

## Reply to Seth Lyman

We thank Seth Lyman for his constructive review and reply as follows:

1. *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 Hg<sup>2+</sup>) can deposit directly to surfaces,*
  - *Hg<sup>2+</sup>, not just elemental Hg, is emitted from sources*
  - *Deposition of elemental Hg and Hg<sup>2+</sup> can occur near sources, not just in remote locales.*

The purpose of the first paragraph of the introduction is to provide a broad picture in a few words. Details can be found in the given references. We modified the text by pointing out that mercury deposition is not confined to remote areas.

2. *My biggest concern with the paper is with definitive conclusions drawn from the GEM data, which were collected using quartz wool traps to remove Hg<sup>2+</sup>. I hope the authors will forgive me for discussing this at some length, and for coming to conclusions that are contrary to theirs:*
  - a) *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 Hg<sup>2+</sup> 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 exhibited by KCl denuders, but is still there. These results are presented in the supplemental information for Lyman and Jaffe (2012).*
  - b) *Quartz wool unfortunately only holds Hg<sup>2+</sup> 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 the 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 the 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 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.*
  - c) *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 Hg<sup>2+</sup> 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.*
  - d) *I'm not insisting the authors remove all discussion of their GEM measurements. But I do think the conclusions made based on 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.*

The discussion on the efficiency and interferences of different sorts of traps and denuders goes on for more than a decade without providing any firm conclusion on a method of separation of GEM from  $\text{Hg}^{2+}$  which is accurate and reliable (Gustin et al., 2015). Consequently, all  $\text{Hg}^{2+}$  data presented so far have a potential for bias. This is now mentioned in the text.

Quartz wool traps were installed in the CARIBIC instrument in August 2014 and their deployment was based on advice from Jaffe's group including the information presented by Lyman and Jaffe (2012). The reviewer states that the way they were used in CARIBIC flights is likely to lead to bias.

In paragraphs 2a and 2c he specifically mentions a conversion of  $\text{Hg}^{2+}$  to GEM on quartz wool and refers to supplement information of Lyman and Jaffe (2012). In it we could find only information on  $\text{Hg}^{2+}$  losses, not on reduction of  $\text{Hg}^{2+}$  to GEM. We think that in highly oxidizing and dry stratospheric environment with more than 100 ppb of  $\text{O}_3$  and less than 100 ppm  $\text{H}_2\text{O}$  a reduction of  $\text{Hg}^{2+}$  to GEM is unlikely. The authors mention 6 – 20% loss of  $\text{Hg}^{2+}$  from quartz wool traps at 30 ppb  $\text{O}_3$  after 30 min pumping, results from a comparable experiment without ozone are not presented. If ozone were the sole cause for mercury bleed-off then the bleed-off should increase with increasing  $\text{O}_3$  mixing ratio. This is, however, inconsistent with  $\sim 500 \text{ pg m}^{-3}$  peak of  $\text{Hg}^{2+}$  at  $\sim 250 \text{ ppb O}_3$  presented in Fig.1 of Lyman and Jaffe (2012) and, if true, would invalidate the conclusions of their paper.

Paragraph 2b mentions problems with humidity at low flight altitudes, bleed-off of captured  $\text{Hg}^{2+}$  and limited capacity of quartz wool traps for retaining  $\text{Hg}^{2+}$ , i.e.  $\text{Hg}^{2+}$  breakthrough after hours of sampling. To prevent contamination of the sampling manifold and the instruments in the vicinity of airports, the CARIBIC sampling pumps and instruments are switched on when the pressure sinks below 500 hPa (Slemr et al., 2016), i.e. mercury was sampled only at altitudes  $> \sim 5 \text{ km}$ . This is now mentioned in the text. At stratospheric water mixing ratios of less than 100 ppm, humidity should not be a problem. Quartz wool traps were used for 2-4 h at a sampling rate of  $1 \text{ l(STP) min}^{-1}$  by Lyman and Jaffe (2012). The CARIBIC sampling flow rate was set to  $0.5 \text{ l(STP) min}^{-1}$ , but because of the limited pumping speed and low inlet pressure at cruising altitude of mostly  $> 10 \text{ km}$ , the effective sampling rate was on average less than  $0.25 \text{ l(STP) min}^{-1}$ . The two intercontinental flights took typically 20 h. The total sampled volume would then correspond to about 5 h sampling duration at a sampling rate used by Lyman and Jaffe (2012), only slightly exceeding their maximum sampling volume collected during 4 h. The reviewer also thinks that the contamination leading to higher GEM than TM values was due to humidity induced bleed-off. In fact, the contamination started to appear after the change of the personnel operating the instrument and we thus attribute the contamination to it.

