Meeting Summary

First Tropospheric Airborne Measurement Evaluation Panel Meeting

Prepared by Gao Chen, Mary Kleb, and Margaret Pippin

A. Overview:

The first Tropospheric Airborne Measurement Evaluation Panel (TAbMEP) meeting was held in Baltimore, MD August 19-21, 2008. This meeting was sponsored by NASA's MEaSUREs (Making Earth System data records for Use in Research Environments) program through the funded project "Creating a Unified Airborne Database for Assessment and Validation of Global Models of Atmospheric Composition," which is conducted by a team of NASA Langley scientists, Drs. Gao Chen, Mary Kleb, Margaret Pippin, Jennifer Olson, and SSAI support contractors (hereafter, refered to as the LaRC MEaSUREs project). The TAbMEP meeting and the LaRC MEaSUREs project have received broad endorsement and participation from NASA, NOAA, NSF, EPA, DOE and IGAC. TAbMEP is a group of measurement and modeling experts representing a broad spectrum of trace gas and particle measurement techniques/ instruments as well as global and regional models (See Appendix A for a list of attendees). TAbMEP serves as a steering committee to guide the project in achieving its overarching goal to generate unified data products for model assessment and validation. The primary objectives of TAbMEP are to provide an unbiased assessment of the measurement uncertainties and measurement consistency for the historical airborne observations, and to establish systematic approaches for combining airborne data sets from multiple instruments/techniques and aircraft platforms. It is also the goal of TAbMEP to provide a forum for direct dialogue between the modeling and measurement communities to facilitate better understanding of measurement uncertainties and modeling needs of the respective communities. We plan to host annual TAbMEP meetings for the duration of the LaRC MEaSUREs project (four years).

Through the series of TAbMEP meetings we intend to assess the historical airborne data sets spanning from the 1980s to the present. Earlier meetings plan to concentrate on the more recent data sets. The 2004 ICARTT observational database was the focus for the first TAbMEP meeting. This choice reflects two important considerations. First, an extensive series of measurement comparison exercises were conducted during the ICARTT field campaigns enabling quantitative assessment of measurement consistency between the deployed instruments and airborne platforms. Second, it addresses the need of the HTAP community in their execution of Experiment Set 3 which bases model assessment on the ICARTT data.

The specific observations addressed during the meeting were:

- Gas phase measurements: O_3, CO, NO, NO_2, HNO_3, PAN, CH_2O, SO_2, H_2O, methane, ethane, ethyne, propane, n-butane, benzene, toluene, and isoprene.
- Particulate phase measurements: total number density, submicron volume, total volume, scattering, absorption, and mass concentrations of sulfate, ammonium, and nitrate.
- Meteorological and radiative measurements: temperature, wind speed, $j(NO_2)$, and $j(O^1D)$.

During the first TAbMEP meeting, each of the trace gas and particle measurements was discussed individually. The discussions were constructive, consensus building, and aimed to address five specific topics: measurement uncertainty (both random and systematic), measurement consistency, suitability for model assessment, ways to unify the multiple instrument/platforms data sets, and action items for further analysis. A summary of the TAbMEP meeting discussions, the follow-up analysis, and panel assessment will be released to the public via the TAbMEP Assessment Report. The Assessment Report will provide sufficient detail to highlight the issues raised during discussions, to support arguments for the panel uncertainty assessment, and to substantiate the panel recommendation regarding the suitability for use in model assessment. It will include a brief description of the ICARTT field campaign including science objectives, airborne platforms, and execution/flight plans. The Assessment Report will also include a summary table for each species (see Appendix B) containing metadata (e.g., data source, PI, platform, etc.), panel assessed uncertainty and consistency, PI uncertainty where available, and panel recommendation for use in models. The lead facilitators for each measurement will have the primary responsibility of drafting the panel assessment and corresponding with the PIs. The draft assessments will be reviewed and revised by the panel and sent back to the PIs for comment. The final TAbMEP Assessment Report is due to be released to the public by June, 2009.

