Responses to RC2: 'Comment on amt-2024-4', Anonymous Referee #2, 06 Mar 2024 reply

This very thorough and detailed manuscript addresses the recalibration of the extensive 12-year NOAA global atmospheric hydrogen (H_2) measurement program to the MPI X2009 calibration scale. It does this by adjusting for the significant and variable rates of growth of H_2 in the secondary and tertiary standard aluminum calibration tanks used for these measurements that were originally based on the NOAA X1996 calibration scale. As the manuscript notes, this has become immensely important work as mankind moves increasingly toward H_2 as an energy storage medium that comes with a large potential to leak into the atmosphere and fundamentally alter the oxidation capacity of the atmosphere with respect to lengthening the atmospheric lifetimes the of methane and other anthropogenic greenhouse gases.

While acknowledging this importance, I had considerable difficulty reading the manuscript because of its distractingly minute detail. One example is the inclusion of the serial numbers of the instruments used to make the measurements. Another is the extensive comparisons of the performance of the three different "MAGICC" instrument systems used to make these measurements over time. This is too much detail for the average reader's attention span, even though it is important that it be recorded somewhere.

One solution would be for the main paper to outline the principles of what was done with a few illustrative examples, and to move the bulk of the details that should be recorded somewhere either to the supplementary information (SI) addendum together with this AMT paper, or to a separate project report published by NOAA and available through the NOAA GML website. I recommend distilling the main paper to something like one third to one half of its present length, while still conveying the rigor of this important work, including a few examples, and providing the full details elsewhere.

This paper aims to fully document the recent revisions affecting the NOAA 2009/2010-present H_2 measurement records, which together make the largest global data set for H_2. We have moved 3 tables to the SI and removed some duplicative or extraneous text.

Thank you for your advice to streamline some of the information or move some content to the SI to not burden the reader.

A few more specific comments follow:

1) Is NOAA GML now using stable stainless steel gas cylinders to propagate its calibrations going forward so that drift adjustments resulting from the use of aluminum cylinders will no longer be an issue?

GML uses multi gas standards in aluminum cylinders in the flask analysis laboratory. We are still exploring options as explained in Reviewer 1, Question #3.

2) The dry air mole fraction ppb H_2 concentration units are not defined when they are first used.

The definition of the mole fraction unit (ppb) has been added in the introduction.

3) Please explain why calibration scales are necessary for atmospheric research, as compared to using individual calibration standards that are not related to a specific scale. This important point is not widely appreciated, especially among national metrology institutes.

We have added the text below at the beginning of section 2 in the main manuscript.

> To reliably infer fluxes and changes in fluxes over time from atmospheric measurements, scientists need to detect small temporal and spatial gradients in the abundance of trace gases. This requires comparable data across time and across monitoring networks to ensure biases are minimized and do not influence interpretation. The use of a common calibration scale among measurement laboratories ensures data are traceable to a common reference. It is the first step in preventing biases that could stem from using different references.

4) The word "best", used in line 170 to describe the post-2009 NOAA H_2 data, is subjective. A better term might be "most precise", or something to that effect.

"Best" as been replaced in the sentence below:

> The ~~best~~ more precise and better calibrated NOAA H_2 measurement records date back to 2009/2010 and are the main focus of this paper.

5) Following on the above comment, NOAA prepared their X1996 calibration scale well before the measurements that are recalibrated in this paper. Are there pre-2009 NOAA H_2 data that could still benefit from being recalibrated to the MPI calibration scale despite their lower precisions?

Unfortunately no. Please see SI section S1 and SI Figures 12 & 13 for our detailed explanation.

6) Should Paul Novelli, and perhaps Ed Dlugokencky, be coauthors of this work? They are both retired from NOAA, but much of this work was done by them.

Paul Novelli retired in 2017 and he was the science lead that started the CO and H2 measurements at NOAA. The paper references Paul's early work and papers and it makes it clear the NOAA H_2 work originated with him. Co-author Andy Crotwell streamlined the calibration scale propagation and Andy Crowell and Brad Hall worked with Paul on the adoption of the new instrumentation and the preparation of the new gravimetric mixtures in Essex tanks (Novelli et al., 2009). None of this work would have been possible without Andy's analytical and technical expertise and leadership. Andy implemented a strict calibration standards hierarchy for CO and H_2 and he led the adoption of best practices to store and manage data files. Andy has been the technical lead in charge of the calibration work for multiple gases (CO_2, CH_4, CO, H_2) at NOAA GML for many years, working closely with Brad Hall and others. Ed Dlugokencky retired from NOAA GML in spring 2023. He was not involved with the H_2 measurements or the CCL calibration scale adoption. We reference two papers by Ed when we give background information about the NOAA GHG measurements in samples from the Cooperative Global Air Sampling Network.

The COVID-19 pandemic restricted lab access for many months and provided Andy and several of us in GML time to work on the H_2 dataset. It was a group effort and the NOAA GML and CIRES staff involved as well as our close collaborators on H_2 in other laboratories (MPI BGC, CSIRO, UCI) are co-authors on the paper.

7) The Teflon o-rings used in the flask stopcocks (line 568) are highly permeable to H_2 and other gases. Since the flasks are pressurized and H_2 permeates much faster than N_2 and O_2, shouldn't the H_2 mole fraction decrease with time between sampling and analysis? Has this been tested?

The global network flasks are pressurized to 17-19 psia. We have not conducted systematic storage tests to evaluate if the permeability of teflon affects the integrity of H_2 in the air samples. The GML measurement team plans to do sampling equipment and flask storage tests in the not too distant future to carefully evaluate the stability of a suite of gases.

For now, we have looked at samples collected close in time, and analyzed at different times leading to different storage times. Please see the response to Reviewer 1, Question 20.

8) Please use the widely accepted spelling of "gases" instead of "gasses". Please also change "data is" to "data are" since "data" is the Latin plural of "datum".

Changes have been made. Thank you for pointing this out.

9) Given the range of coauthors and institutions involved in this work, and the number of years of data that are corrected, I assume that the acknowledgments and the financial support (lines 958-966) may be incomplete.

Please see the earlier response to your question 6.

Summary:

This is important work that should be published, and AMT is an appropriate venue for this. But it should first be distilled to a more readable structure that conveys the principles of what was done and summarizes the results, with the details presented either in the SIs or in a separate reference publication.