TAbMEP Assessment: ICARTT Toluene Measurements

1. Introduction

Here we provide the assessment for the toluene measurements during the summer 2004 ICARTT field campaign [*Fehsenfeld et al.*, 2006, *Singh et al*., 2006]. The inter-platform assessment is based upon the four wing-tip-to-wing-tip intercomparison flights conducted during the field campaign. The intra-platform comparison involving two NOAA WP-3D instruments is conducted using all the available data from the mission. Recommendations provided here offer TAbMEP assessed biases only for each of the WP-3D measurements and a systematic approach to unifying the ICARTT WP-3D data for any integrated analysis. These recommendations are directly derived from the instrument performance demonstrated during the ICARTT measurement comparison exercises and are not to be extrapolated beyond this campaign. Low toluene conditions were encountered during the comparison between the NASA DC-8 and FAAM BAe-146 instruments. This resulted in a very small number of points for analysis, which prevent us from carrying out a meaningful assessment. Thus recommendations are not given in terms of the measurement uncertainties for these measurements. The comparison between the DC-8 and WP-3D instruments similarly was based on very limited number of samples and therefore the comparison was not used for bias assessment. This report serves as a record for the DC-8 and BAe-146 ICARTT toluene measurement comparisons.

2. ICARTT Toluene Measurements

Three whole air sampler instruments and one proton transfer reaction mass spectrometry instrument were deployed on three aircraft. Table 1 summarizes these techniques and gives references for more information.

Table 1. Toluene measurements deployed on aircraft during ICARTT

Aircraft	Instrument	Reference
NASA DC-8	Whole Air Sampler (WAS)	*Colman et al.* [2001]
NOAA WP-3D	Whole Air Sampler (WAS)	Contact PI: eatlas@rsmas.miami.edu
	Proton Transfer Reaction Mass Spectrometry (PTRMS)	*de Gouw et al.* [2007]
FAAM BAe-146	Whole Air Sampler (WAS)	*Hopkins et al.* [2003]

3. Summary of Results

Table 2 summarizes the PI reported uncertainties for each of the four toluene measurements involved in the intercomparisons. Assessed bias is included for the two NOAA WP-3D instruments and is based on the comparison for those instruments during all WP-3D flights only for levels less than 250 pptv. More detailed description is provided to illustrate the process for assessment of bias in Section 4.1. Precision assessment is not provided here as the TAbMEP prescribed procedures cannot be applied for to ICARTT toluene measurements. This is because the reported data have large time gaps and a small data population (see Section 3.1 of the introduction). The assessed bias reported in Table 2 (see Section 4.1 for details) can be applied to maximize the consistency between the data sets, by subtracting the value from the reported data to 'unify' the data sets. Low toluene levels limited the comparison between BAe-146 and DC-8 near the limit of detection (LOD) levels. Five BAe-146 points with values above LOD suggest that the BAe-146 measurement may be up 11 pptv higher than that of the DC-8. A

similar situation was also encountered during the WP-3D/DC-8 comparison. The results from the analysis of a small data population show that DC-8 data is, on average, 30 pptv higher than WP-3D PTRMS values with a maximum difference of 54 pptv for a given point. On average, the DC-8 is 0.1 pptv lower than the WP-3D WAS measurement with maximum difference of 24 pptv for a given point.

Table 2. Recommended ICARTT Toluene measurement treatment

Aircraft/Instrument	Reported 2σ Uncertainty	Assessed Bias (pptv)
NASA DC-8 WAS	10%	N/A
NOAA WP-3D WAS	10 – 15%	-13.82 + 0.103 Toluene $_{\text{WP-3D WAS}}$
NOAA WP-3D PTRMS	15% [a]	17.39 – 0.129 Toluene $_{\text{WP-3D PTRMS}}$
FAAM BAe-146 WAS	Point by Point, average: 58% [b]	N/A

[a] This is the calibration accuracy. The precision is determined by ion counting statistics and is not accounted for here. For full details see: *de Gouw et al.* [2007]
[b] The average encompasses only the comparison period for DC-8/BAe-146.