The overarching goal of the LaRC MEaSUREs project is to create a unified airborne database for tropospheric chemistry observations. The format and content of the unified airborne database were discussed extensively during the first TAbMEP meeting. Both modeling and measurement experts recognized the critical need for a unified airborne database with adequate metadata describing the data source and measurement uncertainties. For each species, the panel recommended including a brief panel assessment summary in the metadata, e.g., panel assessed uncertainty, consistency, and suitability for model assessment. The panel tentatively agreed that each file should correspond to one field campaign (containing all the species/parameters for all the flights). The panel recognized that generating a data product which is user friendly and provides sufficient information for a meaningful model comparison will likely be an iterative process. The modeling experts prefer merged files in netCDF format. Some modeling experts requested 60 sec. merges while others disagreed. The panel did not reach consensus on the time scale for the merged files however they did agree the merged files were needed. Initially a 60 second merge will be provided with the possibility of offering additional flexibility (allowing the modelers to specify the time interval for merge and geographical location of interest) at a future date. The disadvantage of this approach is that it complicates the uncertainty when the merge time interval is smaller than the measurement integration time. The panel suggested that the LaRC MEaSUREs team work closely with the modelers to explore the best method for meeting the needs of the modeling community. The unified database will initially be stored on the LaRC airborne data website and eventually transferred to the LaRC ASDC. Given the strong support from the both the measurement and modeling communities and the recognized need for unified data products for model assessment and validation, panel members suggested we consider seeking further funding sources to sustain this activity beyond the term of the LaRC MEaSUREs project.

B. General Panel Recommendations:

1. The scope of the TAbMEP meeting is to evaluate the implementation of techniques, but not to critique the techniques themselves.

2. The panel views the internal estimate of instrument precision (IEIP) as a useful data-driven independent check on the PI reported uncertainties. This analysis should be performed on all applicable data (i.e., high time resolution and continuous). Detailed description of IEIP analysis procedures are provided in Appendix C.

3. The measurement consistency analysis for the intercomparison data will report the absolute or relative difference between coincident points in addition to the orthogonal distance regression (ODR) slopes and intercepts. Appendix C outlines systematic analysis of the measurement consistency assessment procedures.

4. In several cases, there are significant differences between measurements which cannot be reconciled with our current understanding of the instruments or their calibration procedures. Nevertheless, the measurements provide useful information on atmospheric variability for the measured species. For these

cases, the panel recommended more than one measurement as suitable for model assessment, even though it is not clear which measurement is closer to the actual ambient value. These measurements will be unified by increasing systematic uncertainties so that all the measurements will be encompassed within 2-σ total uncertainty limits. The individual data sets will not be adjusted even when significant differences occur, since the average is unlikely to be closer to the actual ambient value. A detailed algorithm is given in Appendix C.

5. For future field campaigns, the PIs should be required to report uncertainty in the form of +/-(x +y%), representing both percent and constant for uncertainty. For example, WP-3D ozone uncertainty is reported as 0.1 ppb + 3%.

6. For future field campaigns, the PIs should be encouraged to reconcile the measurement differences during or shortly after the field campaign.

7. In future field campaigns when multiple particle measurements are deployed, the PIs are encouraged to discuss and decide before the campaign to set common measurement conditions (e.g., inlet size cut) and to report common quantities (e.g., submicron volume).

C. Action Items and Timetable:

1. Follow-up analysis of ICARTT data (Sept - Dec. 2008).

    a. Gain access to DLR Falcon data sets (Kleb, Evans).

    b. Evaluate temp, wind speed, wind direction, $j(NO_2)$, $j(O^1D)$, dew point (Chen, Kleb, Pippin).

    c. Investigate and if appropriate make comparison plots for NO (Pippin, Chen).

