**S2. Same air comparison with CSIRO for NOAA historical H2 data**

CSIRO started measuring H2 by gas chromatography with mercuric oxide detector (GC-HgO) using a Trace Analytical RGA instrument in 1991. Data are reported in the MPI X2009 H2 calibration scale. CSIRO’s implementation of the scale in recent years is defined by a suite of 5 H2-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 CH4, H2 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 H2. Dilution ratios were determined by measurements of CH4 tied to a gravimetrically defined CH4 calibration scale. They found the instrument response was “significantly non linear” and of similar shape for both H2 and CO (of the form y=ax2+bx+cxd, where x = peak height and a,b,c,d are estimated parameters from the response function fit). A single response function was used for H2 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 H2 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 analysed 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 non-linear response of the H2 analytical system detector, the instability of H2 standards stored in aluminum cylinders (commonly used for CO2 and CH4 standards) and the different calibration scales were presented as likely explanations for the observed time dependent biases between the two labs.