Answers to Reviewer #1:

We want to thank reviewer #1 for the valuable feedback. We improved the manuscript based on these suggestions:

1) A large portion of the manuscript is focused on comparing model results to the observations, but the authors do not provide any numerical measure of agreement between the model and observations. There are several qualitative comparisons, but the lack of numerical comparison makes it very difficult for the readers to draw their own conclusions. I recommend the authors include one or more of the standard metrics (mean bias, mean error, correlation coefficient, etc.) for model-observation comparisons.

A: We agree with the reviewer and added a quantitative comparison to the manuscript. For GEM, we chose to use the mean normalized bias (MNB) and mean normalized error (MNE). We prefer these to the normalized mean bias as it gives more weight to the individual data points. We added three tables to the manuscript: Table 1 gives MNB and MNE for each model separated Europe and North America. Table 2 gives model ensemble MNB and MNE values for altitude slices of 1000m in order to identify whether the models perform better or worse in different altitudes. We found that the model bias and error are mostly uniformly distributed in the troposphere with larger errors in the lower stratosphere. We did find a bias minimum between 2000 and 3000m which we think is an artifact due to the observed decrease of GEM concentrations above the PBL which was mostly in this altitude range. Moreover, for GEM bias and error do not differ significantly between Europe and North America.

Region	Europe		North America	
Species	GEM		GEM	
Model	MNB	MNE	MNB	MNE
GLEMOS	-0.07	0.16	-0.12	0.16
GEOS-Chem	-0.18	0.21	-0.11	0.16
GEM-MACH-Hg	-0.04	0.15	0.08	0.17
ECHMERIT	-0.27	0.34	-0.27	0.28
CMAQ-Hem	-0.20	0.27	-0.23	0.25
WRF-Chem	-0.17	0.25	-	-
CCLM-CMAQ	0.05	0.19	-	-
ENSEMBLE	-0.14	0.21	-0.13	0.20

Table 1: Mean normalized bias and mean normalized error for each model as well as the model ensemble For GEM in Europe and GEM and OM in North America.

altitude	Europe		North America	
	MNB	MNE	MNB	MNE
0 - 1000m	-0.20	0.20	-0.17	0.19
1000 – 2000m	-0.22	0.23	-0.21	0.25
2000 – 3000m	-0.08	0.15	-0.12	0.17
3000 – 4000m	-0.14	0.16	-0.16	0.20
4000 – 5000m	-0.21	0.21	-0.11	0.21
5000 – 6000m	-0.27	0.27	-0.04	0.24
6000 – 7000m	-0.20	0.24	-0.12	0.24
7000 – 8000m	-0.28	0.28	-	-
8000 – 9000m	-0.28	0.28	-	-
9000 – 10000m	-0.24	0.24	-	-
10000 - 11000m	-0.26	0.26	-	-
11000 - 12000m	-0.24	0.25	-	-
> 12000m	0.33	0.41	-	-

Table 2: Model ensemble vertical distribution of model mean normalized bias and mean normalized error for GEM in Europe and North America.

Finally, we calculated the correlation for the vertical oxidized mercury profiles. The results underline the findings already discussed with better performance of Br chemistry for the NOMADSS campaign and better performance of OH and O3 chemistry for the Tullahoma flights.

	Tullahoma flights January and February (Fig. 8)					
	BASE	NOCHEM	BRCHEM1	BRCHEM2	O3CHEM	OHCHEM
GLEMOS	0.76	-0.84	0.46	0.47	0.82	0.56
GEOS-Chem	0.37	0.16	0.37			
GEM-MACH-Hg	0.23					0.23
ECHMERIT	0.77	0.49	0.40	0.40	0.42	0.55
CMAQ-Hem	-0.10				-0.10	
	Tullahoma flights April, May, June (Fig. 9)					
	BASE	NOCHEM	BRCHEM1	BRCHEM2	O3CHEM	OHCHEM
GLEMOS	-0.17	-0.59	-0.80	-0.71	-0.21	0.37
GEOS-Chem	0.39	-0.62	0.39			
GEM-MACH-Hg	0.63					0.63
ECHMERIT	0.93	0.17	0.54	0.52	0.87	0.94
CMAQ-Hem	0.53				0.53	

	NOMADSS flights (Fig. 10)					
	BASE	NOCHEM	BRCHEM1	BRCHEM2	O3CHEM	OHCHEM
GLEMOS	-0.55	-0.60	0.08	0.03	-0.49	-0.54
GEOS-Chem	0.35	-0.49	0.35			
GEM-MACH-Hg	0.07					0.07
ECHMERIT	-0.05	-0.44	0.43	0.39	-0.05	0.03
CMAQ-Hem	0.13					0.13

Table 3: Correlation of individual models for OM profiles depicted in Figures 8, 9, and 10.