2. Lead Facilitators draft TAbMEP assessment report on corresponding species/parameters (Jan. - Feb., 2009).

3. Panel review of the draft report (Mar. - April, 2009).

4. Request comments from PIs (May, 2009).

5. Panel finalizes report with consideration of PI comments (June, 2009) and release to the public.

D. Discussion Summary on each Species/Parameter:

CO (Leads: D. Parrish, G. Diskin, M. Evans)

- DC-8 and WP-3D CO are state of the art and need only slight increases in uncertainties.
- Determine if bias (offset between DC-8 and WP-3D) is a function of concentration level or variable between the different intercomparison flights.
- BAe-146 data may need to be reprocessed with new calibration standard (Evans).
- Include NOAA WP-3D WAS data in intercomparison.
- Review tank intercomparison results (Apel and Ryerson).

O_3 (Leads: M. Avery, T. Ryerson, M. Evans)

- Both O_3 measurements are state of the art, however, there is a few ppb (e.g. 2 ppb) offset between the DC-8 and WP-3D. As analysis progresses a more precise value will be given.
- Minor adjustment (e.g. 1%-2%) of uncertainties is necessary to unify the DC-8 and WP-3D data. Final adjustments will be determined by ongoing analysis.
- BAe-146 data are reasonable even though the difference between the BAe-146 and DC-8 is larger than the difference between the DC-8 and WP-3D. BAe-146 data uncertainty may be subject to a larger adjustment to be consistent with the DC-8 and WP-3D O_3 data. Final adjustments will be determined by ongoing analysis.
- Document H_2O correction and absorption cross section used to derive ozone for each measurement.

NO_2 (Leads: T. Ryerson, R. Cohen, M. Evans)

- Slight adjustments in uncertainties are recommended to encompass both measurements.
- Data should be segregated by altitude for analysis.
- Check for BAe-146 data.
- DC-8 data often reports uncertainties for 1 sec data well over 100%. Panel should state, based on IEIP analysis, whether uncertainties are overstated for DC-8 data.

- Detection limits for both instruments should be determined from the data and given to the modelers for their analysis.

SO_2 (Leads: T. Ryerson, G. Huey)

- The comparison results varied considerably: the first and third comparisons showed agreement within the PI reported uncertainties, but the second comparison indicated DC-8 observations were ~75% higher. Standards were exchanged and no problem was found. Both instruments may have problems with rapid changes in humidity. The panel cannot determine the source of the bias (see slope regression analysis) and will document this difference for future panel assessments (e.g. MILAGRO, ARCTAS data).

- Panel recognizes the need to further investigate difference in detection limits especially on 1 s data.

- Adjust uncertainties to encompass both measurements.

- Investigate to determine if the H_2O correction of the CIMS data could be a source of the disagreement

- Compare ambient SO_2 data collected in a power plant plume by aircraft with CIMS data collected at the power plant to see if robust tests are possible.

HNO_3 (Leads: J. Dibb, T. Ryerson)

- The panel recognizes the challenges in making this measurement and that significant progress has been made in the last decade.

- Significant differences exist between the two instruments on the DC-8 at times and larger differences were found between the DC-8 and the WP-3D data. However, the panel recommends retaining all data for model assessment.

- Significant adjustment in uncertainties will be necessary.

- The panel may not be able to reconcile the differences between measurements but will investigate the nitrate issue (evaporation) to determine if it explains the discrepancy.

CH_2O (Leads: F. Flocke, E. Apel, M. Evans)

- WP-3D data has large fluctuations, larger than that seen on DC-8.

- Add comparison with GTLIF.

- BAe-146 comparison limited in range.
- Quantify BAe-146 results in terms of x% lower than DC-8 TDL and y% higher than DC-8 EFD.

VOCs (Leads: D. Blake, E. Apel, M. Evans, D. Parrish)

- The first TAbMEP considered only ethane, methane, propane, n-butane, ethyne, benzene, and toluene. Ethene analysis deferred until next year.
- Examine issue with BAe-146 butane data.
- Redo comparison to capture all overlap in measurement integration times.
- Based upon test results, may allow for non-overlapping measurements to be considered a 'match' if the time between the two time periods is sufficiently small.
- It may be useful to analyze VOCs vs. CO on respective planes since the sample times do not overlap very well.
- Perform IEIP analysis on PTRMS data.
- For certain species with known systematic differences, the associated error bars may be adjusted asymmetrically (e.g.,propane).
- Review results of canister exchange.
- Include best practice procedures for conditioning cans in final report.
- Investigate relationships between simultaneously measured VOCs in the complete data sets collected on individual platforms. There are some useful robust relationships between VOCs that should be present in all data sets. Deviations from those relationships are important indicators of measurement problems.

