June 17, 2024

Reply to the Revised version of Liu et al:

I thank the authors for their detailed replies to my original review. I have several comments on their replies to my original review, given below. I'll refer (in bold font) to the page number of the pdf document containing their replies in my comments below, to help guide the authors as to which comment of theirs I am addressing.

I sincerely appreciate the passion of the authors and the efforts they have made to justify their results. But I also want to emphasize to the authors that the usefulness of the science is not in the value of a measured parameter, but rather in the ***demonstrable uncertainty*** of a measured parameter.

Recommendation:

I am rejecting the paper again and returning for major revision. As outlined below, the Monte Carlo analysis is incorrect for systematic errors in general and in particular for the nature of the uncertainty in key parameters (atomic oxygen and carbon dioxide, and also the nature of the non-LTE radiative transfer) that are used in the SABER temperature retrieval. These errors cannot be reduced by averaging as they are systematic and not random or quasi-random in nature.

I will give one example here of what I am referring to, and it is discussed below as well. The MSIS 2000 atomic oxygen has no long-term trend or dynamical variability component in it. The model is empirical. One specifies, date, local time, location, F10.7, and Ap to get a profile of O that is used in the SABER temperature retrieval above about 87 km. One would expect to get the same O on Jan 1 2003 and Jan 1 2023 if all the remaining parameters entered in the call to the model are the same. So, from the outset, ***one would expect incorrect trends*** above 95 km where O uncertainty drives the error budget.

In addition, Mlynczak et al., 2023 showed that 15% uncertainty in CO_2 at 110 km led to an 8 K error in global mean temperature. The uncertainty in polar regions to CO_2 accuracy is likely much higher due to the non-LTE nature of the radiative transfer. Furthermore, the 15% uncertainty came from CO_2 used in two different versions of the WACCM model (i.e., different Prandtl numbers). So the uncertainty in model projections of the future CO_2, particularly in the polar regions, are significant sources of uncertainty in SABER temperatures at high altitudes as well. This is both a trend uncertainty and a dynamical uncertainty. From the outset, ***one would expect incorrect trends*** above 95 km where CO_2 also drives the temperature error budget. (https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL102398)

It is more than likely that what is appearing in the SABER data at these high altitudes is the consequence of the lack of correct trends in O and CO_2 coupled with the lack of any dynamical adjustment in the O and likely only partially correct dynamical adjustment in CO_2. The authors must understand that these are systematic errors and not random or quasi random. The values are simply incorrect by an unknown amount. They cannot be reduced by averaging.

For a paper to be acceptable for publication, the authors must remove the incorrect Monte Carlo analysis. It is not valid even as an Appendix. They must confront the stated errors in the SABER data at these ***altitudes and include error assessments of the derived trends*** based on the systematic nature of the errors.

The authors may ask why this has never been done in prior papers. Frankly, only in the last couple of years have we come to an understanding of the intricacies of the algorithms and the sensitivity of the temperatures to the algorithms (See Mlynczak et al., GRL, 2023, on Algorithm stability cited above) with regards to trends. The instrument, mission, and algorithms were designed over 25 years ago for a 2-year mission to examine the annual variability of the mesosphere. The community is now using the data for applications that were never intended, and, as pointed out in my comments below, have multiple facets that can cause large uncertainties in the data, particularly trends. I would write some of my own papers differently if doing so today.

I am encouraging the authors not to be discouraged. A revised paper rigorously discussing the observed 'trends' but then with a rigorous analysis of the errors, which will almost certainly show the large "trends" to be insignificant at the known uncertainty, will be much more useful for the community and for development of future missions than a paper showing a trend that cannot be justified. Such analyses are badly needed for the future of the 'space climate' field and for the development of new measurement techniques.

Page 10. Reply (1) "For the unexpected trends values.."

This point made by the authors about dynamical feedback might be the case, but might it also be just natural variability of the system? How reliable are the modeled Beig values from 2003 and 2011? How would you know the difference? It seems that the authors are attempting to models to justify very uncertain observations.

Page 10, Reply (2), "The measurement uncertainties of the SABER temperature…"

The authors have done an interesting Monte Carlo simulation. However, the simulation essentially is one of random errors and not systematic errors. Random and quasi-random error can be reduced with averaging many data points. Systematic errors cannot be reduced by averaging and their distribution is unknown. You can average all you want but the final product is still going to have a systematic error of a single profile.

This is especially true of the SABER temperature dataset at 1e-04 hPa which depends so strongly on atomic oxygen and carbon dioxide. As I noted in my original review, there is no reason to expect that the trend in atomic oxygen is correct in the MSIS 2000 data that is used in the SABER temperature retrieval. For a given day of year, latitude, local time, and Ap index, it will return the same atomic oxygen in 2024 as it did in 2004. That is how MSIS works. Further, MSIS is totally dependent on the input data used to 'train' it. It does not have correct local time and possibly even correct seasonal or annual variations.

