This manuscript shows some very nice measurements of temperatures in the stratosphere. The measurements compare well with various other sources. While I have several suggestions, I have no major recommendations for changes. My most serious complaint is that, while many of the authors of this manuscript speak excellent English, much of the text is very poorly written. I certainly do not think that it should be the reviewer's role to assist in this task, especially when several of the co-authors are completely capable of doing so.

Below are a series of specific suggested changes:

Page 5 - "Radiosondes reach an altitude of 35km". This, and later statements, gives the impression that all radiosondes from Payerne reach precisely this altitude, but I am skeptical that this is the case.

Figure 3 – Given that, according for Figure 4d, the measurement response falls to well below 50% at ~17km, and that, as far as I have been able to determine, other TEMPERA studies show sensitivity only up to at best ~7km, suggesting that TEMPERA measures into the upper troposphere is very deceptive. It does certainly not, as the text suggests "cover the full troposphere and stratosphere".

Figure 4b – This Figure is a simplification of a very similar Figure 14 in Stähli et al. (2013). In that figure it is stated that "In the center of the lines we use all channels and on the wings of the line we use a binning of 3 channels for data reduction". I am almost certain that this is why the middle channels are noisier, and that this has nothing to do with the Zeeman effect, as is stated in the paper.

Figure 4c – What is meant by "observation error"? Given that there are systematic changes >2K in the dataset, I assume that this must be some kind of random error estimate. If this is the case please label it as such. How is it calculated?

Figure 6 – Given the very large discrepancy between the MLS and other measurements on 4 February above 35km, this clearly warrants some discussion. It is very troubling that neither the LIDAR nor the MWR show the decrease in temperature above 35km. Do nearby (in time and space) MLS profiles show the same structure? Do the authors think that this is a bad MLS profile?

Page 9 – "The measurements presented in the plot show the importance of continuous observations for a fixed location, since the variability of atmospheric parameters such as temperature evinces the necessity of measurements with good temporal resolution." This suggests that there are temperature variations every few hours (a conclusion that can certainly not be reached by looking at Figure 5). If this is the case, please show such. If not, then certainly daily satellite measurements must be adequate, and this statement should be removed.

Page 11 – It is stated that, above 35km the RS profiles were extrapolated using TEMPERA profiles. But the top altitude of RS profiles varies, so exactly what does this mean? Were only RS profiles which reached 35km included. If there was data above this was it included (instead of using the TEMPERA data)?

Figures 7, 9, and 12 – The most piece of information on these is the comparison between TEMPERA and the convolved retrievals from the other instruments. Since this is the case, it would be best to plot the

interpolated line first, and then the convolved and TEMPERA lines on top of this. As the plots are currently shown it is sometimes difficult to tell whether the TEMPERA line coincides with the convolved or the interpolated measurement. Alternatively, the interpolated measurement could even be dropped from these plots.

Figures 7 and 9 – is the black deviation line TEMPERA vs. RS interp or RS conv? In fact, throughout much of the text it is not clear whether convolved or unconvolved RS and/or MLS data is being used.

The step in the data shown in Figure 7 very helpful in that it provides a useful measure of the systematic errors in these retrievals. I applaud the investigators for not homogenizing the data between the two periods.

Page 13 – "near time-coincident". Does this mean that the MLS profile was taken during the period of spectral integration for the TEMPERA measurement? If so, please state this. If not, please state what exactly "near time coincident" means. The same applies to the RS comparisons.

Page 15 and Figure 11 – The authors note that: "the bias and the standard deviation observed between MLS and RS is very similar to the values observed in the comparison between TEMPERA and RS in period 2." If do not know, and it seems to be nowhere stated, whether in Figures 8 and 10, the MLS and RS measurements are convolved before comparison with the MWR. If they are (and I think they should be), then the appropriate comparison in Figure 11 would be convolved MLS with convolved RS profiles. This could be added as a dashed line in Figure 11.

Figure 17 – The legend is a bit confusing. Please make the 4 instrument lines solid and thick enough so that one can distinguish lidar and WACCM. Then separately show two styles of lines, one for Period 1 and one for Period 2.

Table 1 – Since Period 1 and Period 2 are presented everywhere else, why is only Period 2 in this table?