NO (Leads: G. Huey, T. Ryerson)

- Complete full analysis (Pippin).

PAN (Leads: F. Flocke, G. Huey, M. Evans)

- Intercomparison results show general agreement. The lead facilitator will provide additional in depth interpretation of the comparison results.

H_2O (Leads: G. Diskin, T. Ryerson)

- May need to increase stated uncertainty on DLH to 10% above some H_2O concentration (need to look at data to determine where).

- Panel recognizes DLH represents the best H_2O measurement in terms of time resolution.

- Larger uncertainty adjustment may be needed for the WP-3D data.

- Modelers need water vapor and relative humidity.

Aerosol number density and size distribution (Leads: B. Anderson, C. Brock)

- The comparison showed better agreement for integrated quantities than size distributions.

- Modest adjustment is needed for number density uncertainty.

- Larger adjustment is needed for volume density uncertainties.

- Panel recommends archiving data consistently across all instruments (i.e., DMA, OPC, APS).

Particle chemical composition (Leads: J. Dibb, J. Jimenez, T. Quinn)

- The first TAbMEP considered only SO_4^{-2}, NO_3^-, and NH_4^+.

- LOD may not be adequate for AMS NO_3^-.

- SO_4^{-2} and NH_4^+ need larger error bars.

- Modelers still want this information, even with large uncertainties.

- NH_4^+ needs more analysis (DC-8 to DC-8 and WP-3D to WP-3D comparisons) to find out if it is suitable for model assessment.

Optical properties (Leads: B. Anderson, C. Brock)

- Modelers want extinction and absorption information.

- If possible intercompare with J31.

- Carry out IEIP analysis.

Appendix A. List of Attendees

Attendees	Contributions	Affiliation	E-mail	Phone
Bruce Anderson	Aerosol Measurements	NASA LaRC	bruce.e.anderson@nasa.gov	757-864-5850
Eric Apel	Trace Gas Measurements	NCAR	apel@ucar.edu	303-497-1452
Melody Avery	Trace Gas Measurements	NASA LaRC	melody.a.avery@nasa.gov	757-864-5522
Steve Arnold	Global & Reg. Model: Trace Gas	Univ. of Leeds	s.arnold@see.leeds.ac.uk	+44 (0) 113 343 7245
Don Blake	Trace Gas Measurements	Univ. of CA, Irvine	drblake@uci.edu	949-824-4195
Chuck Brock	Aerosol Measurements	NOAA/ESRL	charles.a.brock@noaa.gov	303-497-3795
Greg Carmichael	Reg. Model: Trace Gas & Aerosol	Univ. of IA	gcarmich@engineering.uiowa.edu	319-335-5191
Gao Chen	Organizer, data analysis	NASA LaRC	gao.chen@nasa.gov	757-864-2290
Mian Chin	Global Model: Aerosols	NASA GSFC	mian.chin-1@nasa.gov	301 614-6007
Jack Dibb	Trace Gas Measurements, Aerosol Measurements	Univ. of NH	jack.dibb@unh.edu	603-862-3603
Glenn Diskin	Trace Gas Measurements	NASA LaRC	glenn.s.diskin@nasa.gov	757-864-6268
Louisa Emmons	Global Model: Trace Gas	NCAR	emmons@ucar.edu	303-497-1491
Mat Evans	Global & Reg. Model: Trace Gas	Univ. of Leeds	mat@env.leeds.ac.uk	+44 (0) 113 343 1594
Arlene Fiore	Global Model: Trace Gas	NOAA/GFDL	Arlene.Fiore@noaa.gov	609-452-6525
Frank Flocke	Trace Gas Measurements	NCAR	ffl@ucar.edu	303-497-1457
Jose Jimenez	Aerosol Measurements	Univ. of CO	jose.jimenez@colorado.edu	303-492-3557
Terry Keating	HTAP & EPA Representative	EPA	keating.terry@epa.gov	202-564-1174
Mary Kleb	Organizer, data analysis	NASA LaRC	mary.m.kleb@nasa.gov	757-864-5816
Qing Liang	Global Model: Trace Gas	NASA GSFC	qing.liang-1@nasa.gov	301 614-5993
David McCabe	EPA Representative	AAAS/EPA	McCabe.David@epa.gov	202-564-0016
Pete Parker	Statistician	NASA LaRC	peter.a.parker@nasa.gov	757-864-4709
David Parrish	Trace Gas Measurements	NOAA/ESRL	David.D.Parrish@noaa.gov	303-497-5274
Margaret Pippin	Organizer, data analysis	NASA LaRC	m.pippin@nasa.gov	757-864-9366
Tom Ryerson	Trace Gas Measurements	NOAA/ESRL	thomas.b.ryerson@noaa.gov	303-497-7531
Ariel Stein	Global Model: Aerosols	NOAA/ARL	ariel.stein@noaa.gov	301-713-0295 x119
Jian Wang	Aerosol Measurements	DOE/BNL	jian@bnl.gov	631-344-7920
Margie Hall	Logistics Support	SSAI	Marjorie.F.Hall@nasa.gov	757-951-1607
Absent Panel Members				
Greg Huey	Trace Gas Measurements	GA Tech	greg.huey@eas.gatech.edu	404-894-5541
Trish Quinn	Aerosol Measurements	NOAA/PMEL	patricia.k.quinn@noaa.gov	206-526-6892
Michael Schulz	Global Model: Aerosols	LSCE	michael.schulz@cea.fr	01-69-08-77-31