2.1) The manuscript lacks a discussion of the representativeness of the observations. I understand that these are the best observations we have, but, as the authors also point out on Line 116-117, aircraft based observations are not representative. Yet they seem to ignore the limited temporal and spatial coverage of the observations when they construct vertical profiles of TM and GEM for the northern mid-latitudes (Fig. 1). It is not clear how these profiles were calculated.

A: So far only single profiles for the vertical Hg distribution were published, mostly only considering GEM or TGM. As the presented manuscript is the first comprehensive global analysis of the vertical distribution of mercury we decided to combine all available aircraft based observations to estimate idealized vertical profiles. These profiles represent our current best knowledge of the distribution of mercury species in the atmosphere and we think that they are an important contribution to the scientific discourse.

We added a paragraph discussing this at the end of Section 2.1:

“These flights cover a large horizontal area, namely the mid latitudes in Europe (45°N - 55°N) and North America (30°N – 45°N) and a large vertical area ranging from the surface up to the lower stratosphere (12000 m). Moreover, comparable flights were performed throughout the year between January and October. Finally, all measurements were performed with Tekran instruments allowing for a comparison of all aircraft based measurements as well as the combination with ground based observations which use similar instruments. It is arguable whether this is already enough data to give us a comprehensive and representative picture of the vertical distribution of mercury in the atmosphere. However, we think that there is an adequate amount of data to allow for more than just an anecdotal investigation of a specific episode. Thus, we combined measurements from all flights in Europe and North America as well as ground based observations for the year 2013 in order to construct idealized seasonal average vertical profiles for TM and OM (Fig. 1)….”

2.2) Secondly, it is not clear for what time period was simulated by the models, how were the models sampled, and what steps were taken to address the issue of representativeness when making comparisons between the model and the observations.

A: We added more information to the model evaluation section 2.4:
“For the model evaluation we used hourly model results for the year 2013 for all models, with the exception of ECHMERIT which provided a lower temporal resolution resulting in 3 hourly average concentrations. The grid cell and time step matching each individual measurement were taken using a 4 dimensional bi-linear interpolation to the nearest model space and time coordinate. For the analysis we used three aggregated model species: TM, GEM, and OM = TM – GEM.“

3) It seems to me that the manuscript was not thoroughly proof-read before publication in ACPD. There are several minor errors that often are distracting. For example, the citation ’Lyman and Jaffe, 2012’, was cited at times as ’Lyman and Jaffe, 2011’, ’Lyman and Jaffe et al., 2012’, ’Lymann and Jaffe, 2012’, and ’Jaffe and Lyman, 2012’). There are also a few instances where abbreviations were used without prior definitions. For example, in the main text RM was used first on Line 515, but not defined until Line 663, while in the abstract it was referred to as oxidized mercury, but in the main text as reactive mercury. I recommend that the manuscript be thoroughly proof-read before final submission.

A: I want to apologize for any inconveniences for the reviewer.
We corrected this and other errors in the citations and did a more thorough proof reading of the revised manuscript.

Specific comments:

1) Sect. 2.3: What was the spin-up period for the sensitivity simulations?

A: The specification for the MMTF scenario model runs was a spin up time of at least 2 years starting from the BASE case spin up. We included this information.

2) Lines 452-457: The underestimate in GOM concentrations seems to be related to the ambient absolute humidity and ozone concentrations and is likely not systematic, as stated by the authors.

A: We clarified this and added more statistical analyses to the results section.

“Generally, the model error can be separated into three parts: The bias, which represents any systematic errors, the variance which gives the variability around the mean value, and the covariance which represents the correlation between model and observations (Solazzo and Galmarini, 2016). By using MDPs we completely remove the bias and all systematic errors from our evaluation. Combining MDP and correlation coefficient, we are able to investigate the models capabilities to reproduce areas with high and low production of oxidized mercury and the influence of different chemistry schemes. The idea behind this is that even if the absolute measurements are not correct, we can use them to identify regions with mercury oxidation in the vertical column.”

3) Line 543: Do the authors mean measured ’variability’ instead of ’uncertainty’? Same for Lines 699, 797.

A: No we refer to the actual uncertainty of the observations as published in the ETMEP, CARIBIC, and NOMADSS datasets. For ETMEP the uncertainty is based on measurements from two co-located Tekran instruments on board of the air-craft for GEM and on denuder blank measurements for GOM. For NOMADSS the uncertainty is based on the lower limit of detection of the DOGHS instrument.

