of Molecular Hydrogen, Env. Sci. Technol. (43), 2431-2436, doi: 10.1021/es803180g, 2009.

WMO/GAW, 17th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2013) (Beijing, China, 10 - 13 June 2013) Edited by Pieter Tans and Christoph Zellweger, report 213. Accessible at https://www.uncclearn.org/wp-content/uploads/library/gaw_213_en.pdf. (Last accessed November 10, 2023), 2014.

Tables

Table S1: List of instruments and reference air tanks used for H_2 in air sample measurements in NOAA GML prior to 2010.

Tank Air Analysis						
Dates of operation	System	Instr. ID	Model	Response	Secondary standard tank ID	Notes
1993-1997	rgd2	R2	RGD2	Linear	CC73110*, CC71607	No electronic records of calibration prior to 2001. Later used as Target air for H9.
1997-2006	rgd2	R7	RGA3	Non-linear	CC73110*, CC71607	
2006-2008	cocal-1	P2	PP1	Non-linear	CC119811	See Figure S12
Flask Air Analysis						
Dates	System	Instr. ID	Model	Response	Working standard tank ID	Notes
1988-1990	rgd2	R2	RGA3	Linear	AAL-17262, CC68734*	* H_2 was assigned against 1996 gravimetric standards and early data was reprocessed (Novelli et al., 1998).
1990-1995	carle	R4	RGA3	Non-linear	AAL-17269*, AAL-17270*, CC105871*	
1995-1997	carle	R7	RGA3	Non-linear	CC105871, AAL-17259	Assignments for later working standards were mostly inferred from earlier tanks, assuming no drift.
1997-2010	MAGICC-1	R5	RGA3	Non-linear	CA02439,CA01493, CA02952,CA01777, CC61344, CA06593, CC105928, CC71649	
2004-2009	MAGICC-2	R6	RGA3	Non-linear	CA02439, CA06527, CC68676, CA06591, CC305198	

Table S2: H_2 secondary standards used in the tank calibration laboratory and H_2 tertiary standards used on the MAGICC-1 and MAGICC-2 systems (2009 to 2019). Assignments made on 2023-01-24 for all tanks except CA08107 and CB11090 for which assignments were entered on 2023-04-24.

Tank Calibration / H9								
Tank ID (fill)	Time of use	t0	Assignment at t0 (ppb)	C1 (ppb/yr)	C2 (ppb/yr^2)	N	Residuals standard deviation (ppb)	Fill date
CC119811 (A)	2/05/2008 to 6/02/2013	2010.0689	549.4	2.0	0	47	0.50	01/01/2006 SM*
CA03233 (B)	6/02/2013 to 11/01/2018	2016.7106	502.8	0	0	19	0.23	08/12/2010 NWR
MAGICC-1 / H11								
Tank ID (fill)	Time of use	t0	Assignment at t0 (ppb)	C1 (ppb/yr)	C2 (ppb/yr^2)	N	Assignment uncertainty (ppb)**	Fill date
CA08107 (D)	7/22 to 8/7/2019	2019.2959	562.9	15.4	0	6	0.6	11/9/2018 NWR
CB11090 (B)	10/18/2018 to 7/19/2019	2019.1482	576.3	6.9	0	4a	0.6 After 2019-06-21: 1.5	9/30/2016 NWR
CB11551 (A)	2/13 to 10/17/2018	2018.1878	548.8	6.7	0	3a,b, c	0.5 After 2018-08-27: 1.5	1/1/2015 SM*
CC91285 (C)	6/19/2017 to 2/13/2018	2017.1711	538.4	0	0	8	0.5	8/14/2015 NWR
CA08165 (B)	10/13/2016 to 06/16/2017	2016.9137	535.7	4.5	0	3c	0.5	12/16/2011 NWR
CC302566 (B)	3/21/2016 to 10/12/2016	2016.3645	540.2	4.4	0	5	0.5	8/14/2015 NWR
CC105491 (B)	8/10/2015 to 3/18/2016	2015.1506	522.3	0	0	5d	1.0	1/16/2014 NWR
ND33801 (B)	8/4/2014 to 8/7/2015	2013.8771	509.3	0.9	0	6e	0.5 After 2015-05-14: 1.0	12/27/2012 NWR
CB09117 (A)	2/18 to 8/1/ 2014	2013.8912	635.3	28.7	0	5	2	12/17/2012 SM
ND46735 (A)	9/10/2012 to 2/13/2014	2012.9158	527.4	2.5	-1.0	7e,f	0.5 After 2013-12-11: 1.0	1/1/2011 estimated
CA04505 (B)	12/9/2011 to 9/7/2012	2011.4593	540.6	1.7	0	3c,d	1.0	8/12/2010 NWR
ND38963 (A)	8/12/2010 to 12/7/2011	2011.704	586.0	6.2	0	4	0.5	1/1/2009 estimated
CC71649	1/22 to 8/6	2009.1184	507.1	8.4	0	7b,e	1.5	9/19/2008

(E)	2010							
MAGICC-2 / H8								
Tank ID (fill)	Time of use		Assignment at t0 (ppb)	C1	C2	N	Assignment uncertainty (ppb)**	Fill date
ND38954 (B)	3/26/2013 to 3/21/2014	2014.2094	516.6	2.0	0	5	0.5	12/9/2012 NWR
CA03409 (B)	5/23/2011 to 3/25/2013	2011.6278	526.6	0	0	5e	0.5 After 2013-01-21: 1.0	1/1/2010 estimated
ND38415 (A)	4/5/2010 to 5/20/2011	2010.2502	566.1	20.9	-8.7	6	0.5	1/1/2009 estimated
CC305198 (A)	11/2//2009 to 4/3/2010	2009.7211	557.9	65.8	0	3a,b	1.5 After 2010-01-31: 2.5	1/1/2009 SM*

* Gravimetric blends with CO, H_2, CO_2, CH_4 and N_2O in zero air purchases from Scott Marrin.

** Uncertainty estimates listed for the tertiary standard assignments assume a 0.5 ppb uncertainty for each calibration result on H9 and do not formally include the uncertainty on the secondary standard assignments.

a. Assignment does not use existing post-use calibration results that show larger drift
b. Drift change towards end of use, additional drift correction applied.
c. Force linear fit in drift calculation code
d. Only predeployment calibrations
e. No end-of-use or post-use calibration
f. Force quadratic fit in drift calculation code

Table S3: H_2 standards used on the MAGICC-3 system. Best polynomial curve fit coefficients to the August 2019-December 2022 calibration histories. Assignments made on 2023-01-24.

Tank ID (fill)	t0	Assignment at t0 (ppb)	C1 (ppb/yr)	C2 (ppb/yr^2)	N	Assignment uncertainty (ppb)	Fill date
CA01414 (I)	2020.0964	238.4	10.0	-1.9	9	0.5 ppb	12/29/2017 NWR
CA04403 (F)	2020.1052	474.6	10.2	-1.7	9	0.5 ppb	12/1/2017 NWR
CB11270 (A)	2020.0012	515.0	2.9	-0.5	9	0.5 ppb	12/1/2017 NWR
CA06388 (H)	2019.9423	551.2	1.1	0	9	0.5 ppb	2/23/2018 NWR
CA05773 (F)	2020.2585	565.6	1.4	0	8	0.5 ppb	5/17/2018 NWR
CB11034 (B)	2020.0783	580.1	8.3	-1.2	9	0.5 ppb	5/17/2018 NWR
CA05680 (H)	2020.0904	588.1	1.9	0	9	0.5 ppb	12/1/2017 NWR
CB11405 (C)	2020.1474	605.6	23.3	-1.6	9	0.5 ppb	5/17/2018 NWR

Table S4: H9 target tanks and the polynomial best fits to their calibration histories between 2009 and 2022.

Tank ID (fill)	Calibration date range on H9	t0	Assignment at t0 (ppb)	C1 (ppb/yr)	C2 (ppb/yr^2)	N	Residual standard deviation (ppb)	Fill date (location if known) (R=Refilled)
CC311842 (A)	2019-2022	2020.9878	478.6	0	0	8	0.32	2009-09-04 (NWR)
ND33960 (C)	2018-2022	2019.9289	529.5	0	0	11	0.43	2014-03-05 (NWR)
CC121971 (G)	2019-2022	2021.0834	546.5	0	0	9	0.30	2012-05-10 (NWR)
CA06194 (B)	2019-2022	2020.7726	578.4	0	0	10	0.49	2008-09-25 (NWR)
ND16439 (A)	2008-2015	2009.66673	635.9	0	0	9	0.54	2002-01-01 (R)
CA08247 (A)	2020-2022	2021.2483	675.1	0	0	7	0.73	2008-10-01 (NWR)
CA05278 (A)	2008-2014	2011.8239	675.2	0	0	7	0.56	2007-03-01 (MPI) (R)
CA05300 (A)	2008-2014	2011.8667	596.8	0.84	0	7	0.31	2007-03-01 (MPI) (R)
CC71607 (A)	2008-2021	2016.889	537.9	0.44	0	18	0.34	1991-10-01
CC73110 (A)	2008-2021	2016.1309	563.8	0.79	0	19	0.41	1990-01-01 (NWR, SM Luxfer)
CA04551 (F)	2012-2016	2014.9953	523.18	4.55	0	42	0.32	2011-12-21 (NWR)
CA07328 (A)	2008-2010	2009.2785	598.7	2.83	0	6	0.20	2006-10-02 (SM, grav blend)
CB10910 (B)	2018-2022	2019.8396	577.28	3.51	0	11	0.40	2016-02-18
CC71579 (F)	2008-2012	2011.3385	605.6	7.74	0	26	0.36	2008-09-19 (NWR) (R)
CA08145 (C)	2016–2017	2016.7627	646.5	27.2	0	20	0.48	2015-08-14 (NWR)
ALM-065166 (A)	2008-2022	2014.6308	659.0	0.26	0	8	0.69	2006-01-01
CC309852 (A)	2009-2019	2015.1105	227.5	2.23	-0.39	9	0.93	2009-10-01 (SM, grav blend)
CC309852 (A)*	2011-2019	2015.7837	226.8	1.66	-0.16	8	0.36	2009-10-01 (SM, grav blend)
CC327035 (C)	2019-2022	2020.7333	370.5	5.76	-0.48	10	0.23	2017-10-13 (NWR)
CA07339 (B)	2018-2022	2019.9513	365.0	4.777	-0.32	11	0.37	2010-03-01 (BLD, CO grav blend)