In summary, the quartz wool traps onboard CARIBIC were used nearly in the same way as described by Lyman and Jaffe (2012) and our results, if biased, should have the same bias as theirs.

3. *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.*

a) *To me, the evidence offered in Slemr et al. (2016) that the CARIBIC sampling line quantitatively transmits  $\text{Hg}^{2+}$  is circumstantial and not conclusive. It would be easy enough to permeate  $\text{HgBr}_2$  or something into the line under simulated conditions on the*

*ground and measure the difference in Hg<sup>2+</sup> concentration at the beginning and end of the line. This would provide the evidence needed to make a fairly firm statement about Hg<sup>2+</sup> transmission.*

- b) *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 cut point and aspiration efficiency.*
  
- c) *Quartz wool test 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 QC test made with the CARIBIC instrument are described in detail by Slemr et al. (2016).

In paragraph 3a the reviewer states that the evidence for quantitative transmission of Hg<sup>2+</sup> through the sampling line is circumstantial and not conclusive. We agree with him about the circumstantial evidence but see no easy experimental way to arrive at conclusive evidence. The major problem is the on-ground simulation of the sampling conditions at the cruise level, i.e. ca -50°C, 200 hPa, less than 100 ppm water, and high flow rate through the manifold induced by the ram pressure at the cruising speed of some 900 km h<sup>-1</sup>. The permeation test would thus make sense only if HgBr<sub>2</sub> were added in flight at the inlet described in detail by Brenninkmeijer et al. (2007). Without construction changes and costly recertification of the inlet and tubing in the aircraft such addition is not possible. A partial test of the tubing in the container would not provide the conclusive evidence required by the reviewer. We have thus to live with the circumstantial evidence.

The determination of the particulate cut point for the trace gas inlet mentioned in paragraph 3b is not easy as stated by the reviewer. In principle, the cut point can be determined by wind tunnel tests, in-flight comparison to a well-known system or by relying on CFD (computational fluid dynamics) models. The first method requires an immense effort which would far exceed the financial and personnel capacities of the project. We attempted an in-flight comparison with the Jaffe's group during the NOMADSS (Nitrogen, Oxidants, Mercury, and Aerosol Distributions, Sources, and Sinks) flights but this attempt failed because of bad weather. The third approach was tried by one of the coauthors several years ago, but it always turned out that for transitional flows (from highly turbulent to laminar) all tested models could not represent experimental results from a wind tunnel inlet experiment. We thus rely on similarity approach, i.e. for the CARIBIC trace gas inlet we assume a particle transmission similar to the CARIBIC aerosol inlet (the same design principle) and use empirical equations from one of the standard aerosol measurement text books to estimate the particle transmission. This entails, of course, larger uncertainties than results from a wind tunnel, but the resulting numbers are nevertheless reliable.

4. *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 therefrom.*

As discussed before we believe that our data are as good as those of Lyman and Jaffe (2012). Because being the largest data set available for the UT/LS such the data should be used.

5. *I am concerned that the authors have misinterpreted some of the information presented by Lyman and Jaffe (2012):*
  - a) *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).”*
  - b) *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.”*
  - c) *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 the stratospheric lifetime and to not note the long lifetime of sulfate particles in the stratosphere.*

We acknowledge the criticism of the reviewer in paragraphs 5a and 5b and have modified the text accordingly.

