I'm afraid I don't think the author's responses and changes they made to the manuscript adequately address the concerns raised by both reviewers. Thus, I don't think this article should be published unless major changes are made.

The main conclusion of the article, that Hg(II) concentrations in the stratosphere are low, is driven by a measurement method that is very likely to have biased the Hg(II) measurements low. The authors mention in the article that GEM concentrations were often higher than TM concentrations, but they don't acknowledge the most likely reason for this—that the GEM (and thus Hg(II)) measurement method is biased. Evidence exists in the literature and from unpublished work that quartz wool traps, at the very best, can only be used for very short-term measurements, and that they off-gas captured Hg into the sample line (1) over the passage of time and (2) when humidity increases. The authors did not explain this in the article and they did not account for this known bias in the conclusions they draw from their measurements.

I still think this paper can be published. I think the authors need to make two general changes:

1. They need to provide more information about the quartz wool method they used and, most importantly, they need to talk openly in the text about the fact that it is well known that quartz wool traps suffer from bias under at least some conditions (possibly all conditions).
2. They need to tone down and qualify the discussion and conclusions that stem from the quartz wool trap-based measurements. The paper should be shorter and more focused on the TM results, which seem robust. The conclusions that involve Hg speciation should be stated in light of the high uncertainty of the measurements, and many of them probably need to be removed.

I think the major changes mentioned above are the most important, but I have a few specific comments below. Line numbers are from the revised manuscript that has changes tracked.

- Lines 194-196: I don't think this is true or accurate.
    - First, the implicit assumption is that the Lyman and Jaffe data are of high quality. In light of what we now know about quartz wool and its ability to rapidly re-release Hg, I think we should view the data of Lyman and Jaffe with skepticism.
    - Second, I don't think it is true that the quartz wool scrubbers were operated according to the method of Lyman and Jaffe. I view the results of Lyman and Jaffe with skepticism, and several key differences between the Lyman and Jaffe method and the method used in the current study cause me to view the results presented in the current study with even more skepticism:
        - Lyman and Jaffe switched out quartz wool scrubbers every ~2 hours and as needed to minimize release of accumulated Hg from the scrubbers, but in the current study the same scrubbers were used for long periods of time.
        - Lyman and Jaffe performed multiple zero air checks of their measurement system on each flight to ensure that loss of Hg from quartz wool scrubbers was minimized, but this was not done in the current study.
        - Water vapor has been shown to facilitate the loss of Hg from quartz wool scrubbers. Lyman and Jaffe showed that no change in water vapor concentrations occurred as they traveled in and out of air influenced by the UT/LS, making it unlikely that the patterns they observed were due to the

effects of humidity. If high ozone in the UT/LS-influenced air they observed drove Hg from the quartz wool traps, it would have dampened the measured Hg(II) signal, rather than enhanced it—so the Hg(II) measured by Lyman and Jaffe in UT/LS-influenced air is likely a lower-bound. In the current study, not attempt has been made to determine periods when the quartz wool traps may have been operating poorly.

- Lines 196-202: If GEM was greater than TM, this was almost certainly because Hg accumulated on quartz wool scrubbers was being released. We can expect that some Hg was being released from the quartz wool during periods when GEM was lower than TM. One way the authors could deal with this is to look at only the first 1-4 hours or so of data collected after a new quartz wool trap was installed. Comparing quartz wool-derived Hg(II) data over different intervals after installation of a new trap would provide information about the level of bias in the results.
- Lines 426-429: Brooks et al. used KCl denuders, which are also known to be biased low. I think this should be pointed out. Also, Brooks et al only went to 6 km, far below the UT/LS, so I don't think those data can be used to tell much about Hg(II) in the UT/LS.
- Lines 431-432: Gratz et al. did not target UT/LS air either.
- Many ground-based studies show high Hg(II) in air masses influenced by UT/LS—it seems the authors need to include a discussion of some of that work as well.
- Lines 432-435: I think it could be true that the Lyman and Jaffe results are an "event phenomenon." However, I don't think the current study is adequate to show that this is the case.