Figure 1a displays the PI reported uncertainties for the four toluene instruments, and Figure 1b shows the recommended bias for two of the toluene instruments for which a bias could be found using the intercomparisons. The BAe-146 uncertainty was reported for each measurement point and the measured values were all under 150 pptv.

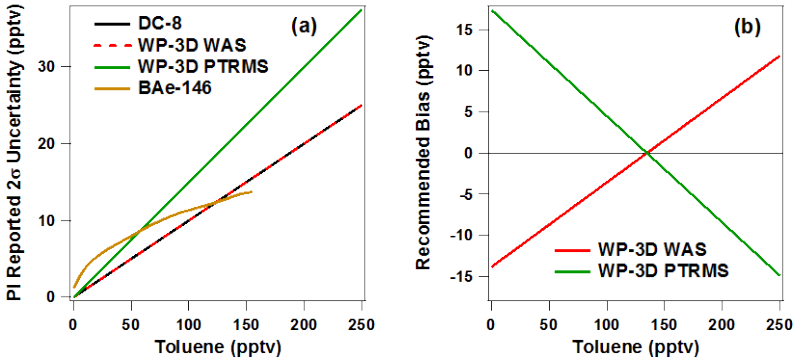

Figure 1. Recommended bias (panel a) and PI reported 2σ uncertainty (panel b) for DC-8 WAS (black), WP-3D WAS (red), WP-3D PTRMS (green), and BAe-146 (gold) as a function of toluene level. Values were calculated based upon data shown in Table 2. The BAe-146 reported uncertainty was calculated using a function derived from the 60 second merge file.

4. Results and Discussion

4.1 Bias Analysis

Figure 2 shows the time series plots for the DC-8/WP-3D comparisons. All three comparison flights are at low toluene levels. The WP-3D PTRMS in particular has several points below the detection limit. There are very few overlapping points for the comparisons. It is not statistically significant to show the linear regression for each individual flight. Therefore, linear regression is performed over the data combined from all three DC-8/WP-3D flights. The integration time difference in WAS techniques presents some challenges in analyzing the data. The DC-8 data have an integration time of approximately 60-70 seconds, while the WP-3D data have an

integration time between 6-11 seconds. For these measurements to be considered simultaneous, the start and stop times of the WP-3D data must fall within the start and stop times of the DC-8 data. In order to maximize the data coverage for statistical analysis, one exception is made to this rule. If the shorter (WP-3D) integration time falls outside the longer integration time by no more than two seconds, the data points are also considered to be simultaneous. It is noted here the integration time difference may potentially be another factor leading to the difference between measurements.

As shown in Figure 3, all DC-8 reported values are under LOD within the comparison period and six BAe-146 points are above LOD for this period. These BAe-146 values are less than 11 pptv with an average level of about 6 pptv, which suggests that the BAe-146 potentially can be up to 11 pptv higher than DC-8. BAe-146 integration times range from approximately 30-60 seconds. Since the DC-8 and BAe-146 have similar integration times, the measurements are considered correlated if the midpoint of DC-8 or BAe-146 fall within the start and stop time of the other measurement.

Figures 4a and 4b show both of the DC-8/WP-3D comparisons. Both cases show a small number of points. The regression between DC-8 and WP-3D PTRMS in Figure 4a yields a significant R^2 value with large scatter. The regression line shown in Figure 4b is largely driven by the two points with high values. We do not believe the regression results shown in Figure 4 can be considered as statistically robust. Therefore, the regressions were not included in Table 3 and were not used to determine the best estimate bias, as discussed below. As shown in Figure 5, there is an overall offset between the two instruments of about 30 pptv, PTRMS being higher at an average PTRMS level of 69 pptv. The comparison between the DC-8 and WP-3D WAS has only six comparison points. There were several points where the DC-8 reports LOD and the WP-3D WAS does not. Figure 6 shows the residuals for the comparison and there is an overall average near zero.

