This is a very important paper. A common methodology in stratospheric research involves treating the results of reanalyses products as being analogous to atmospheric data from measurements. There are several reasons for this. One being the relative paucity of stratospheric measuerments, and another being the ease of use of reanalyses (i. e., no missing data, evenly spaced information). One need to examine the conclusions of papers such as Randel et al. (2004) and later works to see both the advantages and shortcomings of reanalyses representation of measurement data. This paper by Wright and Hindley does a very careful mapping of reanalyses temperatures from several modern reanalyses (CFSR, ERA-5, ERA-Interim, JRA-55, JRA-55C, and MERRA-2) onto the HIRDLS, SABER, and AIRS satellite instrument and the COSMIC weighted measurement volumes for comparison with HIRDLS, SABER, AND COSMIC retrieved temperatures, as well as AIRS radiances at various altitudes in the stratosphere and mesosphere. Comparisons are done using global data by means of scatter plots and the resulting correlations, and also by comparing time series for equatorial and high latitude zonal means at different altitudes. In general, the principal conclusions are that the reanalyses closely reproduce the measurements at 30 km (correlations in excess of 0.98) except for JRA-55C with correlations about 0.97, despite the fact that JRA-55C assimilates no stratospheric data. The correlations fall off substantially at 50 km, with the lowest correlations being for JRA-55C, followed by CFSR. Interestingly, the comparisons with COSMIC give lower correlations even though COSMIC stratospheric measurements are included in the assimilation process.

Examining time series comparisons shows several things. SABER displays a high bias of 1-2 ^{0}K at 30 km in equatorial latitudes, and HIRDLS shows a similar low bias at 50 km near the Equator. All of the reanalyses show less fidelity to the observations at 50 km than at 30 km, but this is particularly true for the CFSR, with lesser fidelity to the observations during disturbed periods.

An interesting result shown in the Taylor diagrams is that all of the reanalyses are more similar to each other than to the observations. A cluster analysis is carried out showing that the reanalyses are more correlated with each other than with any of the observations (including COSMIC, whose data is assimilated in all the reanalyses). The suggestion here is that the reanalyses may be being “tuned” to each other excessively. The HIRDLS, COSMIC, and SABER temperatures are noticeably less similar to each other than are the reanalyses temperatures to the other reanalyses. The COSMIC temperatures are particularly dissimilar to the HRDLS and SABER temperatures. The authors indicate that this “suggests either that the COSMIC temperature retrieval introduces large additional errors relative to the assimilated form, or that the relative importance of COSMIC data in the reanalyses schemes used is too low.”

As is apparent in the previous discussion, I liked this paper a lot, and I think it brings out some very important points for users of stratospheric reanalyses products in their research. Nevertheless, I do have some critical comments, which follow.

1. On page 1, line 5, COSMIC is described as an instrument. It is not. See page 3, lines 6 and 7.
2. On page 1, line 7, I suggest "use cases" be deleted, in favor of the word "usage."
3. On page 1, reference is made several times to "full input reanalyses" without definition.
4. On page 3, on line 2 the acronym COSMIC is defined incorrectly. This is done correctly lower down on line 28.
5. On page 3, line 20, "they" should be "this."
6. Page 4, line 1. Aren't stratospheric temperatures always "dry?"
7. I found the description of figure (4b) to be confusing. POerhaps, the authors might revise the text in this regard.
8. I think the paper would benefit from a more "broad-brush" description of the methodology in the main text with the details being in either an appendix or a supplement. While it is important to describe their procedures, I tended to get bogged down in the detailed procedure descriptions. This detailed description should not detract from the paper's important conclusions.
9. Page 23, lines 11-15. I find this statement confusing. The authors are attributing summer pole problems to the cold-pole problem. Isn't this a winter pole problem? Aren't the authors referring to the need for more gravity wave drag around 60 ^{0}S? Certainly, this was the concern of the cited papers.
10. Would the authors say a few words describing the stand-alone black dots in figures 16 and 17?
11. Page 17, line 26, satellite should be singular.
12. Page 29, lines 12-14, is this due to the large vertical extent of the AIRS weighting function, its fine horizontal resolution, or both?
13. Page 29, line 20. The zonal mean results are very interesting. I think it might also be interesting to compare results for various wavenumbers since planetary wave diagnostics are quite common in stratospheric research.

References

Randel et al., 2004: The SPARC Intercomparison of Middle-Atmosphere Climatologies. *J. Climate*, *doi.org/10.1175/15200442(2004)017<0986:TSIOMC>2.0.CO;2*.