I enjoyed reading this paper and I congratulate the authors on a very thorough investigation of mercury in the UT/LS. It is a shame that CARIBIC mercury measurements are not happening right now. I think this paper needs to be published and it is a useful and important contribution to scientific understanding of atmospheric mercury distribution and chemistry. I think a few substantial changes need to be made to the manuscript before it can be accepted for publication. I point these out below along with more minor issues.

- The description of atmospheric mercury cycling on lines 63-70 is misleading, I think. It could be improved by pointing out that
    - Elemental Hg (not just Hg2+) can deposit directly to surfaces,
    - Hg2+, not just elemental Hg, is emitted from sources
    - Deposition of elemental Hg and Hg2+ can occur near sources, not just in remote locales.

- I think some of the evidences for the conclusions drawn by the paper are perhaps too detailed, and that some of the discussion could be moved to supporting information. I feel like this would improve the paper's readability. But this is a minor thing and I think it is up to the discretion of the authors. I just offer it as a suggestion.

- My biggest concern with the paper is with the definitive conclusions drawn from the GEM data, which were collected using quartz wool traps to remove Hg2+. I hope the authors will forgive me for discussing this at some length, and for coming to conclusions that are contrary to theirs:
    - The authors state that Ambrose et al. declare quartz wool traps do not suffer from bias in the presence of ozone. My own experiments show that Hg2+ compounds are, in fact, converted to elemental Hg on quartz wool traps in the presence of ozone. The loss in the presence of ozone is less than that exhibited by KCL denuders, but it is still there. These results are presented in the supplemental information for Lyman and Jaffe (2012).
    - Quartz wool unfortunately only holds Hg2+ compounds temporarily. The longer the flow time through the tube, the more Hg will be released. Also, I have seen that a relatively short pulse of high humidity, such as one might encounter during the beginning of CARIBIC flights when the aircraft is at low altitude, can lead to long periods of bleed-off of elemental Hg. This is likely what led to GEM numbers that were higher than TM numbers as reported in the manuscript. The fact that GEM is lower than TM doesn't necessarily mean that bleed-off of elemental Hg is not happening. It just means it is low enough that it isn't higher than the TM value. During the flights described by Lyman and Jaffe (2012), we switched among quartz wool traps every 2-4 hours so we always had relatively fresh traps. We also pumped zero air into the manifold every few hours to make sure we weren't having bleed-off of elemental Hg from the quartz wool traps. In spite of this, our flights still had problems at low altitudes in humid air. The strat intrusion data used by Lyman and Jaffe (2012) were for a new trap in low-humidity air, and the humidity didn't vary between the strat-influenced air and trop-influenced air, so that gave us confidence in our conclusions.
    - I think the evidence we have about quartz wool traps indicates that the way they were used in CARIBIC flights is likely to lead to bias. Specifically, it would likely lead to conversion of trapped Hg2+ to elemental Hg, followed by bleed-off of elemental Hg,

which would cause a low bias in the GEM measurement. The extent and timing of this bias on a given flight is probably unknowable and uncorrectable.
    - I'm not insisting the authors remove all discussion of their GEM measurements. But I do think the conclusions made based on the GEM data need to be watered down and qualified in terms of the very real and very large uncertainty that comes from using quartz wool traps the way they were used on the CARIBIC flights. I also think the manuscript needs to discuss the known and possible biases in quartz wool traps in detail and clearly acknowledge the uncertainties they present.

- I would have loved to see some detailed QC tests of the measurement setup. I think that without these, some of the conclusions drawn in the paper (including the ones that rely on quartz wool traps) are less than convincing, and I would like to see the authors qualify those conclusions better in the paper. Even better, I would like to see them conduct some on-the ground experiments to test the measurement setup.
    - To me, the evidence offered in Slemr et al. (2016) that the CARIBIC sampling line quantitatively transmits Hg^{2+} is circumstantial and not conclusive. It would be easy enough to permeate $HgBr_2$ or something into the line under simulated conditions on the ground and measure the difference in Hg^{2+} concentration at the beginning and end of the line. This would provide the evidence needed to make a fairly firm statement about Hg^{2+} transmission.
    - Particulate cut point tests are more difficult, but it is surprising to me that there are no experimental results available from CARIBIC about the cut point of the trace gas inlet used in this study. This is definitely needed. It seems premature to make strong conclusions about TM/GEM ratios based on an inlet with an untested cutpoint and aspiration efficiency.
    - Quartz wool tests on the ground in simulated flight conditions would be similarly useful. If the authors feel their GEM measurements were high quality and unbiased, they need to provide strong evidence that this is the case, especially since other studies (supporting information in Lyman and Jaffe, 2012 and Ambrose et al., 2015) show that quartz wool traps only work under very specific conditions. Many of these tests would be relatively straightforward to put together.

- The strength of the CARIBIC measurements is in their quantity and duration, not in the speciation data collected, and I recommend the authors refocus the paper on the long term trends in TM and de-emphasize the speciation measurements and conclusions drawn therefrom.

- I am concerned that the authors have misinterpreted some of the information presented by Lyman and Jaffe (2012):
    - The authors state that the empirical model developed by Lyman and Jaffe predicts a total depletion of Hg in the stratosphere, and they criticize the Lyman and Jaffe article extensively because this is demonstrably untrue. In reality, Lyman and Jaffe were careful to point out that they understood that a total depletion of Hg in the stratosphere is unrealistic. I quote the article here: "Whereas the model shows a total depletion of

mercury about 5 km above the tropopause, previous measurements show some mercury even 8 km above the tropopause. This discrepancy probably exists because the linear relationships used in the model become less applicable deep in the stratosphere (note that the ozonesonde data used in the model exhibited higher ozone mixing ratios at the thermal tropopause than is typical; see Supplementary Information)."

  - The authors also state that the conditions encountered by Lyman and Jaffe were not likely representative of the UT/LS as a whole, with the implication that Lyman and Jaffe didn't point this out. In fact, this was pointed out in the above quote, and also here: "Although it is unlikely that this model is quantitatively representative of the stratosphere as a whole, it does exhibit the qualitative features of mercury distribution in the stratosphere that have been demonstrated in this and other studies."
  - I am glad the authors pointed out the error of one of Lyman and Jaffe's main conclusions, namely that Hg in the stratosphere has a relatively short lifetime. It was a gross oversight on Lyman's and Jaffe's parts not to calculate a stratospheric lifetime and to not note the long lifetime of sulfate particles in the stratosphere.

- The authors use data from Brooks et al. (2014) to demonstrate that GOM is relatively low at altitude. The authors should point out that those measurements were collected with KCl denuders, which only retain about 1/3 of total gaseous Hg2+ in a typical humid atmosphere.

- I disagree with the first two sentences of the "conclusions and outlook" section.
  - The authors state that "The obvious implication of the long stratospheric TM and GEM lifetimes is that most atmospheric mercury is oxidized in the troposphere."
    - I don't think this is obvious at all. The long TM lifetime in the stratosphere is likely caused by a lack of Hg loss processes. There are few clouds and no surfaces for Hg to deposit to, so it has no choice but to stick around for a long time.
  - The authors also state that "The second direct implication is that if the lifetime of GEM in the stratosphere with its very high O3 mixing ratios (typically 1 ppm and more) is quite long, then the GEM + O3 reaction cannot be important in the troposphere with its low O3 mixing ratios."
    - I don't agree with this conclusion either. Even if most of the Hg in the stratosphere is elemental, which I don't think this study demonstrates in a conclusive way, it could be thus because Hg2+ reduces back to elemental Hg after it deposits on particles, keeping Hg2+ concentrations from building up.