The comparison between the two WP-3D instruments uses all available flights during ICARTT as shown in Figure 7. Table 3 includes the linear relationships from both Figure 7a and 7b. The two equations are not dramatically different. The PTRMS reading is below the detection limit for about 53% of the comparison points, where the WP-3D WAS instrument reported a value above LOD. These points are included in all of the figures, but not included in any calculations for the WP-3D WAS/WP-3D PTRMS comparison. This comparison is also published in *de Gouw et al.* [2006] where LOD points were included, yielding slightly different results. Figure 8 shows the residuals for the two instruments. Many of the comparison points that are not LODs were well outside of the range of results expected from the reported measurement uncertainties. However, our measurement uncertainties do not take into account ion counting for the PTRMS, and as shown in *de Gouw et al.* [2007] the scatter was as expected with the inclusion of ion counting statistics.

The difference between the instruments ranges from about -100 to 100 pptv with an average difference of about 10 pptv. The relative residuals are shown in Figure 9, and although there are several large relative residuals, the bulk of the comparison lies within the -1 to 1 range. Bias correction was made to the relative residuals using the assessed bias described below for the WP-3D WAS and WP-3D PTRMS instruments.

Section 3.3 in the introduction describes the process used to determine the best estimate bias. The reference standard for comparison (RSC), as defined in the introduction, is constructed by averaging the NOAA WP-3D measurements with equal weights. The linear regression for points less than 250 pptv (shown in Figure 4) was used for the WP-3D WAS/WP-3D PTRMS bias calculations because it better represents the where the bulk of the data was measured throughout the mission. The regression including points above 250 pptv is driven by six high points. The resulting RSC can be expressed as a function of the WP-3D WAS toluene measurements as the following:

$$RSC_{Toluene} = 13.82 + 0.897\ Toluene_{\text{-WP-3D WAS}}$$

The RSC is then used to calculate the best estimate bias as described in Section 3.3 of the introduction. Table 3 summarizes the assessed measurement bias for each of the two WP-3D ICARTT toluene measurements. Note that additional decimal places were carried in the calculations to ensure better precision. It is also noted that the intercept in the equations listed in Table 3 should not be viewed as an offset. These linear equations are used to best describe the linear relation between the measurements.

Table 3. ICARTT Toluene bias estimates

Aircraft/ Instrument	**Linear Relationships**	**Best Estimate Bias (a + b Toluene) (pptv)**
NOAA WP-3D WAS	$Toluene_{\text{WP-3D WAS}} = 0.00 + 1.00\ Toluene_{\text{WP-3D WAS}}$	$-13.8 + 0.103\ Toluene_{\text{WP-3D WAS}}$
NOAA WP-3D PTRMS	Toluene < 250 pptv $Toluene_{\text{PTRMS}} = 27.6 + 0.795\ Toluene_{\text{WP-3D WAS}}$	$17.4 – 0.129\ Toluene_{\text{WP-3D PTRMS}}$
NOAA WP-3D PTRMS	All data $Toluene_{\text{PTRMS}} = 24.3 + 0.836\ Toluene_{\text{WP-3D WAS}}$	N/A

As part of ICARTT intercomparison standard exchange exercises, University of California, Irvine (UCI) prepared the common VOC samples that were sent to University of Miami (Miami), University of New Hampshire (UNH), and University of York (York) for their lab analyses. Some of these same institutions had instruments on the following planes during ICARTT: UCI on the DC-8, Miami on the WP-3D with the WAS instrument, and York on the BAe-146. Toluene was one the 9 species involved in the comparison. We believe that the inclusion of this comparison result in this report will help the readers better understand the airborne intercomparison analysis. The difference in this lab comparison between the DC-8 and WP-3D WAS instruments was 27 pptv, DC-8 being higher, with a DC-8 instrument reading of 277 pptv. From the same lab comparison, the difference between the DC-8 and BAe-146 was 42 pptv, DC-8 being higher. Comparing the ICARTT flights to this lab comparison shows different results in the case of DC-8/WP-3D WAS. The airborne flights provide few data points to compare and on average the difference between the instruments was approximately zero. Although there was not a large difference between the two instruments in flight, there is limited data and the flights measured low toluene levels (< 150 pptv). In the case of the DC-8/BAe-146 comparison, no conclusions could be made from the flight data because there were no comparison points.

4.2 Precision Analysis

A detailed description of the precision assessment is given in Section 3.1 of the introduction. The IEIP precision, expected variability, and adjusted precision could not be calculated for toluene because of the small number of points and large time gaps between measurements.