CA06827 (I)	2019-2022	2021.1466	433.5	1.91	-0.30	15	0.27	2018-11-09 (NWR)
CA06327 (D)	2019-2022	2021.3555	437.0	2.94	-0.56	16	0.22	2018-11-09 (NWR)
ND15749 (A)	2008-2022	2014.5413	563.6	0.40	-0.02	22	0.27	2001-01-01
CC310014 (B)	2018-2022	2019.6369	572.9	-0.03	0.19	26	0.24	2010-04-29 (NWR)
ND16443 (A)	2008-2022	2015.0192	604.6	0.45	-0.03	20	0.32	2001-01-01
ND17445 (A)	2008-2022	2014.9725	632.9	0.99	-0.07	22	0.46	2001-01-01
ND17435 (A)	2008-2022	2015.3295	686.9	0.47	-0.05	19	0.76	2001-01-01
CA05554 (B)	2010-2016	2014.7948	699.67	0.85	0.46	53	0.83	2009-10-23 (NWR)

* Alternative assignment when the tank first calibration result, 5 weeks after its fill date in 2009, is dropped from the fit.

Table S5: MAGICC systems target tanks and the polynomial best fits to their calibration histories between 2009 and 2022.

Tank ID (fill)	Calibration date range on H9	t0	Assignment at t0 (ppb)	C1 (ppb/yr)	C2 (ppb/yr^2)	N	Residual standard deviation (ppb)	Fill date (location if known)
CC1824 (H)	2009-2011	2010.1738	574.5	6.22	0	4	0.51	2006-07-06 (NWR)
CB08834 (B)	2011-2018	2015.6272	537.8	4.06	-0.50	10	0.57	2011-10-20 (NWR)
CC303036 (A)	2010-2017	2013.1491	588.3	21.31	0.47	10	0.44	2008-12-04 (NWR)
CB11143 (C)	2019-2022	2020.6759	534.7	1.91	0	9	0.54	2018-11-01 (NWR)
ALMX067998 (C)	2016-2022	2019.4574	542.1	0.62	0	13	0.28	2016-02-12 (NWR)
CB10292 (B)	2020-2022	2021.4553	597.4	0.95	0	5	0.44	2019-10-17 (NWR)
SX-1009237 (A)	2022-2023	2021.1697	526.5	0	0	2	0.24	2022-11-16 (BLD)

Table S6: Summary statistics for SPO flask pair H_2 differences. Npairs= Number of flask pairs.

System/ Instrument	SPO “P” flasks Absolute differences			SPO “S” flasks Absolute differences			SPO ”S”-”P” Pair mean differences		
	Mean (ppb)	Std dev (ppb)	Npairs	Mean (ppb)	Std dev (ppb)	Npairs	Mean (ppb)	Std dev (ppb)	Npairs
MAGICC-2 / H8	1.3	1.0	165	1.1	0.9	87	-0.4	1.5	81
MAGICC-1 / H11	0.9	0.8	292	0.9	0.8	143	-0.2	1.3	144
MAGICC-3 / H8	1.6	1.3	45	1.2	1.2	25	-0.1	1.7	25
MAGICC-3 / H11	0.7	0.6	76	0.8	0.6	35	-0.5	0.8	43

Table S7: List of sampling sites in the NOAA Cooperative Global Air Sampling Network with revised H_2 measurement records

Site Code	Name	Country	Lat (°)	Lon (°)	Elevation (meters)	Time from GMT
ALT	Alert, Nunavut	Canada	82.451	-62.507	185.0	-5 hours
AMY	Anmyeon-do	Republic of Korea	36.539	126.329	47.0	+9 hours
ASC	Ascension Island	United Kingdom	-7.967	-14.400	85.0	-1 hour
ASK	Assekrem	Algeria	23.262	5.632	2710.0	+1 hour
AZR	Terceira Island, Azores	Portugal	38.766	-27.375	19.0	-1 hour
BKT	Bukit Kototabang	Indonesia	-0.202	100.318	845.0	+7 hours
BMW	Tudor Hill, Bermuda	United Kingdom	32.265	-64.879	30.0	-4 hours
BRW	Barrow Atmospheric Baseline Observatory	United States	71.323	-156.611	11.0	-9 hours
CBA	Cold Bay, Alaska	United States	55.210	-162.720	21.3	-9 hours
CGO	Cape Grim, Tasmania	Australia	-40.683	144.690	94.0	+10 hours
CHR	Christmas Island	Republic of Kiribati	1.700	-157.152	0.0	-10 hours
CIB	Centro de Investigacion de la Baja Atmosfera (CIBA)	Spain	41.810	-4.930	845.0	+1 hour
CPT	Cape Point	South Africa	-34.352	18.489	230.0	+2 hours
CRZ	Crozet Island	France	-46.434	51.848	197.0	+5 hours

DSI	Dongsha Island	Taiwan	20.699	116.730	3.0	+8 hours
EIC	Easter Island	Chile	-27.160	-109.428	47.0	-7 hours
GMI	Mariana Islands	Guam	13.386	144.656	0.0	+10 hours
HBA	Halley Station, Antarctica	United Kingdom	-75.605	-26.210	30.0	-2 hours
HPB	Hohenpeissenberg	Germany	47.801	11.024	985.0	+1 hour
HUN	Hegyhatsal	Hungary	46.956	16.652	248.0	+1 hour
ICE	Storhofdi, Vestmannaeyjar	Iceland	63.400	-20.288	118.0	+0 hour
IZO	Izana, Tenerife, Canary Islands	Spain	28.309	-16.499	2372.9	+0 hour
KEY	Key Biscayne, Florida	United States	25.665	-80.158	1.0	-5 hours
KUM	Cape Kumukahi, Hawaii	United States	19.561	-154.888	8.0	-10 hours
LLN	Lulin	Taiwan	23.470	120.870	2862.0	+8 hours
LMP	Lampedusa	Italy	35.518	12.632	45.0	+1 hours
MEX	High Altitude Global Climate Observation Center	Mexico	18.984	-97.311	4464.0	-6 hours
MHD	Mace Head, County Galway	Ireland	53.326	-9.899	5.0	+0 hour
MID	Sand Island, Midway	United States	28.219	-177.368	4.6	-11 hours
MLO	Mauna Loa, Hawaii	United States	19.536	-155.576	3397.0	-10 hours
NMB	Gobabeb	Namibia	-23.580	15.030	456.0	+1 hour

NWR	Niwot Ridge, Colorado	United States	40.053	-105.586	3523.0	-7 hours
OXK	Ochsenkopf	Germany	50.030	11.808	1022.0	+1 hour
PAL	Pallas-Sammaltunturi, GAW Station	Finland	67.973	24.116	565.0	+2 hours
PSA	Palmer Station, Antarctica	United States	-64.774	-64.053	10.0	-3 hours
RPB	Ragged Point	Barbados	13.165	-59.432	15.0	-4 hours
SEY	Mahe Island	Seychelles	-4.682	55.532	2.0	+4 hours
SGP	Southern Great Plains, Oklahoma	United States	36.607	-97.489	314.0	-6 hours
SHM	Shemya Island, Alaska	United States	52.711	174.126	23.0	-10 hours
SMO	Tutuila	American Samoa	-14.247	-170.564	42.0	-11 hours
SPO	South Pole, Antarctica	United States	-89.980	-24.800	2810.0	+12 hours
SUM	Summit	Greenland	72.596	-38.422	3209.5	-2 hours
SYO	Syowa Station, Antarctica	Japan	-69.013	39.590	14.0	+3 hours
TAP	Tae-ahn Peninsula	Republic of Korea	36.738	126.133	16.0	+9 hours
TPI	Taiping Island	Taiwan	10.379	114.371	4.0	+8 hours
USH	Ushuaia	Argentina	-54.848	-68.311	12.0	-3 hours
UTA	Wendover, Utah	United States	39.902	-113.718	1327.0	-7 hours
UUM	Ulaan Uul	Mongolia	44.452	111.096	1007.0	+8 hours

WIS	Weizmann Institute of Science at the Arava Institute, Ketura	Israel	29.965	35.060	151.0	+2 hours
WLG	Mt. Waliguan	Peoples Republic of China	36.288	100.896	3810.0	+8 hours
ZEP	Ny-Alesund, Svalbard	Norway and Sweden	78.907	11.888	474.0	+1 hour

Figures

Figure S1. 2018-2022 H_2 calibration histories of eight MAGICC-3 standards

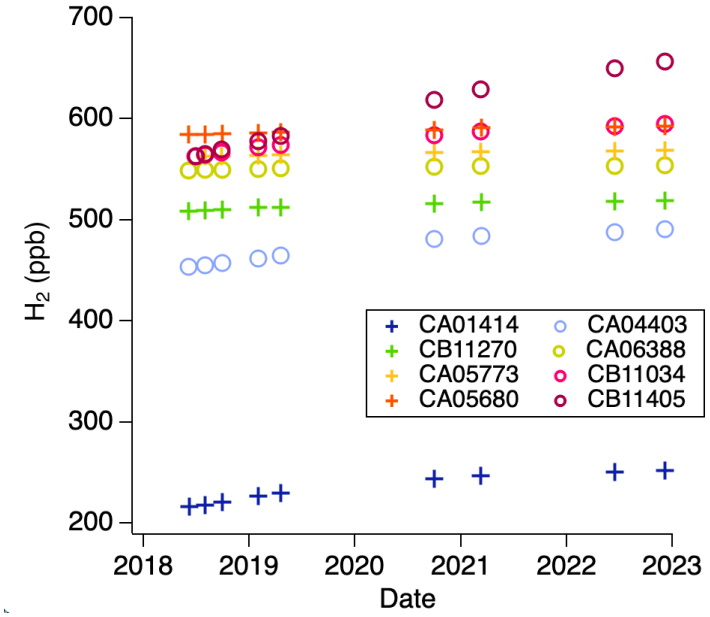


Figure S2: H_2 residuals from the 2018-2022 calibration history trend function for eight MAGICC-3 working standards (see Table S3)

