Reply (*in italics*) to Reviewer 2

Review of Remsberg et al:

Given that there is only one long-term balloonborne stratospheric water vapor record from Boulder and few satellite measurements from the 1980's and 1990's, understanding and interpreting the trends present in these records is extremely important. One of the more puzzling findings that has emerged from the literature is whether or not the positive water vapor trend observed in the Boulder water vapor record is real. Several studies have highlighted that satellite measurements over the late 1990's/2000's (i.e., HALOE and SAGE II) don't seem to show the positive trend present in Boulder data. It has been suggested that the trend difference is due to some unknown drift of the Boulder record, or perhaps that the measurements from a single location are not representative of the true zonal mean. Within this context, the present work seeks a better understanding of these differences. The works shows that prior to 2002 the HALOE and Boulder records are in reasonable agreement, and provides some evidence that the change in HALOE sampling in late winter/spring after 2002 could contribute to the divergence of the trends between HALOE and Boulder when considering the longer time period through 2005. I find this idea and the presented analysis intriguing, but the work seems somewhat incomplete. I believe that more work could pretty easily be done to test the hypothesis that the trend difference is due to sampling changes in HALOE. Below, I provide some specific comments on the manuscript, and offer a few suggestions geared towards improving this analysis to make it better able to address the hypothesis it proposes and the underlying questions regarding trend differences between HALOE and FPH.

Thank you for your careful review of the manuscript.

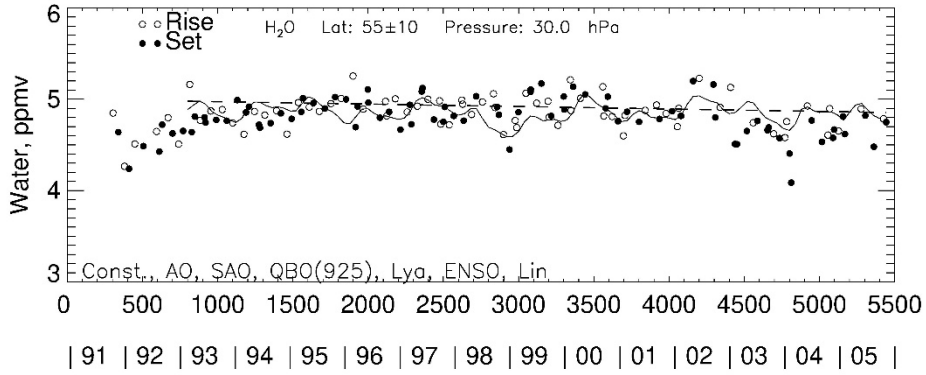
General comments:

If the continued positive trend in WV at Boulder for and end time of 2005 (as opposed to 2002) is caused by vortex air sampling, shouldn't this show up in equivalent latitude changes in Boulder? This is indeed what Scherer et al. 2008 attempted to include, and they were unable to see any change that could explain the WV trends at Boulder. You could probably look at equivalent latitude as a predictor, or just changes in equivalent latitude at Boulder, to address this. On a related note, Lossow looked pretty thoroughly at the representativeness of Boulder and concluded that there was a real discrepant trend between HALOE and FPH that couldn't be explained simply by the fact that the FPH measurements come from a specific site. At the very least, some more thorough discussion/comparison to the work by Scherer and Lossow is warranted here.

If the HALOE and FPH trend differences are explainable by sampling of vortex air at Boulder during late winter/spring, then one could pretty easily look at seasonal trend estimates to see if the agreement is better during other seasons. This can be done very easily with MLR (see, e.g., the LOTUS activity for ozone).

I do not wish to expand this Technical Note into a larger-scale study, along the lines of the work of Scherer et al. and Lossow et al., or to carry out extensive detailed MLR studies like those in the LOTUS Report. It is my conclusion that the trend differences between HALOE and FPH are not due to the specific approaches in the MLR modeling, but due to the differing natures of the HALOE (large-scale) versus FPH

(local) measurements and to the sampling of HALOE toward the end of its data record. In particular, the FPH record includes smaller-scale variations, while the HALOE measurements cannot resolve them. Regarding the reduced HALOE sampling after 2001 in Fig. 1 at the latitude of Boulder, I plan to add the following figure that shows the HALOE time series at 55±10N latitude for comparison. It shows that the HALOE data for this higher latitude zone provides samples in all seasons even after 2001 and that the water vapor values are rather high in 2002. The 55° latitude bin spans 45N to 65N, to attain a minimum of four profiles for generating each average data point within this northern hemisphere sector.



Specific comments:

Lines 44-45: This statement about the Lossow paper seems incomplete and should be more nuance, in my opinion. They showed that Boulder is representative of the zonal mean for trends over a few decades (e.g., late 1980's to 2010), but that at a decadal scale the trend differences between a single site and the zonal mean could be different.

I will make better note of their findings in the revised text.

Line 63: It would be nice to provide a brief summary of your MLR hereso that readers don't have to dig into multiple other papers, and to be as transparent as possible about what you've done. I see that the MLR is described in the next paragraph, so maybe you could just mention that it is described in detail in the next paragraph.

I will add more information about my MLR method in the revised text.

Line 69-70: You say HALOE WV values for SR events are larger than SS for 2002-2005, but the following sentence doesn't actually explain why. Please explain.