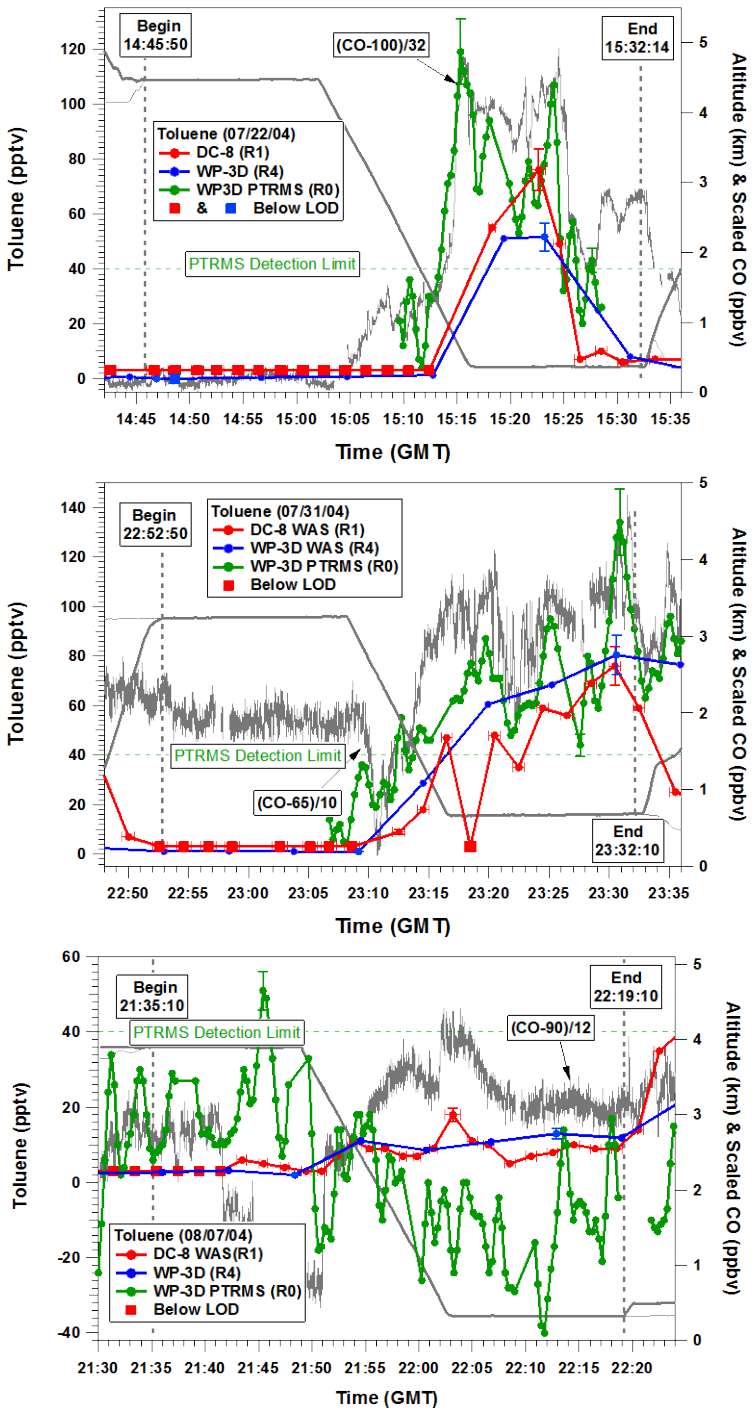

Figure 2. Time series of toluene measurements and aircraft altitudes from two aircraft on the three intercomparison flights between the NASA DC-8 and the NOAA WP-3D. Error bars represent the PI reported uncertainty. In parenthesis next to the plane is the data version number.