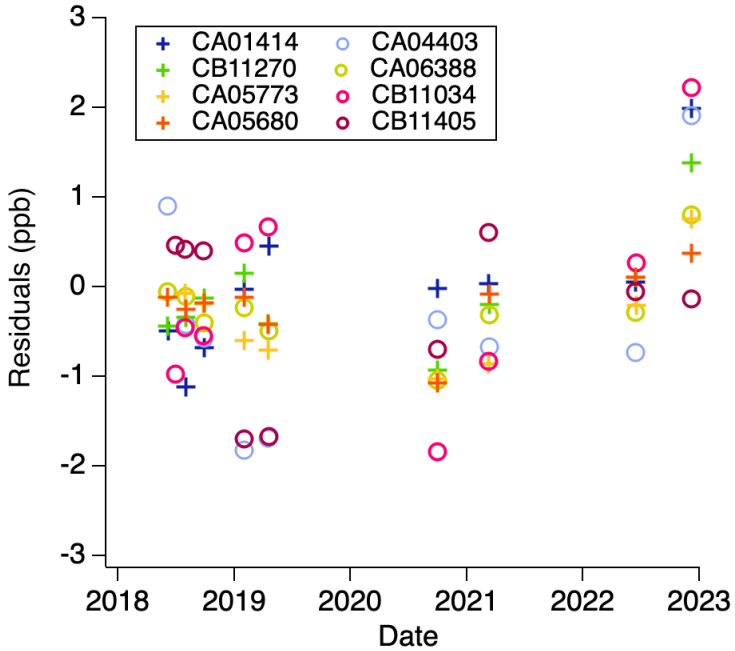


Figure S3. MAGICC-3 reference air deviation over time from mean H_2 derived from response curves with x=1 (in ppb).

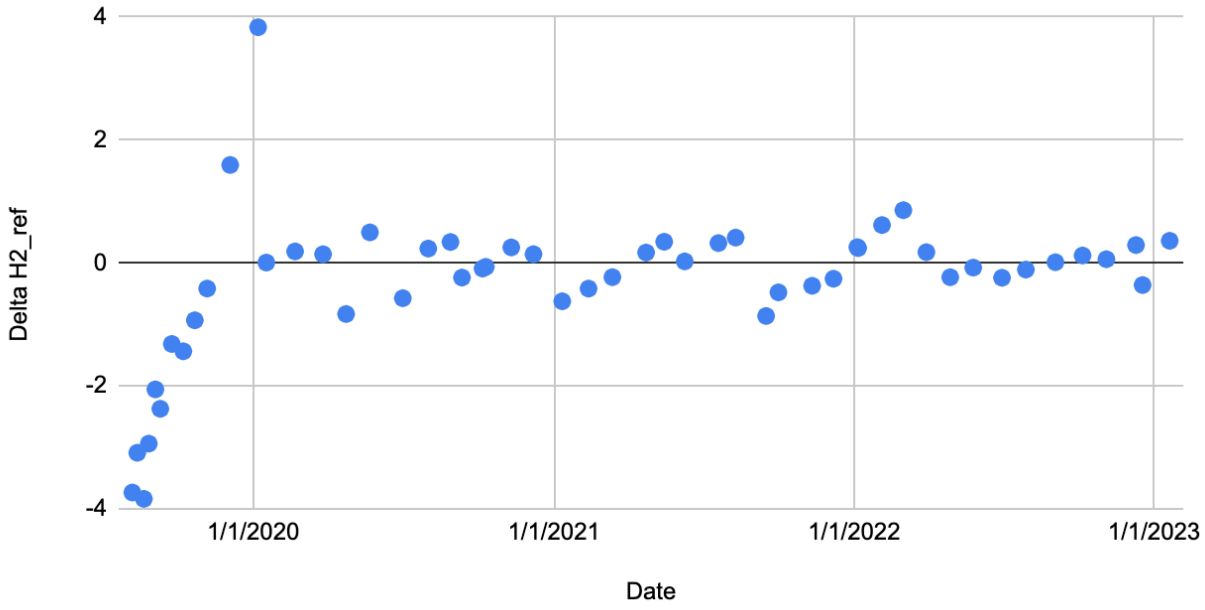


Figure S4. H9 Target tanks with quadratic polynomial fits to their calibration histories shown in plot a). Residuals from each tank best fit are shown in b) as a function of the initial assignment and c) as a function of the tank analysis date. d) Residuals standard deviation versus initial assignments (coef0) for all H9 Target tanks. For the lowest H_2 target tank (CC309852), we show the residuals to the first assignment (in plot c) and the standard deviation for residuals for the 2 assignments in Table S4 (in plot d). All values are in ppb.

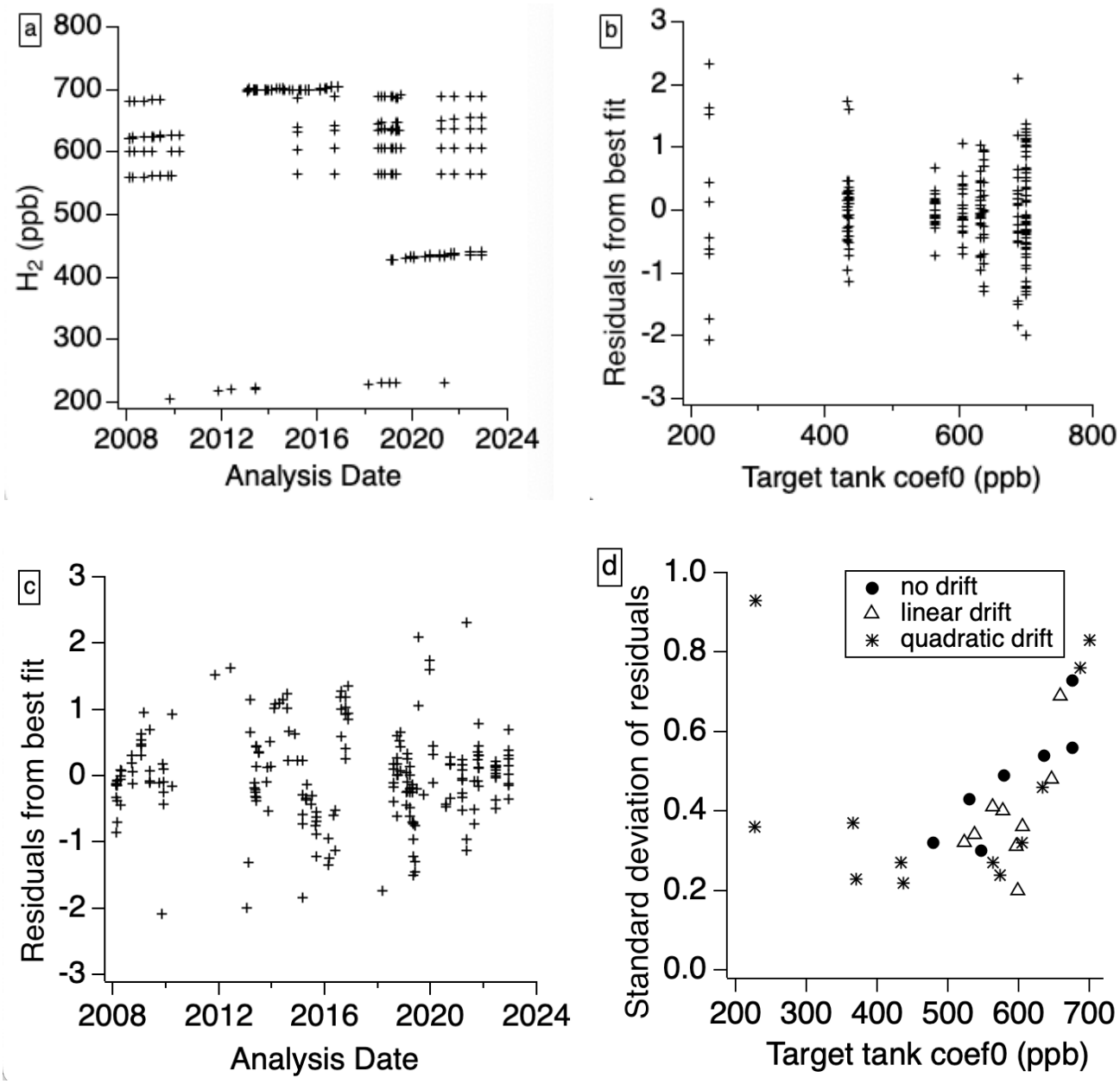
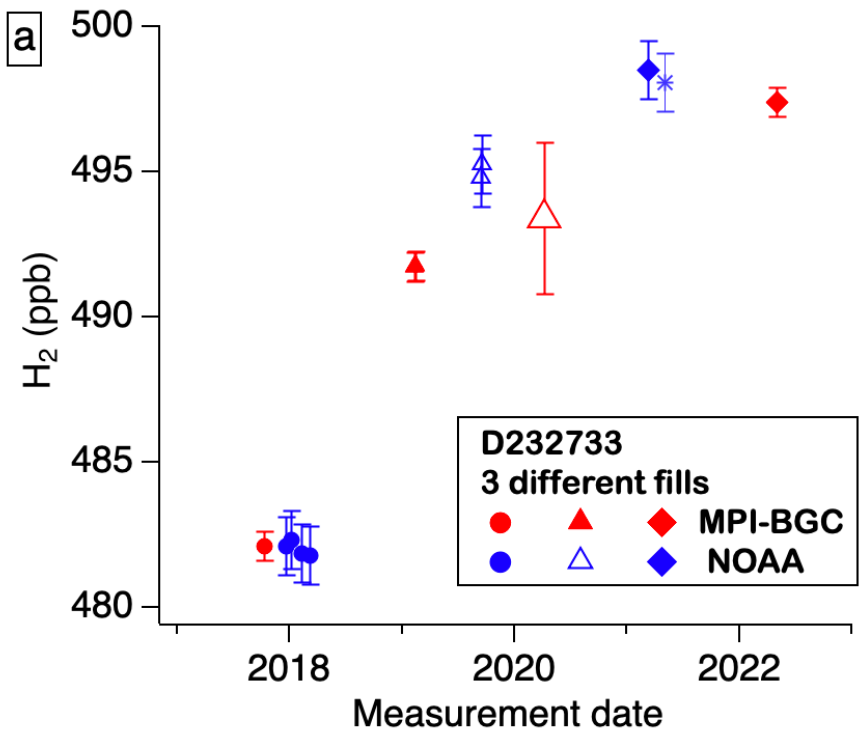
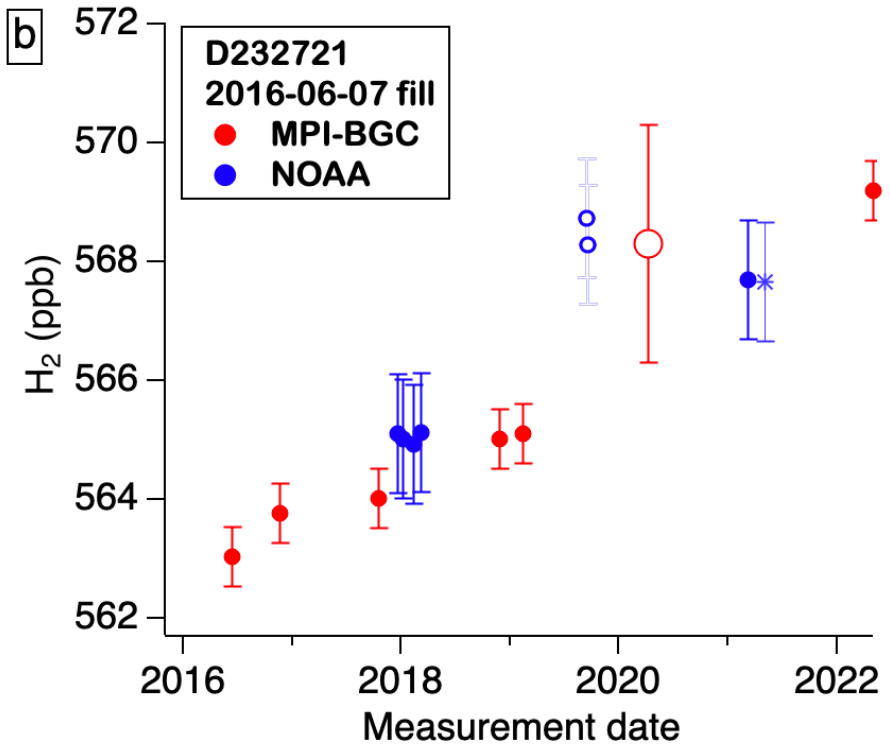