If dynamical processes are thought to be contributing to the observed temperature changes, then there is even further reason to doubt the MSIS 2000 atomic oxygen being correct over 22 years. Furthermore, for CO2, the trend must also be correct in the SABER algorithms, and for that to be the case, the monthly WACCM values of CO2 must be correctly replicating the dynamical changes in the atmosphere, as well as any natural time changing dynamics of the polar summer mesosphere region. That is, the WACCM simulation used in the SABER retrieval has to accurately simulate both the dynamical changes and the trend in CO2 correctly. There is no evidence that can be given to support such a requirement. And as noted in Mlynczak et al, 2023, different versions of WACCM have significantly different values of CO2, **it is expected to have large, systematic uncertainties in temperature**.

Neither the MSIS behavior nor the WACCM behavior can be justified as accurately replicating the time dependent changes in the polar atmosphere. Consequently, the SABER temperature systematic errors are large and cannot be reduced by averaging. It is much more likely that what SABER is observing are effects due to incorrect trends and/or values of key inputs to the temperature algorithm.

There is an even more subtle effect that I'd say is 'secondary' but would still be essential to demonstrate an observed trend, especially at higher altitudes and latitudes. The non-local thermodynamic equilibrium (non-LTE) radiative transfer in CO2 couples the vibrational temperatures at all altitudes due to exchange of radiation between all layers. If there is an error in O or CO2 at one altitude, it affects the temperature at all altitudes to some degree. I mentioned in my previous review that between 80 and 100 km, there were no observations of atomic oxygen used to train MSIS 2000. So there is no reason to expect the trends in O to be correct between 80 and 100 km. Due to this, the temperatures **above** 100 km also have uncertainty due to uncertainties in parameters **below** 100 km. This is why it is called "non-local" thermodynamic equilibrium.

The authors state that the Monte Carlo analysis is included in the Appendix. This discussion is not correct with respect to the nature of SABER errors and should not be included anywhere in the revised paper, not even an Appendix, given all the reasons I have presented above.

I admire the authors' tenacity here, but the nature of the data and how the algorithms work simply does not admit the value of these trends within the known SABER measurement uncertainties.

In summary for this section, the simplest way to think of the role of uncertainty here is that you are differencing two numbers taken 20 years apart and each has a systematic uncertainty of 10 K to 25 K? What is the uncertainty of the difference?

Page 11 Recommendation:

The authors appear to have re-arranged some sections, but still retain the Appendix. As I note above, the Appendix is incorrect with regards to SABER data uncertainties. The authors need to remove this and any reference to these analyses throughout the paper.

Page 12, the authors' "Response", blue text, that extends to the top of page 13.

The authors seem to have misunderstood my point about 'climate sensitivity'. Or maybe I am mis-understanding their reply. The point is not when CO2 doubles at the surface, but when CO2 doubles in the MLT, the temperature should decrease 6 to 8 K from pre-industrial times. This will happen over a century or more. The authors are reporting trends of this magnitude in a decade when CO2 will have increased 5% to 6%. And from what I can tell in their reply, they are still not considering any issues related to the uncertainty of the data even at 1e-03 hPa.

I do not see any reply to my original comment that addresses the effects of uncertainty in the data on trends. The authors would have to justify that the large systematic errors all cancel out, which as we discussed above from the basics of the SABER algorithm, cannot be justified.

In the next version of this paper, I expect to see trends and conservative trend uncertainty estimates accompanying any trend value, as well as the statistical significance of the uncertainty. Note that all SABER uncertainties are 1-sigma values. So a 25 K uncertainty at 110 km 1-sigma corresponds to a 2-sigma uncertainty (95% confidence) of 50 K at that altitude.

Page 14 – Page 18, all of the blue text in the pdf document with the author's replies.

Items numbered (1) and (2) in this section indicate the reality of the SABER data uncertainties but do not address the implications of these uncertainties for the results presented in the paper.

Item (3) largely addresses the Monte Carlo analyses discussed by the authors in earlier replies. As noted, this analysis does not address the nature of the systematic errors in SABER data and the near-certain lack of any correct trends in key parameters (O, CO2) employed in the SABER temperature retrieval algorithms.

Page 18-19, my question regarding de-trending of the QBO parameters.

Note that the temperature trend analysis of Garcia et al, (2019) did detrend the QBO parameters. Perhaps a more important question is why include QBO as a predictor in the first place? Is it demonstrated to play a role in the temperature trends of the polar mesosphere? In both Garcia et al., (2019) and Mlynczak et al., (2022) the QBO was not a significant predictor.

Marty Mlynczak
NASA Langley Research Center