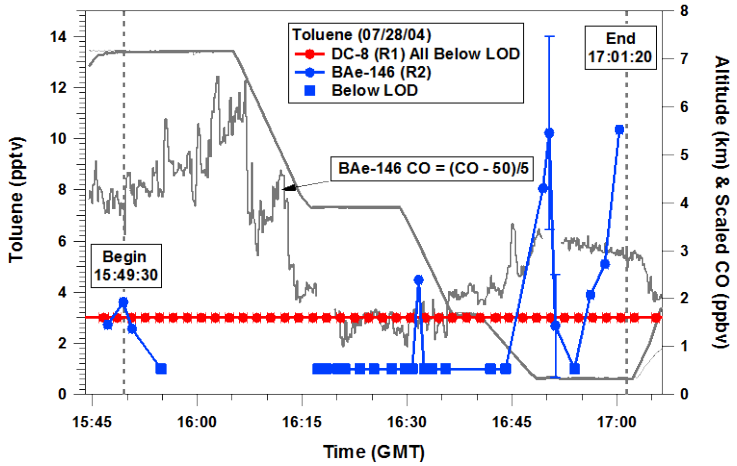

Figure 3. Time series of toluene measurements and aircraft altitudes from the intercomparison flight between the NASA DC-8 and the FAAM BAe-146. In parenthesis next to the plane is the data version number. Error bars represent the PI reported uncertainty.

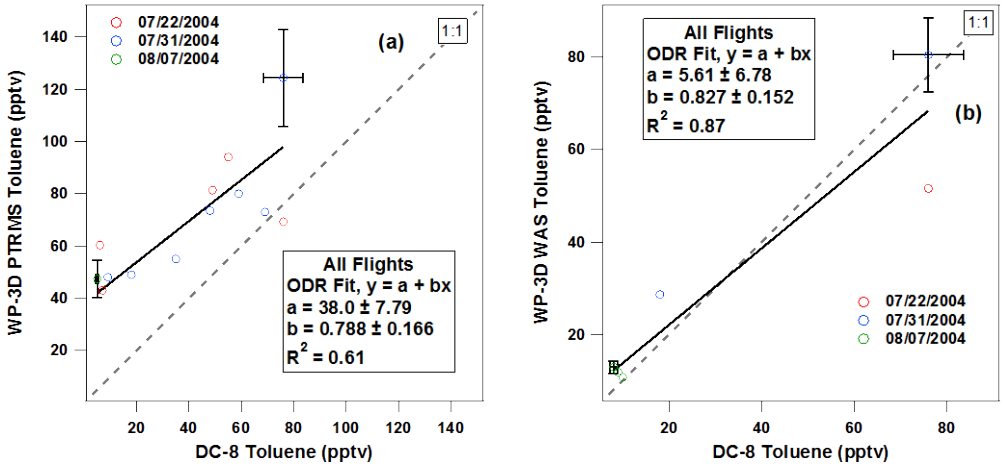

Figure 4. Combined correlation for the toluene measurements on NASA DC-8 and the NOAA WP-3D for 7/22, 7/31, and 8/07 2004. Error bars represent the PI reported uncertainty.

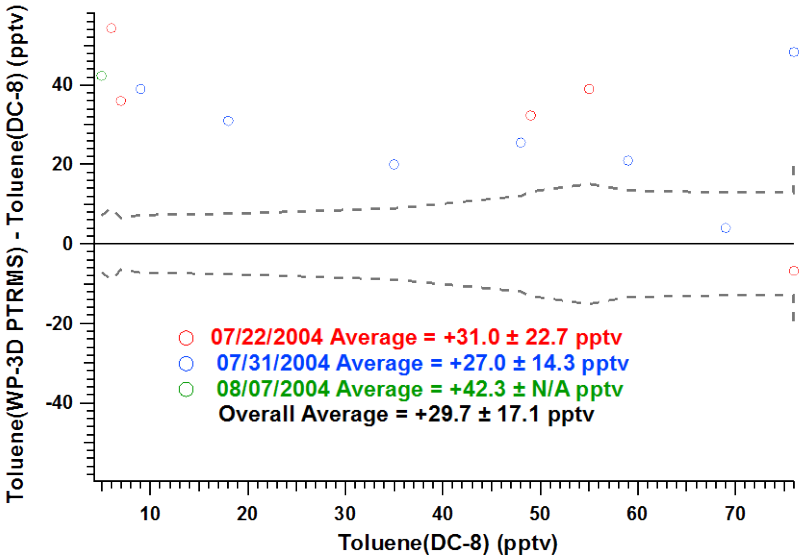

Figure 5. Difference between the toluene measurements from the three DC-8/WP-3D PTRMS intercomparison flights as a function of DC-8 toluene. The dashed lines indicate the range of results expected from the reported measurement uncertainties.