Figure S5: NOAA and MPI-BGC H_2 results for the MPI-BGC GasLab led MENI tank air measurement round robin comparisons (Jordan and Damak, 2022). NOAA measurement results on H9 are shown in blue. Blue open symbols and asterisks show rejected NOAA results due to poor instrument performance or the use of an alternate calibration strategy respectively. All H9 tank air results for the period September 12-18, 2019 were biased high by a few ppbs. The reason is unknown at this point. Most MPI-BGC results (red symbols) are on their GC-PDD instrument, except the April 2020 results are from their GC-RGA instrument (open red symbols). a) Cylinder D232733 is a blind sample and is refilled with different air after each round robin analysis loop. b) Ambient H_2 cylinder D232733 (~565 ppb) and c) low H_2 cylinder D232717 (~ 335 ppb) have slightly increasing H_2. The NOAA and MPI-BGC H_2 results agree well for the ambient and blind H_2 MENI tanks (< 1 ppb difference). The error bars indicate the reported reproducibility or the standard deviation when it is larger than the reported reproducibility.




b
D232721
2016-06-07 fill
MPI-BGC
NOAA
572
570
568
566
564
562
H2 (ppb)
2016
2018
2020
2022
Measurement date

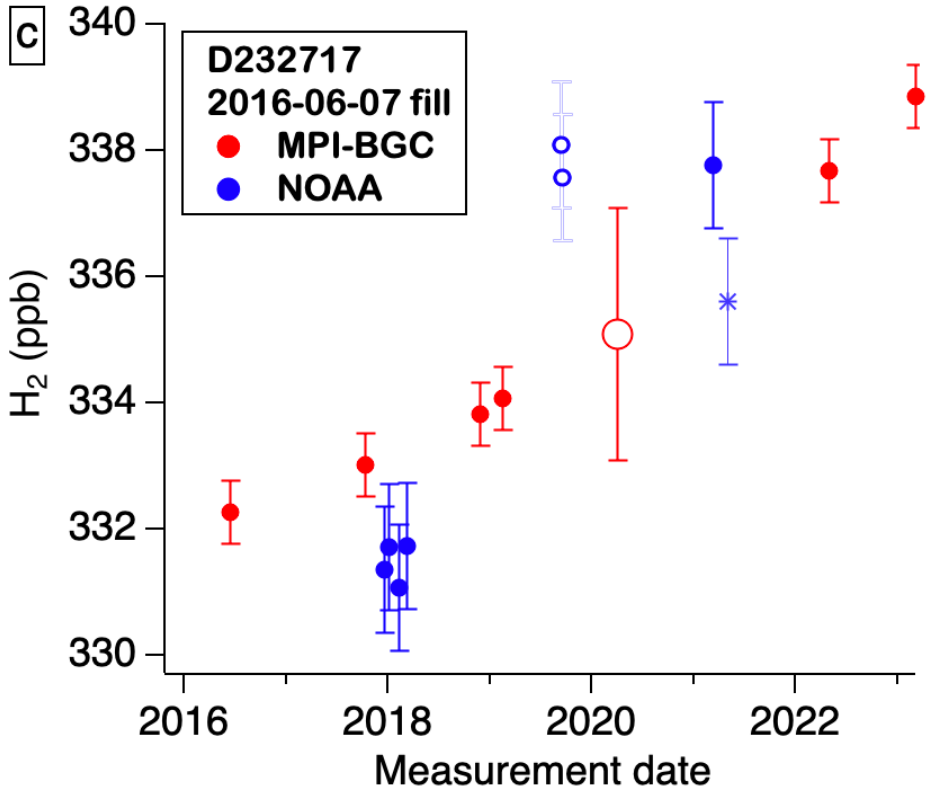

c
D232717
2016-06-07 fill
MPI-BGC
NOAA
340
338
336
334
332
330
H2 (ppb)
2016
2018
2020
2022
Measurement date

Figure S6. H_2 calibration histories of test air tanks 2008-2022. Each test air cylinder has a different color and different tank fills are shown with different symbols.