We thank for the praise in the paragraph 5c but think that the stratospheric lifetime, if calculated from the Lyman's and Jaffe's data, would be too short. The reason for our belief is based on the Figure 6 of our manuscript. In such plot, their data would be probably located at  $N_2O > 315$  ppb and as such would provide much shorter lifetimes.

6. *The authors used 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  $Hg^{2+}$  in a typical humid atmosphere.*

About Brooks et al. (2014) data we state only that they found GOM maxima at  $\sim 4$  km above ground. This remains true even if the absolute GOM concentrations were 3 times as large as stated by Brooks et al. (2014).

7. *I disagree with the first two sentences of the “conclusions and outlook” section.*
  - a) *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 lifetime in the stratosphere is likely caused by 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.*

- b) *The authors also state that "The second direct implication is that if the lifetime of GEM in the stratosphere with its very high O<sub>3</sub> mixing ratios (typically 1 ppm and more) is quite long, then the GEM + O<sub>3</sub> reaction cannot be important in the troposphere with its low O<sub>3</sub> 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 Hg<sup>2+</sup> reduces back to elemental Hg after it deposits on particles, keeping Hg<sup>2+</sup> concentrations from building up.*

Our statement about the "obvious implication" is based simply on the ratio of relatively short tropospheric and relatively long stratospheric mercury lifetimes and the ratio of stratospheric to tropospheric air mass. The explanation of long stratospheric lifetime in paragraph 7a provided by the reviewer is incorrect. There may be few clouds but there is a lot of particle surface to which Hg<sup>2+</sup> will attach. In fact, measurements by CARIBIC optical particle size spectrometer (OPSS) show always an increase of particle surface concentration just above the tropopause. This is consistent with the maxima of particle abundance found by Wilson et al. (2008) at 200 – 300 ppb N<sub>2</sub>O, depending on whether the aerosol originates from COS oxidation only or additionally from volcanic eruptions. Consequently, there are more than enough particles in the stratosphere to serve as vehicles for mercury removal. This is corroborated by the particulate mercury measurements by Murphy et al. (1998, 2006).

In paragraph 7b the reviewer invokes again the reduction of Hg<sup>2+</sup> to GEM, now on stratospheric particles, as an established and well-known process. This is simply not the case – the reduction is frequently postulated by the modelers only to tune their model results to closer agreement with the observations. Horowitz et al. (2017) proposed a reduction mechanism based on photolysis of aqueous-phase Hg<sup>2+</sup> organic complexes on particles and in clouds in the troposphere. In dry stratospheric air without clouds the occurrence of aqueous-phase Hg<sup>2+</sup> organic complexes is unlikely. Even if such reduction mechanism exists in the stratosphere, it would be much slower than in the troposphere and the resulting elemental Hg would be at O<sub>3</sub> > 150 ppb much more rapidly oxidized to Hg<sup>2+</sup>. Consequently, stratospheric mercury would be all Hg<sup>2+</sup> and attached to particles.

If stratospheric mercury were all Hg<sup>2+</sup> on particles its stratospheric lifetime would be given by the lifetime of particles. According to Hg + Br and Hg + OH/O<sub>3</sub> models of Holmes et al. (2010) the oxidation of elemental mercury stops a few km above the tropopause, somewhat higher for Hg + OH/O<sub>3</sub> than for Hg + Br chemistry. Particles injected into the stratosphere in the tropics live on average several years until they reach the tropopause at mid-latitudes or the polar vortex (Vaugh and Hall, 2002). Substantially longer stratospheric lifetime of mercury calculated by us thus implies that a) some elemental mercury escapes from the troposphere into the stratosphere and b) its oxidation occurs at higher altitudes above the tropopause. Ozone is thus an unlikely oxidant as discussed above. More likely oxidants at higher altitudes are Br atoms released at higher altitudes from halocarbons and reactive O atoms. This is now mentioned in the manuscript.

Waugh, D.W., and Hall, T.M.: Age of stratospheric air: Theory, observations, and models, Rev. Geophys. 40, doi:10.1029/2000RG000101, 2002.