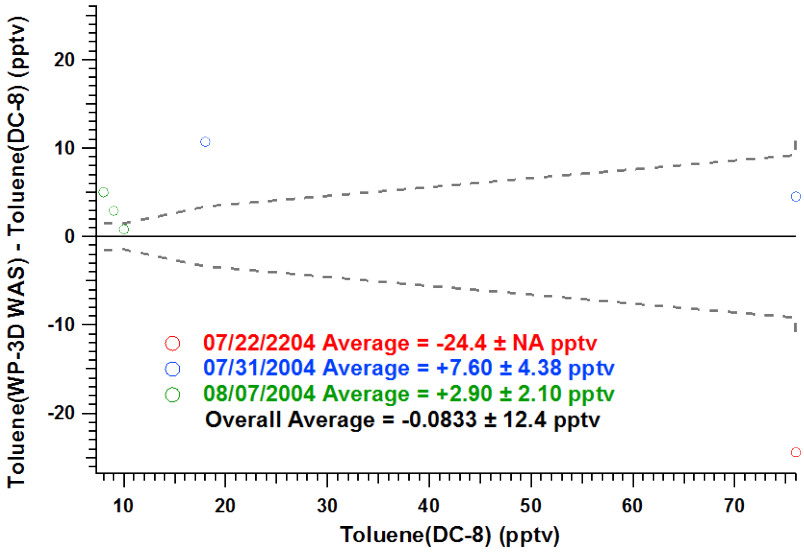

Figure 6. Difference between the toluene measurements from the three DC-8/WP-3D WAS intercomparison flights as a function of DC-8 toluene. The dashed lines indicate the range of results expected from the reported measurement uncertainties.

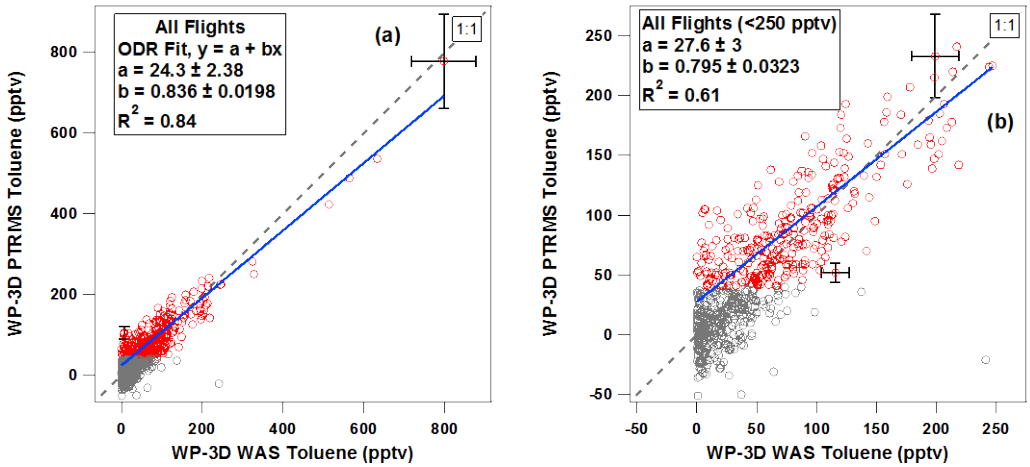

Figure 7. (left panel) Correlation of WP-3D WAS and PTRMS toluene measurements for all ICARTT flights. (right panel) Correlation of WP-3D WAS and PTRMS toluene measurements for all ICARTT flights for toluene levels less than 250 pptv. In both graphs, error bars represent the PI reported uncertainty and gray points represent PTRMS points below the detection limit. These LOD points were not included in regression analysis. Similar analysis was published in *de Gouw et al.* [2006] in which LOD points were included.