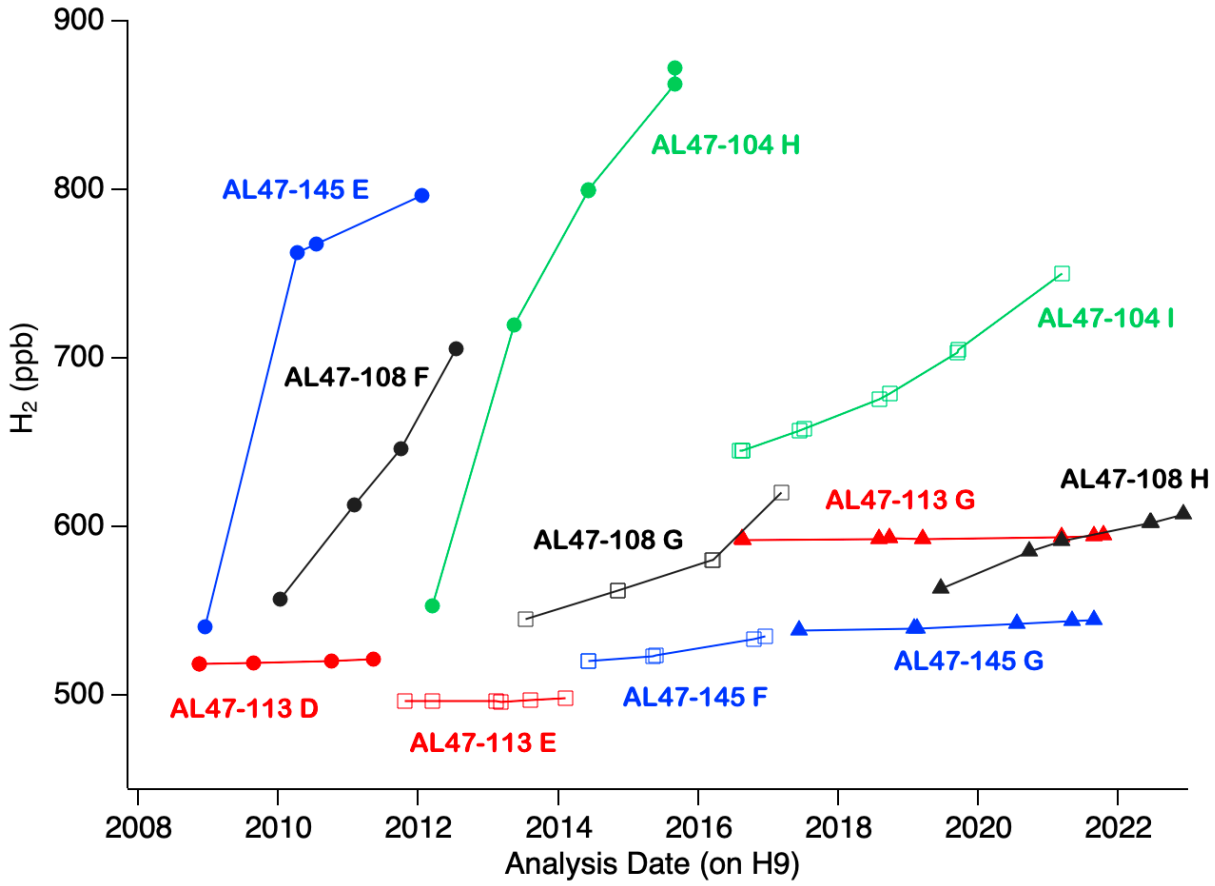
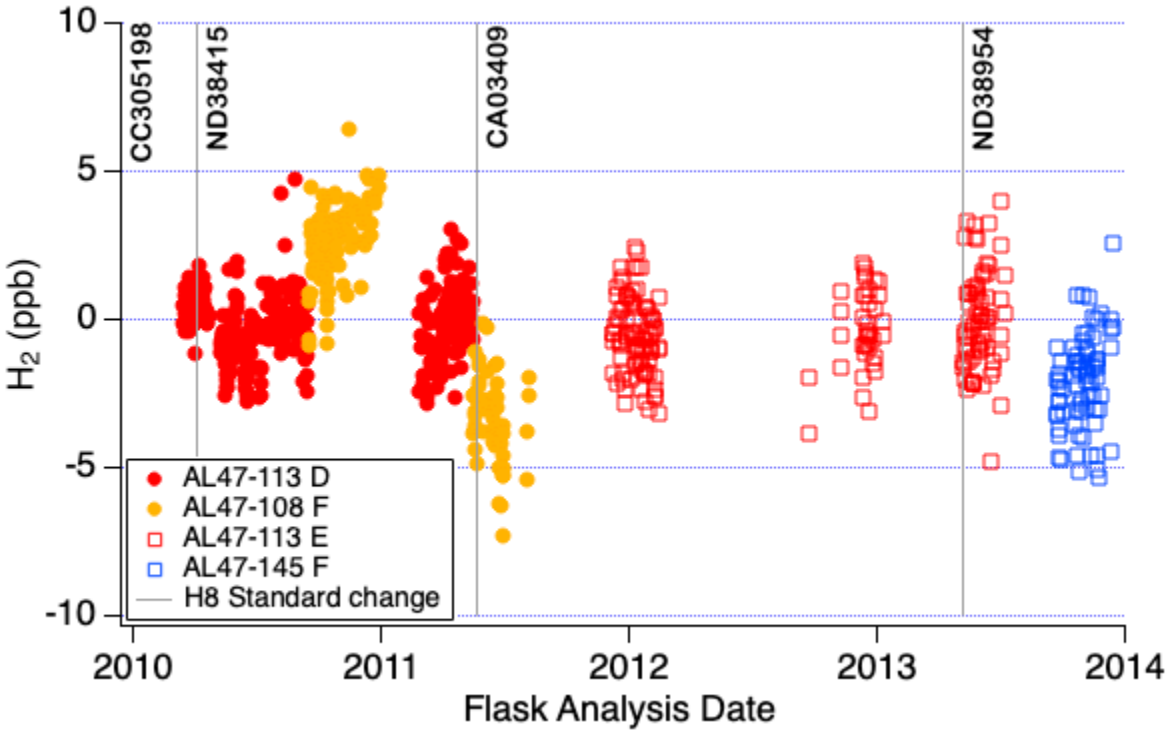
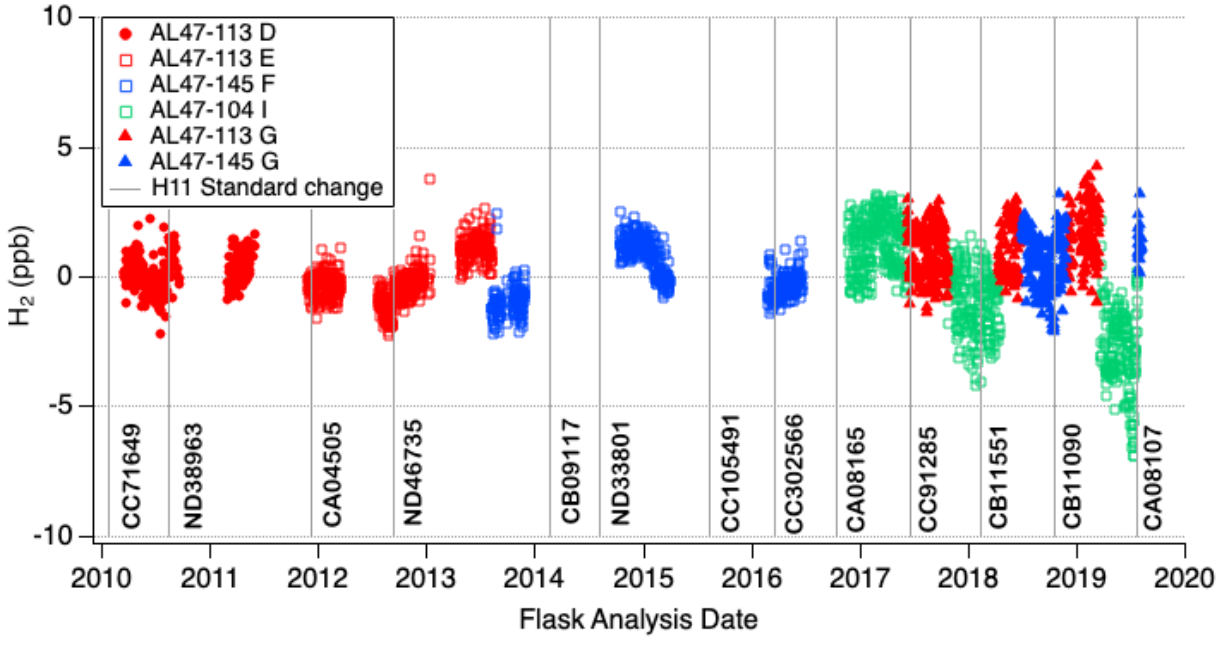


Figure S7. Test air (TST) flask analysis results : differences from test air tank time-dependent H_2 assignment: a) on H8, b) on H11 and c) on MAGICC-3.

a)



b)



c)

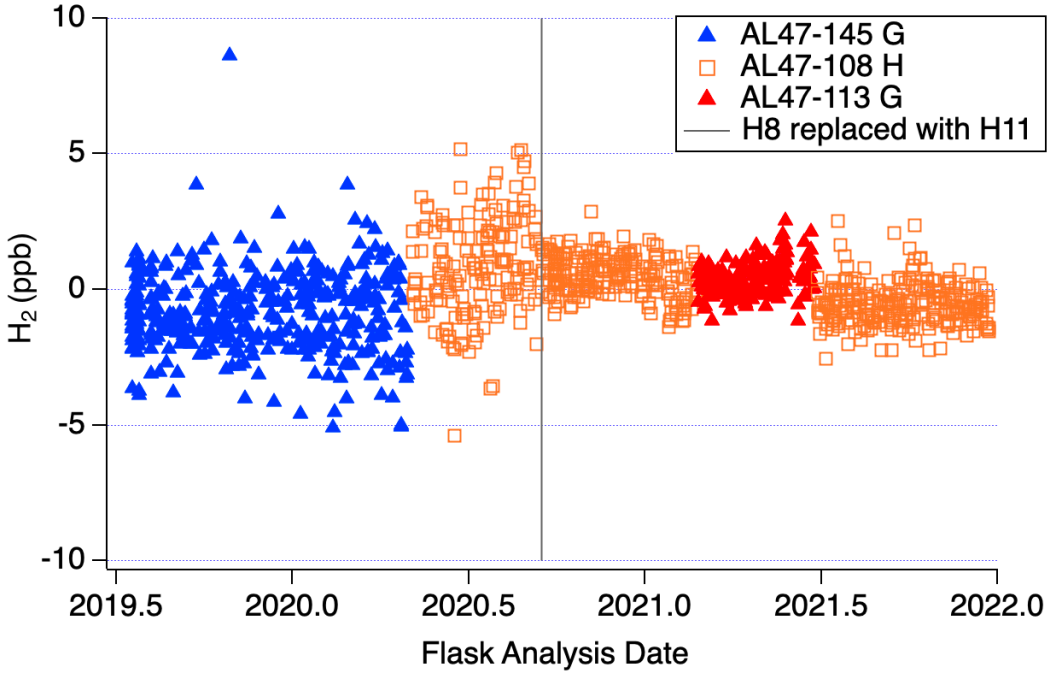


Figure S8: NOAA Cooperative Global Air Sampling Network site map (https://gml.noaa.gov/dv/site/). The four NOAA atmospheric baseline observatories (BRW, MLO, SMO, SPO) are shown in blue.

