We completely agree with the comments and changes in lines 78, 79, 88, 95, 129, 158, 219, 232, 251. These will be implemented in the revised manuscript. Please find the responses for the other comments below.

Comment 5 (Line 103): Calibration uncertainty cannot be calculated with 1 point calibration.

R: We calculated the uncertainty of the single point calibration for the Manual AFS as the standard uncertainty which was calculated from the repeatability of the standard. The system has been checked previously for its linearity in this concentration range.

Comment 6 (Line 117): How many calibration points?
R: 8 calibration points.

Comment 8 (134): This contradicts to the observation of impaired signals with longer sampling. Unless there's sensitivity drift the longer the sampling the lower should be the RSD. Can the authors clarify, please. Is the ambient Hg changing while sampling?

R: This has been done due to the fact that there is a drift when the autozero is done at higher intervals of time. In order to not have this issue we did bracketing sequences of autozero – blank – standard and so on. The Hg concentration changes when sampling ambient air.

Comments 9-10 (147-148): Not clear

R: We will change the phrase to : ″As the $N_2$ flow is lower than the flow of the pump inside the Zeeman AAS, a T-piece was placed between the impinger and the device. This allows for the air from outside to enter the system in order to compensate for difference in flows and pressure. The air passes through a Hg filter in order to assure that it does not interfere with the analysis.″ which was also suggested by Reviewer 2.

12 (213). What is data was analysed with ANOVA?

R: The tests were used to determine whether significant differences exist between atmospheric Hg concentration obtained by different analysers.

14 (224). Saturation issues will be missed with this approach. The calibration standards need to be randomised.

R: Yes, in the line 234 we state that this was tested.

15 (225). Can this information be supplied please. Usually blanks would have lower sd but higher RSD.

R:Blank data, together with corresponding SD and RSD%, will be supplied with the comments in excel format. Indeed RSD% is higher for blanks than for standard especially

when comparing to higher concentration standard. The statement was intended for the standards in the lower concentration range as it can be seen in the attached data.

16 (227). This contradicts to the statement about the stability of the analyser. Blank drift modeling would be more efficient than single point corrections.

R: Single point calibration was used only in the AFS-M method as its linearity has been continuously verified before and during these experiments in this concentration range. For the rest it was multiple point calibration. For the rest it was multiple point calibration.

18 (233). was statistical test for difference performed? If yes, the results should be stated here.

R: We performed statistical test for difference between calibration curves and the results are presented in the text. We have also performer Parallel Lines Analysis (PLA) which is not included in the manuscript between two calibration curves which were in the same concentration range (one with randomised calibration points and one calibration points analysed from the lowest to the highest concentration) and there is no statistical difference between them with P = 0.1995 for the slopes and P = 0.5493 for the intercepts. The equation for the randomised calibration curve is y = 0.6616x - 0.0088 and for the other is y = 0.6412x + 0.22.

20 (254). The figure shows two distinctive slopes. Also, how the authors would explain the fact that the 100 ng m-3 standard signal with uncertainty is not within the calibration confidence interval?.

R: Yes, this is what we wanted to show that even if the standards were analysed in the same experiment, there are issues in the lower range as can be seen by the difference in the slopes that is over 30%. The confidence interval was calculated using all 10 calibration points, and being so many in the lower range of the calibration curse, the confidence interval is shorter that in the case in which we would not use the lowest 5 calibration points.

21 (276). What does it mean "extended" uncertainty?

R: It is a mistake. We will change it to ″expanded uncertainty″.

22 (312). This uncertainty is quite high. Limit of quantification values should be stated with acceptable uncertainty. Lower values should not be reported since they are not informative.

R: The LOQ is approximately 3 ng $m^{-3}$ with slight variations in different days, analysed by using a carbon trap. Due to the fact that the expanded uncertainty falls above the LOQ, we wanted to include the data, given that there are lots of measurements in this range of concentrations reported in other publications.