4) Line 567: It is stated that CCLM-CMAQ has the tropopause as its upper boundary, but Fig. 2 shows model values for CCLM-CMAQ up to 18 km.

A: This is only an artifact in the plot which we corrected.

5) Line 580: ’Linear TM’. This term is not clear.

A: We mean the constant TM concentrations (→ the missing trend) inside the free troposphere. We corrected this.

6) Lines 910-920: The authors interpret the high RM above 6 km as being related to stratospheric transport of Br and cite Gratz et al. (2015). However, Gratz et al. (2015) did not find evidence of stratospheric intrusion, and the authors’ conclusions seem contradictory to that study. Is it possible to reconcile these two interpretations?

A: This is correct and we modified our conclusions accordingly.
We still think that stratospheric intrusions are an important source for Br in the upper troposphere during spring time. However, in the episode during July 2013 as described by Gratz et al. (2015) the water vapor concentrations seem too high to indicate a stratospheric origin of the air mass (the low ozone concentrations however could also be explained by depletion due to high Br concentrations).

“Our interpretation of the observations is that stratospheric intrusions and tropopause folds, which mainly occur during spring time, play an important role for elevated OM concentrations in the upper FT at altitudes above 6000m. The frequency of stratosphere to troposphere transport is regionally variable and has shown to be most common in the latitudes where the measurements were performed. However, also long range transport of marine bromine species as observed by Gratz et al. (2015) during the NOMADSS flights can be an important source of stratospheric Br. Thus, we emphasize the importance of further research regarding the atmospheric bromine cycle to better understand the oxidation pathways of mercury. Besides bromine species, stratosphere to troposphere transport could also be a source for OM already formed in the lower stratosphere. This could also explain the missing correlation of ozone concentrations and GEM/TM ratios measured by the CARIBIC aircraft in the upper FT.”

7) Line 927: ’OH seems a plausible explanation’. How about O3? The models with O3 chemistry had better agreement with the observed concentrations for the NOMADSS profile.

A: This is correct, we adjusted the paragraph accordingly. Moreover, we included the calculated correlation coefficients into the discussion.

8) Lines 940-942: I am not sure how the higher effective height of emissions would affect GEM concentrations in the upper troposphere.

A: Emissions to high altitudes, especially when the effective emission height is above the PBL, have a longer atmospheric lifetime. On average, it will simply take longer for the substance to collide with a surface and undergo dry deposition. Moreover, is will not be scavenged by low altitude precipitation. This is only a hypothesis, but has been shown to be true for other pollutants e.g. SO4 emissions from coal fired power plants (Bieser et al., 2011) which we think is a good proxy for Hg emissions from the same source.

Bieser J, Aulinger A, Matthias V, Quante M, Denier van der Gon HAC. 2011. Vertical emission profiles for Europe based on plume rise calculations. Environ Pollut 159: 2935–2946. doi: 10.1016/j.envpol.2011.04.030

9) Table 1: Is OH an aqueous phase oxidant of GEM in the GEOS-Chem model?

A: Yes, OH is only used in the aqueous phase. We improved Table 1 so it is easier to read.

10) Table 2: Since the results of the sensitivity simulations were not available for all models, I suggest adding a column specifying which models participated in each sensitivity run.

A: We added this to Table 2.

11) Title: Perhaps the title should contain “Vertical and inter-hemispheric distribution of mercury species”.

A: This is a very good suggestion and we changed the title accordingly.

Technical comments:

A: Thanks for the detailed list of minor errors. We corrected these in the revised version of the manuscript.

1) Lines 83-84: Grammar

2) Lines 98-102: Grammar

3) Line 143: Spelling, ’Woll’

4) Line 144: What does DOHGS stand for?
University of Washington Detector for Oxidized Hg Species

5) Line 620: ’14th June’ seems to be typo.

6) Line 665: Do the authors mean equilibrium between GOM and PBM?
yes

7) Table 1: No emission speciation information for WRF-Chem

8) Table 1: No reference for natural emissions for the first six models.

9) Table 1: Missing footnote ’a’.
we corrected table 1

10) Table 2: Refers to the emission set as UNEP2010, while Table 1 refers to it as AMAP.
Both should be AMAP/UNEP

11) Fig. 3: Lower panel. Is MDP for TM or GEM?
Fig 3 to 6 give MDP for GEM only.

12) Fig.7-9: Is Hg2+ different from RM?
Hg2+ is identical to oxidized mercury (OM) We think this is correct as we now use the term oxidized mercury instead of reactive mercury (RM).

13) Figs. 8-9: Units for Hg2+.
Corrected to pg/m^3

14) Fig. 8: Should the x-axis label be MDP instead of PMB?
yes