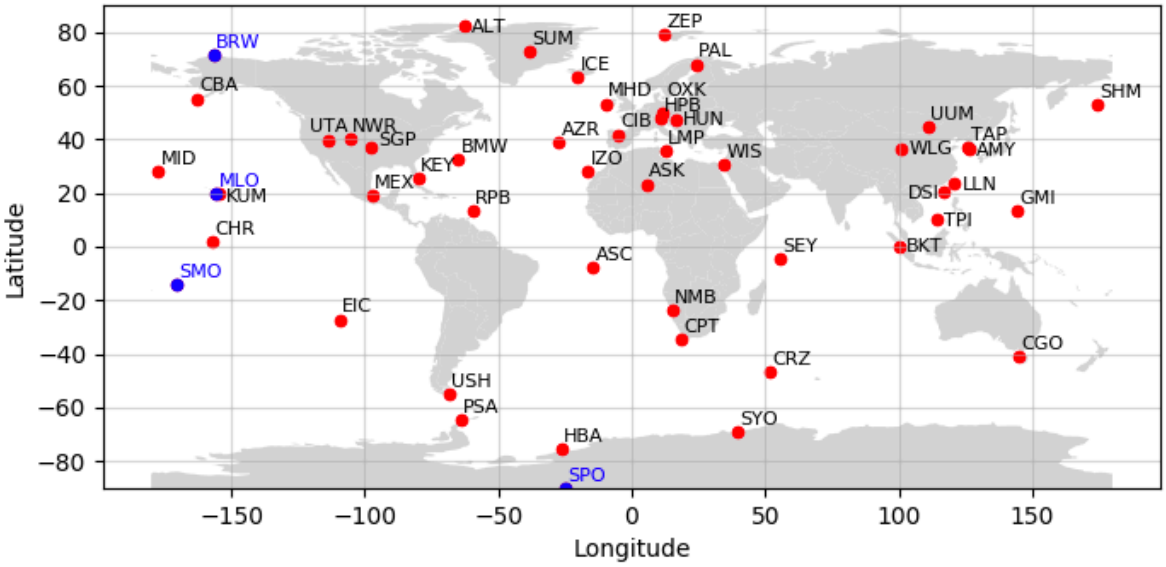
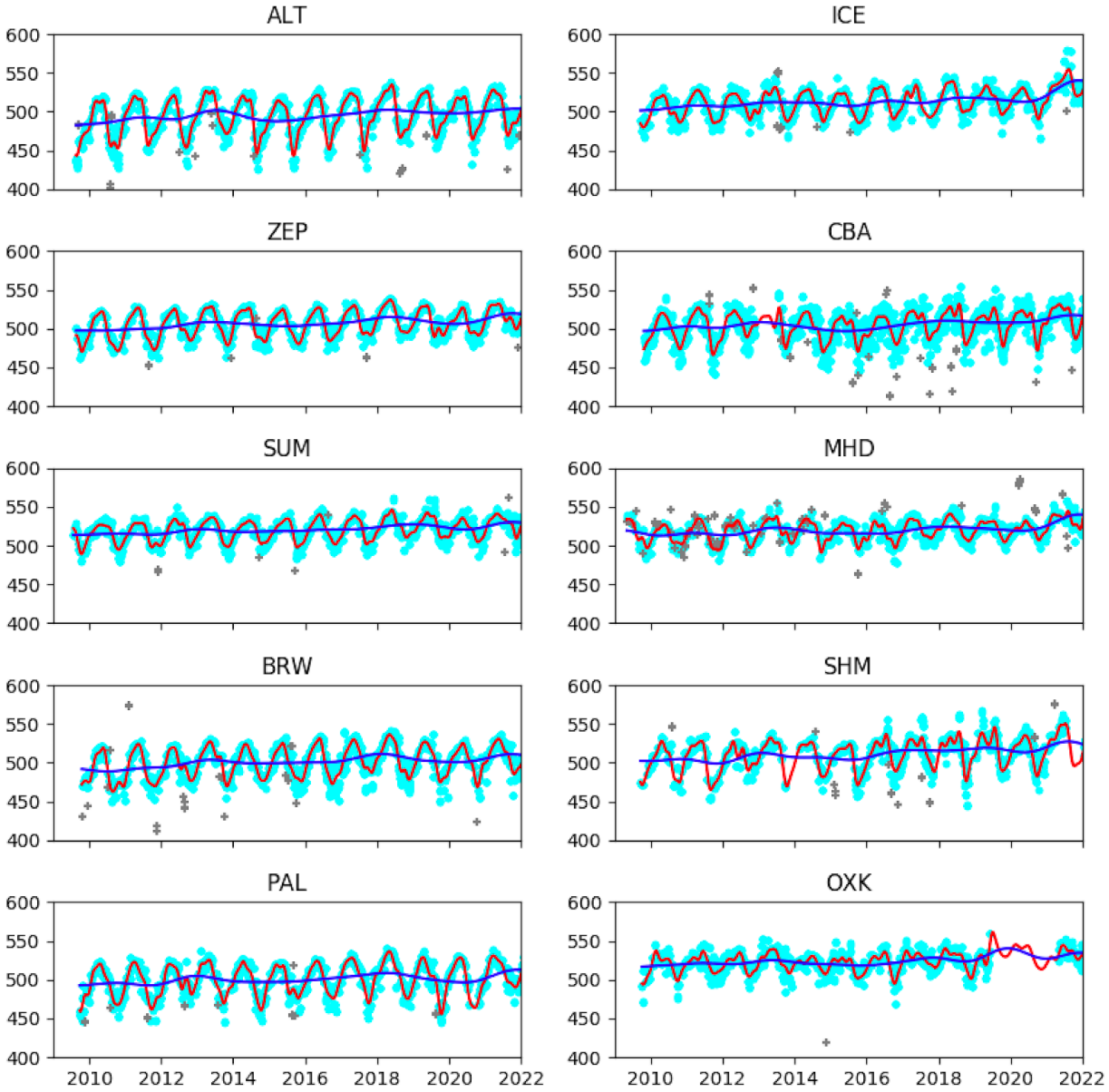
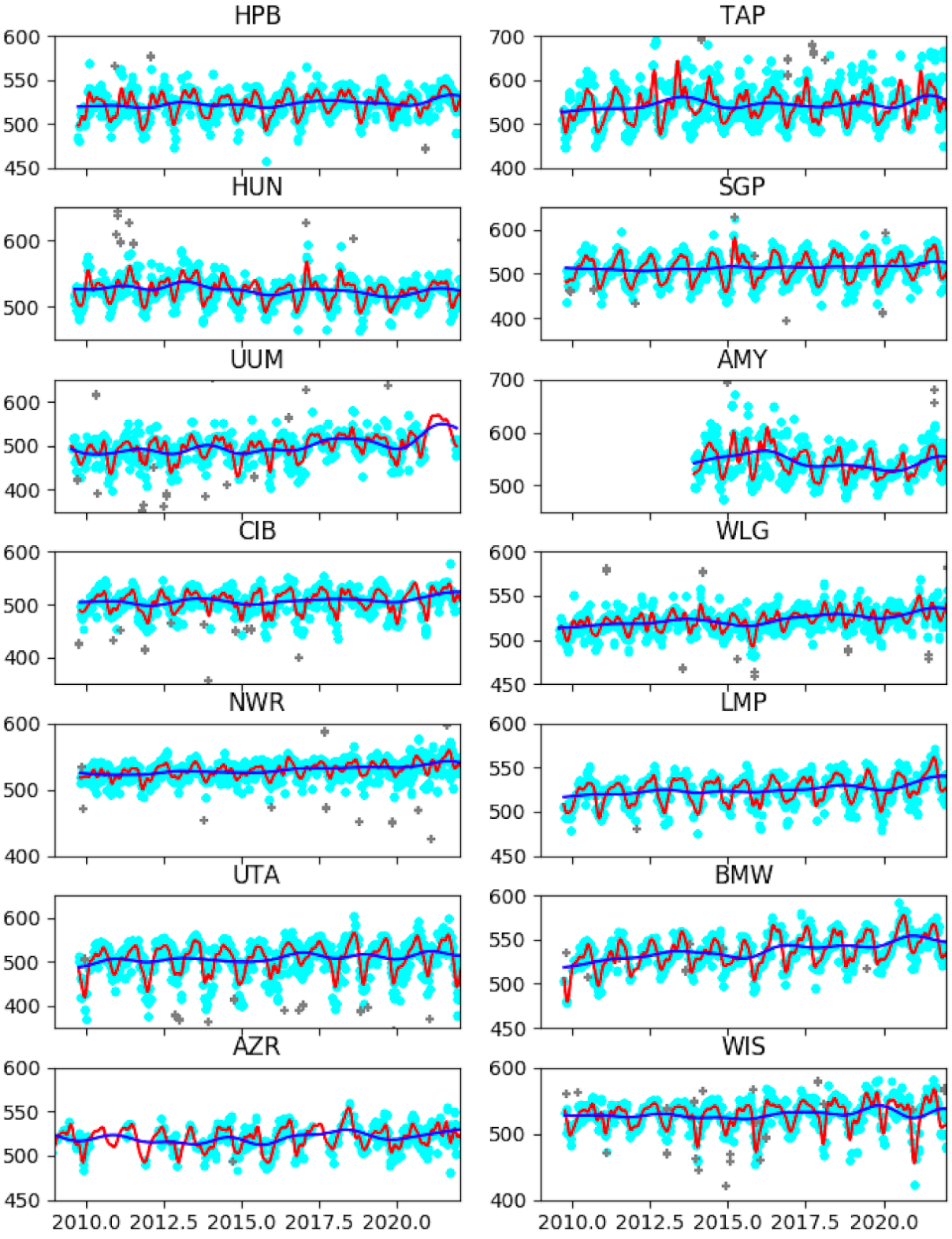
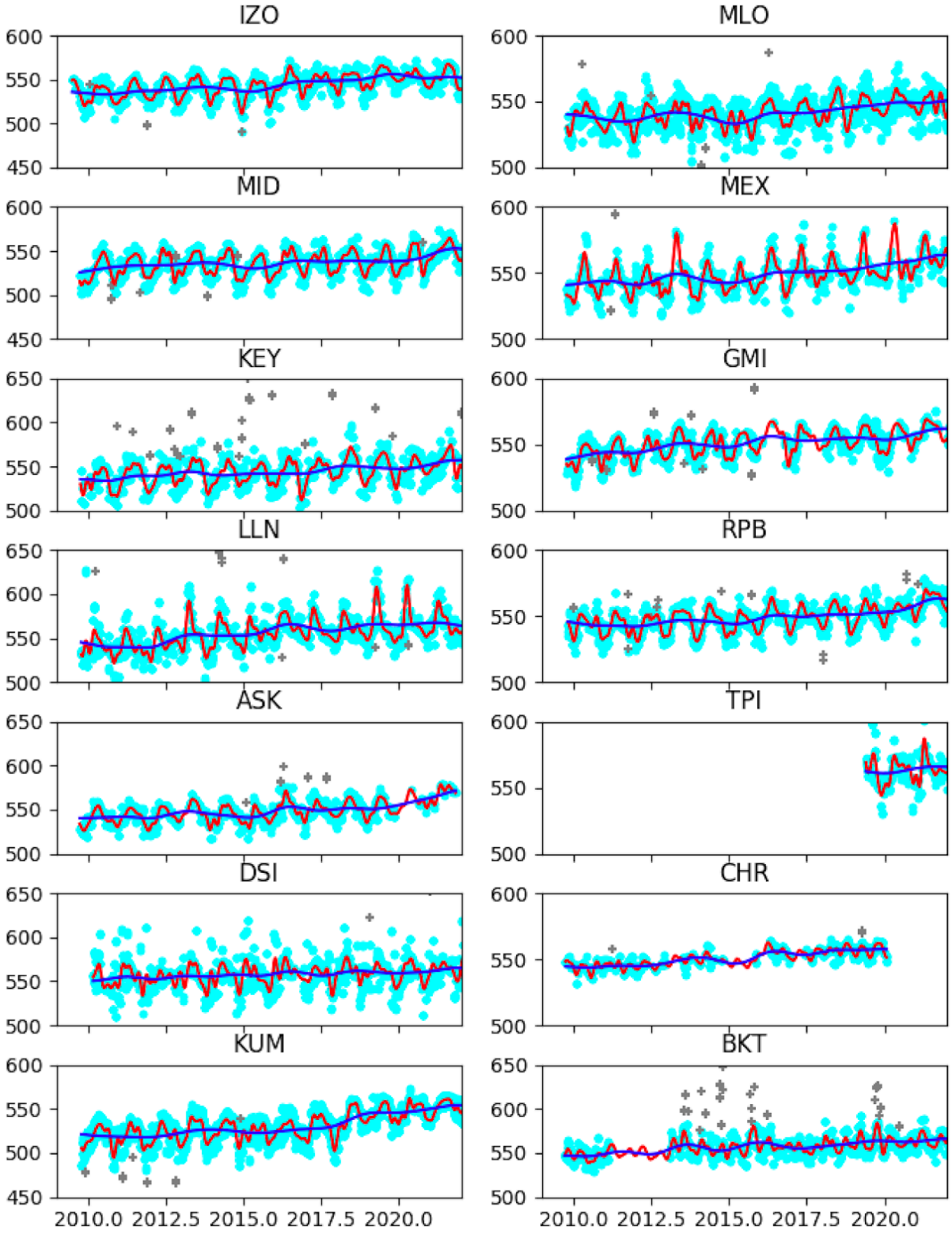