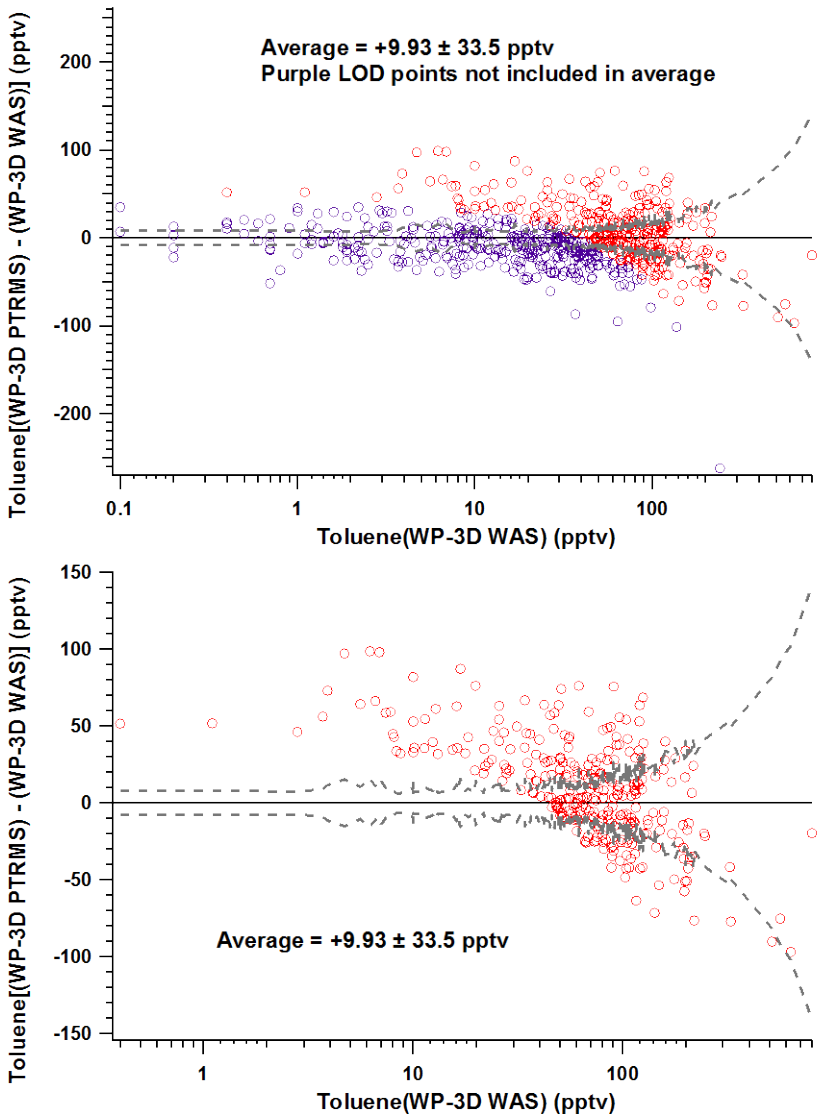

Figure 8. (top graph) Difference between WP-3D WAS and PTRMS toluene measurements from all available WP-3D flights as a function of WP-3D WAS toluene including points below LOD. Purple points represent PTRMS points below the detection limit. (bottom graph) Difference between WP-3D WAS and PTRMS toluene measurements from all available WP-3D flights as a function of WP-3D WAS toluene for only points above LOD. In both graphs, the dashed lines indicate the range of results expected from the reported measurement uncertainties.