The solar lockdown position and subsequent altitude registrations for the SR versus SS measurements differ slightly and affect their retrieved water vapor profiles.

Line 80-82: You are using a 70° wide longitude range around Boulder and say that it can resolve wave-1/2 features in the SWV field. The 70° range represents 1/5th of the longitude values so I don't see how this is possible.

You are correct that a 70° range does not resolve zonal wave-2; I will change the sentence.

Line 82: Is the MLR model run with monthly means or daily data?

I did not interpolate the HALOE data to monthly means because the resulting spacing of the time series points would be altered, along with the amplitudes of the periodic terms from the MLR modeling. Typically, each data point in the series is an average over only several successive days within a month.

Line 83: What do you mean the QBO is approximated as a 28 month cycle? Are you not just using a QBO index like the Singapore winds at a certain level (or two), or the NCEP indices? Please explain.

I do not use the Singapore wind index as a QBO proxy, in part to avoid having to consider its lag at middle latitudes. Instead, initially I fit semi-annual and annual cycles to the time series of data points and analyze the residual for any remaining periodic terms. I find that there is a significant ~28-month cycle in the data in the middle stratosphere (see Remsberg, 2015), so I've used that periodic term for my final MLR modeling. However, I also note that the nominal, interannual cycle has a period closer to 30 months at 55N, 30 hPa (see figure above).

Line 86: I assume the AO is also significant? It is not listed here.

No, it is not highly significant (95%) at 40N and 30 hPa.

Line 89: I appreciate that the trend units of %/decade are useful, but can you please also give trend numbers in terms of ppmv/decade here and throughout the paper. A number of other studies (most, I thnk) have expressed water vapor trends this way, so it makes it much easier to cross-reference and compare with other studies.

I will include trends in those units.

Line 95-96: What do you mean about the MLR terms being non-orthogonal due to uneven spacing of the data? This doesn't make sense to me.

You are correct. I meant to say that the uneven spacing of the data points affects the estimated of error for each of the terms of the MLR model, although only marginally.

Line 98-99: I really think the first description of this drop was from a Randel et al 2006 paper, so that should be cited here.

I will add the reference.

Line 110: You interpolate FPH profiles, but do you apply any smoothing? I think it is appropriate to do so here, particularly if you are going to make any comparison between the noise/scatter of the FPH and that of HALOE as in Line 117. As FPH is much higher vertical resolution, it needs to be degraded to something approaching the HALOE resolution in order to make a reasonable comparison.

I do not smooth the FPH profiles, preferring to show the true nature of the FPH measurements instead. While they indicate a fair amount of scatter, I think it is likely that some of it is representative of the smaller-scale variations in atmospheric water vapor, especially where it has a significant meridional gradient and an incomplete mixing of its fields across latitudes.

Line 117: See above comment.

See my response to your comment about line 110 above.

Line 119: The SPARC 2000 report is pretty out of date at this point, and improved FPH estimates have been provided by Hall et al. Please cite this and use numbers from that paper, as they superceed anything that might have been in the old SPARC assessment.

I will delete SPARC 2000 and add your suggested reference and its numbers.

Hall, Emrys G, Allen F Jordan, Dale F Hurst, Samuel J Oltmans, Holger Vömel, Benjamin Kühnreich, and Volker Ebert. "Advancements, Measurement Uncertainties, and Recent Comparisons of the NOAA Frost Point Hygrometer." Atmospheric Measurement Techniques 9, no. 9 (September 5, 2016): 4295–4310. https://doi.org/10.5194/amt-9-4295-2016.

Line 145 (and elsewhere): These uncertainties seem extremely small to me. 1% is something like 0.05 ppmv, and the residuals (e.g. in Figure 3) are ~0.5 ppmv, so it seems unlikely that the trend values are really this precise. Has lag-1 autocorrelation of the residuals been taken into account for computing the trend significance? If the error bars are significantly larger, it would substantially change how one views the agreement (or lack thereof) between HALOE and FPH.

The slope of the linear term from the FPH time series is +5.8±1.2 %/decade, meaning it has a likely range of 4.6 to 7.0 %/decade. The range for the linear term from HALOE is 3.6 to 5.2 %/decade, which overlaps that from FPH. The corresponding term coefficients are uncertain by 21% (FPH) and by 18% (HALOE). On the other hand, the amplitudes of the periodic terms are small compared to their uncertainties from the MLR analyses of both data sets. The MLR modeling herein corrects for lag-1 effects, although the AR-1

coefficient is small (almost no memory between adjacent points) for the FPH time series because of the scatter of its points.

Line 162: What does “qualitatively accurate” mean here?

A poor choice of words, I agree. I will say ‘qualitative correct” based on the decay of the volcanic aerosol extinction from the HALOE data versus that of the aerosol extinction estimates from balloon-borne particle measurements over Laramie, WY (Hervig et al., 1996, Plate 1).

Figures:

A number of figures are really zoomed out on the y-axis, making it hard to see the data. E.g., in Figure 1 the data range goes from ~4.5-5 ppm but the y axis goes from 3 to 6 ppmv. Please zoom in the axes here and elsewhere.

The data range on the y-axis is intentional, so that one can easily see the differing natures of the HALOE versus the FPH time series data.

A number of the figures would make more sense if they were combined. E.g., Figures 1-2 together, 4-5, 9-10.

I will combine several of the figures into single figures for ease of comparison.