Figure S9: Discrete air H_2 mole fraction (in ppb) time series at 51 sites from the NOAA Cooperative Global Air Sampling Network. Data in light blue symbols are retained and data shown in gray crosses are deemed to be non-background. Rejected data are not shown but are present in the site data files. A curve fit python code is run for each site H_2 time series based on Thoning et al. (1989). First the code optimizes parameters for a function made of a four-term harmonic and a cubic polynomial. The resulting residuals are then smoothed with a low-pass filter with a 667 day cutoff and are added to the polynomial part of the function to produce the "trend curve" shown as the dark blue line. The residuals are also smoothed with a low-pass filter with a 80 day cutoff and are added to the function to produce a "smooth curve". The last plot shows all retained H_2 measurements from the Pacific Ocean Shipboard (POC).




HPB
TAP
HUN
SGP
UUM
AMY
CIB
WLG
NWR
LMP
UTA
BMW
AZR
WIS
2010.0 2012.5 2015.0 2017.5 2020.0
2010.0 2012.5 2015.0 2017.5 2020.0


IZO
MLO
MID
MEX
KEY
GMI
LLN
RPB
ASK
TPI
DSI
CHR
KUM
BKT
2010.0 2012.5 2015.0 2017.5 2020.0
2010.0 2012.5 2015.0 2017.5 2020.0

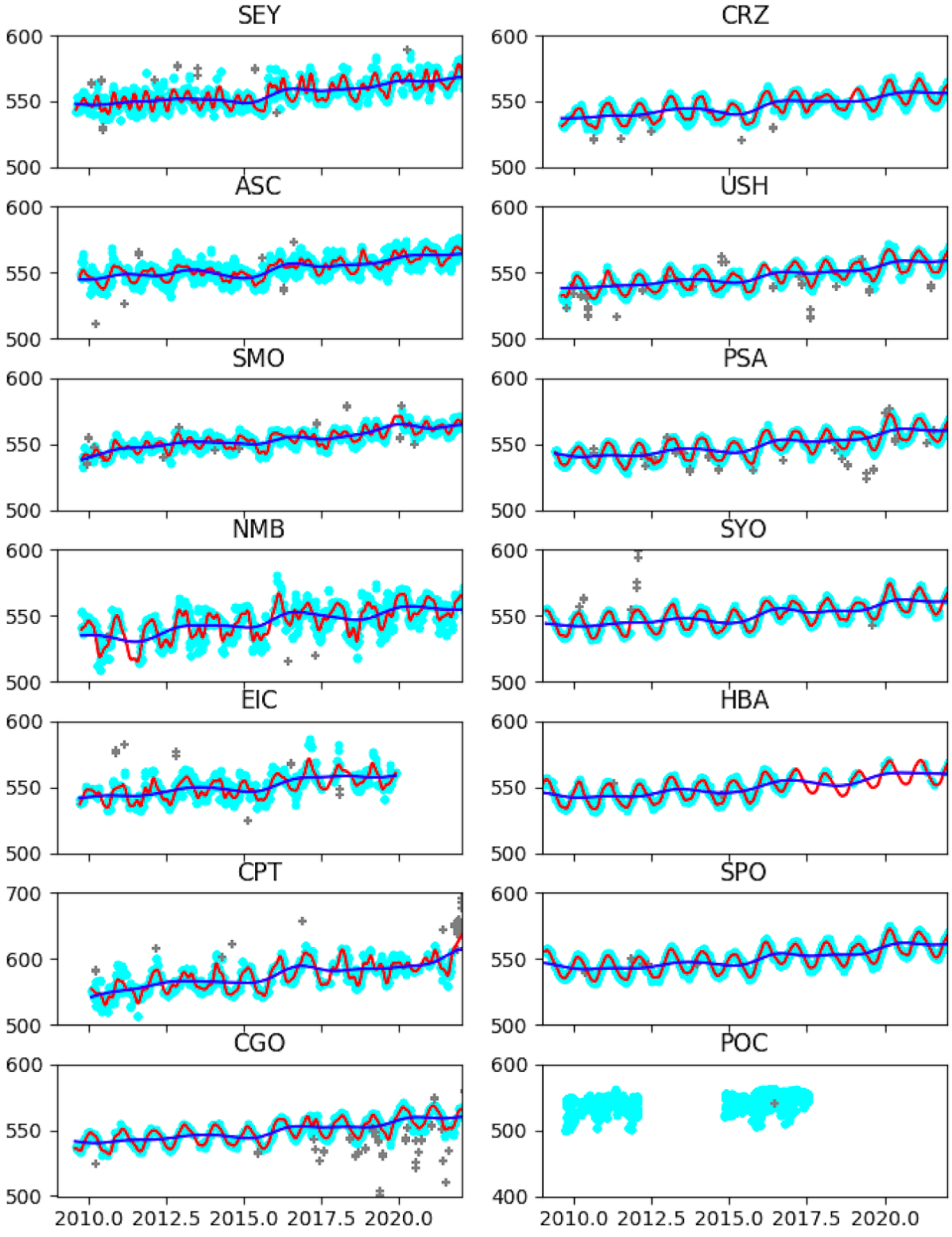

SEY
CRZ
ASC
USH
SMO
PSA
NMB
SYO
EIC
HBA
CPT
SPO
CGO
POC
600
550
500
700
400
2010.0
2012.5
2015.0
2017.5
2020.0

Figure S10: Marine boundary layer global mean and zonal mean H_2 (black, left side y axis) and CO (dashed blue line, right y axis) 2010-2021 time series