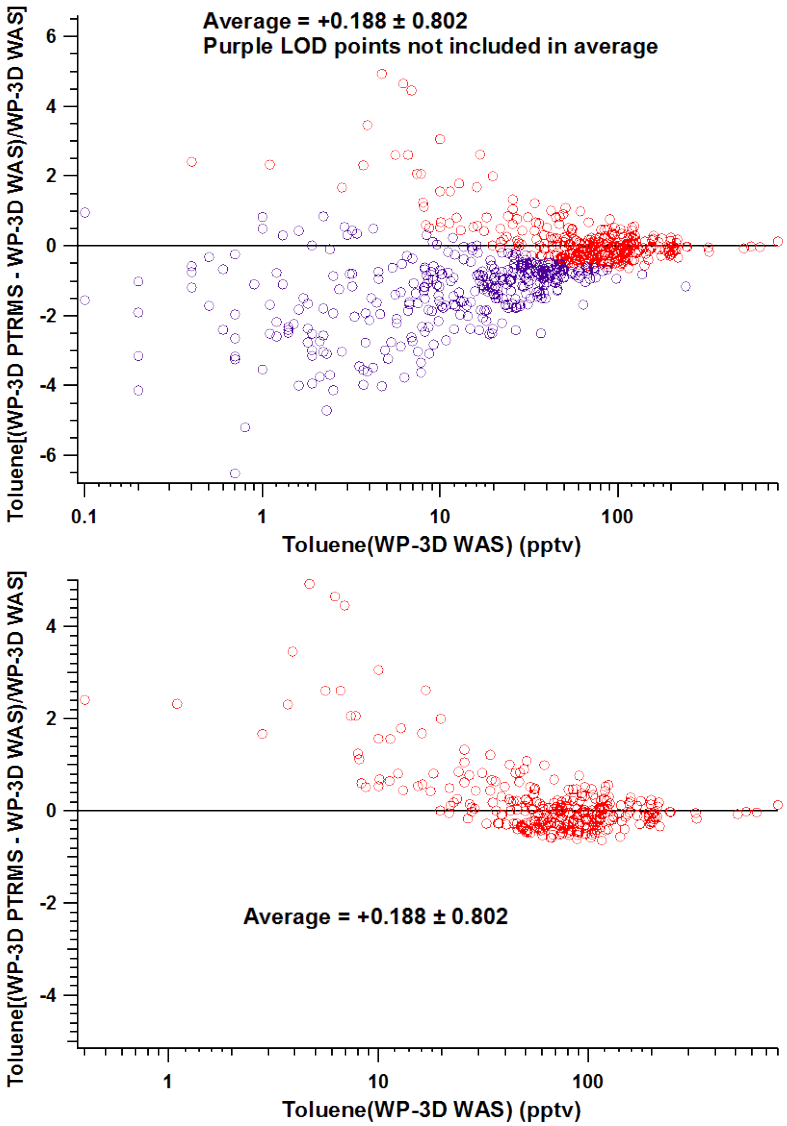

Figure 9. (top graph) Relative difference between toluene measurements from WP-3D WAS and WP-3D PTRMS for all flights as a function of WP-3D WAS toluene including points below LOD. Purple points represent PTRMS points below the detection limit. (bottom graph) Relative difference between toluene measurements from WP-3D WAS and WP-3D PTRMS for all flights as a function of WP-3D WAS toluene for only points above LOD. For both graphs, a correction was made to account for bias.

References

Colman, J.J., et al. (2001), Description of the analysis of a wide range of volatile organic compounds in whole air samples collected during PEM-Tropics A and B, *Anal. Chem* ., 73, 3723-3731.

de Gouw, J. A., and Warneke, C. (2007), Measurements of volatile organic compounds in the Earth's atmosphere using proton-transfer-reaction mass spectrometry, *Mass Spec. Rev.*, 26, 223-257.

de Gouw, J. A., et al. (2006), Volatile Organic Compounds Composition of Merged and Aged Forest Fire Plumes from Alaska and Western Canada, *J. Geophys. Res.*,*111*, D10303, doi:10.1029/2005JD006175.

Fehsenfeld, F. C., et al. (2006), International Consortium for Atmospheric Research on Transport and Transformation (ICARTT): North America to Europe—Overview of the 2004 summer field study, *J. Geophys. Res.*, *111*, D23S01, doi:10.1029/2006JD007829.

Hopkins, J. R., K. A. Read, and A. C. Lewis. (2003), A Column Method For Long-term Monitoring Of Non-Methane Hydrocarbons (NMHCs) and Oxygenated Volatile Organic Compounds, *Journal of Environmental Monitoring*, 5 (1), 8-13.

Singh, H. B., et al. (2006), Overview of the summer 2004 Intercontinental Chemical Transport Experiment-North America (INTEX-A), *J. Geophys. Res.*, *111*, D24S01, doi:10.1029/2006JD007905.