Comments on "The SPARC water vapour assessment II: Biases and drifts of water vapour satellite data records with respect to frost point dygrometer records" by Kiefer et al.

The long-term changes in the stratospheric water vapour and its climate impacts have long been one of the key science questions for the climate community. Validation of satellite data is fundamental for it. The authors provide a comprehensive and careful comparison of all available satellite data records from 2000 to 2016 to FP data. These results would be very useful for future studies on stratospheric water vapour using satellite measurements. I am in favor of publishing the paper after the following points have been considered.

Specific comments:

Line 26: "Ding-Zhu et al. 2015" in text and its reference are in incorrect format, which should be changed to "Hu et al., 2015"

Figure 1: The legend is confusing. It seems only RVM has two locations due to the ship cruise. In this case, it is better to only use one marker for each station except two for RVM in the legend.

Line 154: "Therefore we have decided to minimize the contribution of natural variability using this method…" What is the natural variability here? Is it the internal variability of water vapour itself? Why using the closest satellite profile can minimize the contribution of natural variability?

Line 202: "Comparison of the two profiles, with the FP data simply interpolated to the coarser grid of the satellite instrument, but not smoothed (black diamonds), is misleading since the MIPAS instrument is unable to solve the sharp feature in the profile." I understand the reason, but the "truth" is compromised due to MIPAS current capability.

Figure 4: The grey and blue lines and shadings overlapped with each other, making it not easy to distinguish between the two.

Line 273: "Below 100 hPa, close to the tropopause it has a sharp peak of -25%." Why MLS has a negative bias, instead of positive, relative to the FP when it cannot resolve the sharp water vapour decrease (as suggested by Figure 3)?

Figure 7: It may be worth to put the symbols for FP stations on the plot to save the readers' effort to go back to Figure 1, and likewise for Figure 5 and Figure 6.

Line 321: Add a space between "within" and "±".

Figure A1-A3: These plots in Appendix seem to be essential to the paper.

Figure 8: Some of the bars exceed the x-axis limit.

Figure 11 and Figure 12 are somewhat in contradiction. Figure 11 suggests that "MLS-FP time series show little or no evidence of drift until ~2010" at 68 hPa, but Figure 12 is showing significant positive drifts at this level for the full records from 2004 to 2016. It would be helpful to provide the profiles of drifts for each of the two periods.

Line 899: "Four SATs had mean drifts with magnitudes $> 2\%$ yr^{-1} in one pressure interval ..."
It is more straightforward to just name them here.