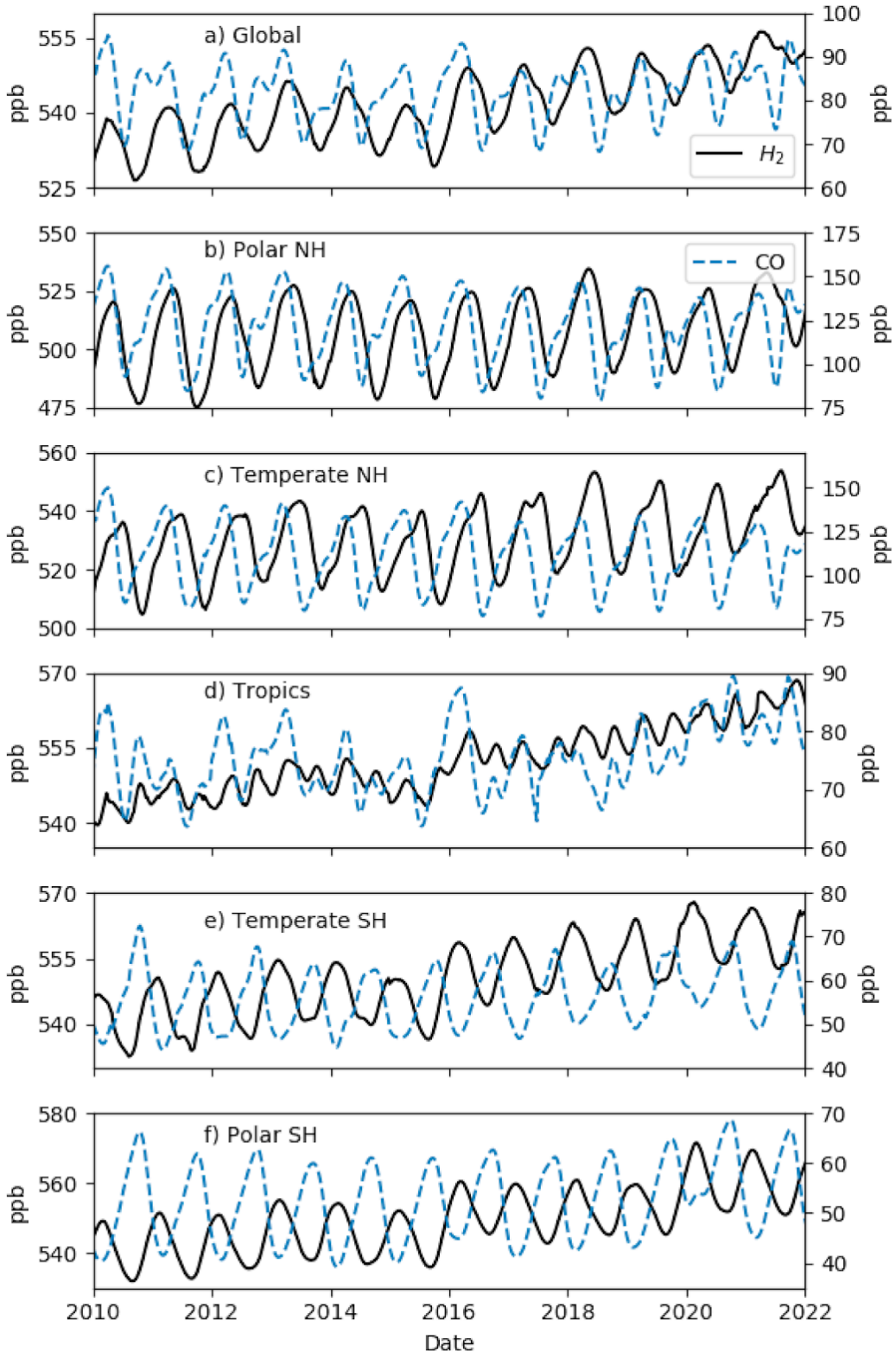


Figure S11: NOAA H_2 (blue, left axis) and CO (orange, right axis) measurement times series for three Cooperative Global Air Sampling Network sites in Iceland (ICE: 63.400°N, 20.288° W, 118 masl), Indonesia (BKT: 0.202° S, 100.318° E, 845 masl) and Tasmania, Australia (CGO: 40.683° S, 144.690° E, 94 masl).

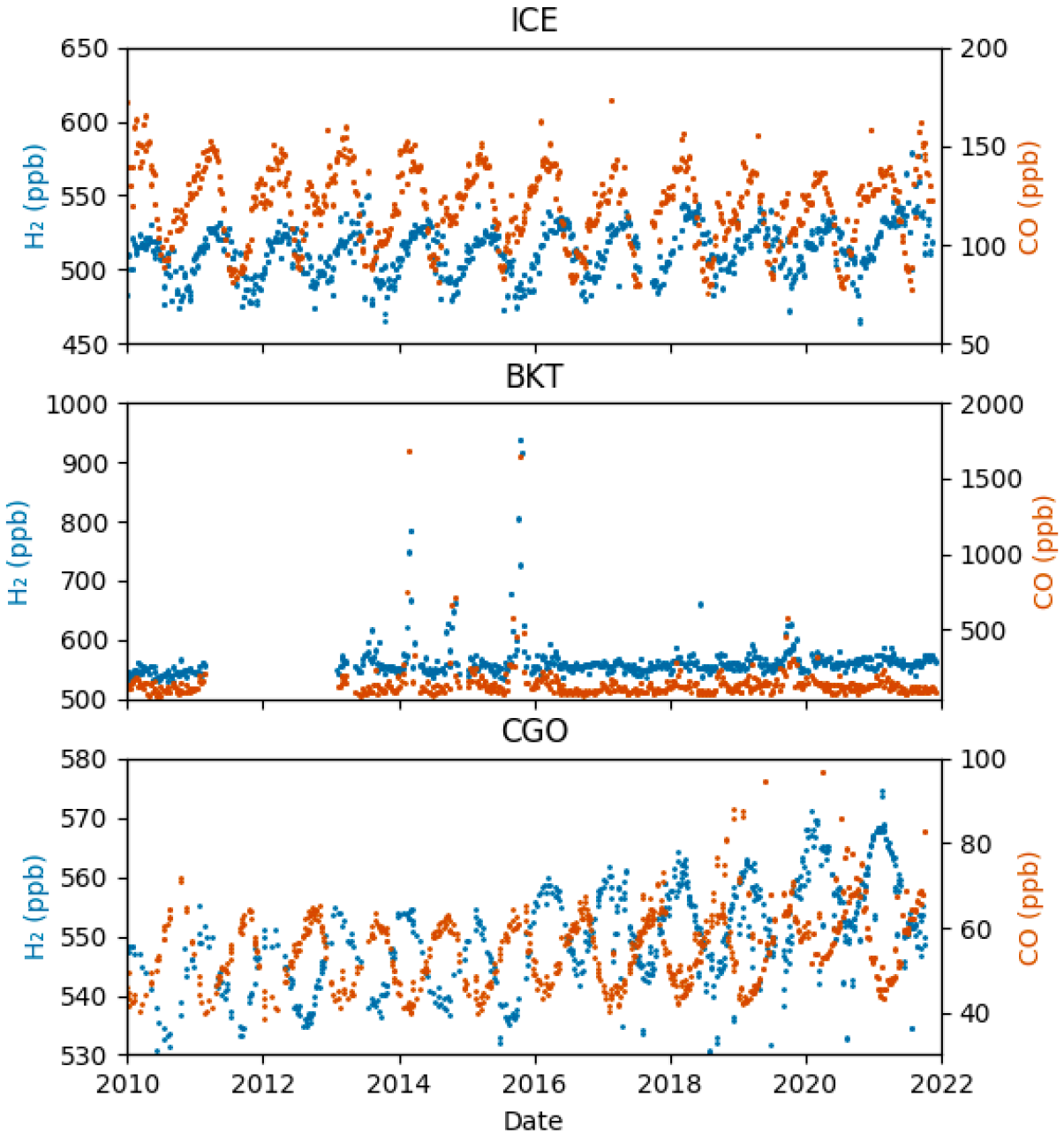


Figure S12: NOAA H_2 secondary standard CC119811 results on Peak Labs instrument (P2) and on GC-HePDD H9 using one point calibration against one of the primary standards.

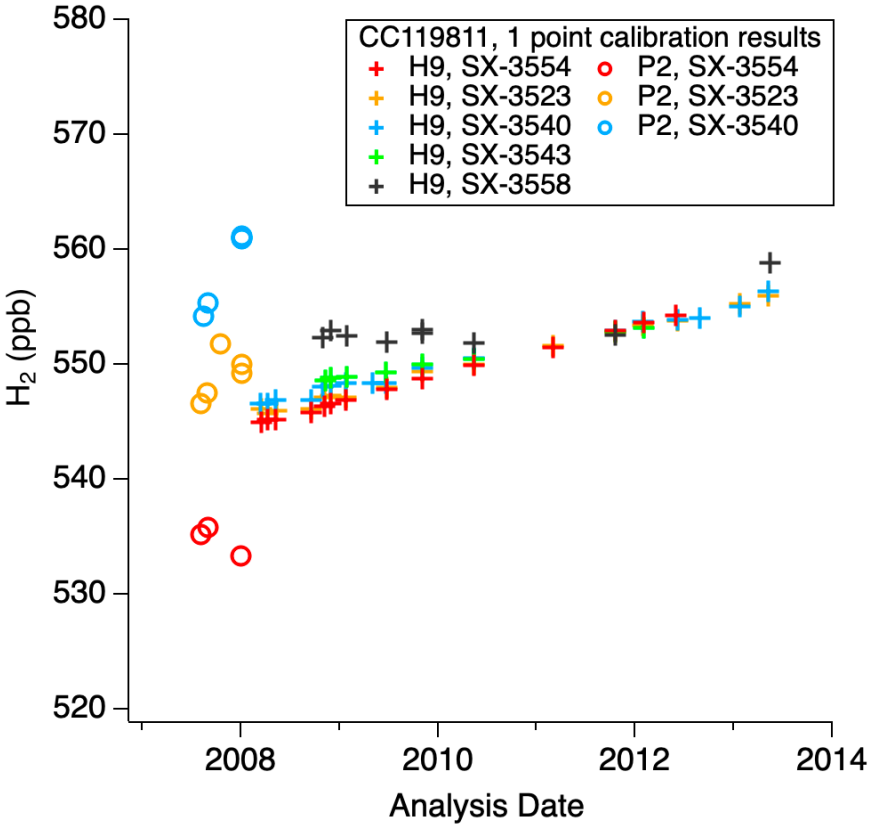


Figure S13: Early target tanks measurement records on different instruments using one point calibration. The working standard/reference tank ID for the measurements on RGA instruments is indicated in the legend.