Appendix B. Sample of the Assessment Summary Table.

Species					
Campaign/Year					
Airborne Platform					
Instrument					
PI/Institution					
Random Uncertainty	**PI**				
	Panel				
Systematic Uncertainty	**PI**				
	Panel				
Total Uncertainty	**PI**				
	Panel				
Comparison Results		NASA DC-8	NOAA WP-3D	FAAM BAE-146	DLR Falcon
NASA DC-8		N/A			
NOAA WP-3D			N/A		
FAAM BAE-146				N/A	
DLR Falcon					N/A
Panel Recommendation					
Panel Notes					

Appendix C. 1st TAbMEP Meeting Recommended Statistical Data Analysis Procedures

Note these recommended procedures are subject to further refinement as the LaRC MEaSUREs project progresses and issues arise.

I. Internal Estimate of Instrument Precision (IEIP).

IEIP is an objective and data-driven approach to assess absolute and/or relative instrument precisions. This approach directly estimates, under a few assumptions, the instrument precision, which is the variance over a small time interval, Δt.

For species x, the total variance can be expressed as: $V[x] = \sigma^2_x + \sigma^2_{\varepsilon\text{-}x}$

where $\sigma^2_{\varepsilon\text{-}x}$ is the random instrument variability which corresponds to instrument precision for measurement of species x and σ^2_x represents the natural ambient variability.

V(x) can be a reasonable estimate for $\sigma^2_{\varepsilon\text{-}x}$ if Δt is small enough such that σ^2_x is negligible compared to $\sigma^2_{\varepsilon\text{-}x}$. At the same time, Δt must be large enough to minimize the effect of autocorrelation. $\sigma^2_{\varepsilon\text{-}x}$ can be assessed by following procedures listed below:

- Compute standard deviation over Δt and generate frequency distribution or histograms.
- Vary Δt and repeat the previous step, then look for the values of the modes, which should be relatively constant over a limited range of Δt values.
- How long should Δt be? In principle, it should be long enough to overcome any significant autocorrelation impact and short enough such that σ^2_x is negligible.
    - depends on the temporal and spatial variability of the species or parameter of interest.
    - depends on instrument sampling rate.
    - requires expert judgment.

The IEIP analysis is typically applied over an entire flight and/or a large segment of data with fairly constant values. It should be noted that this approach may or may not be feasible for the measurements with long integration times and/or significant gaps between the data points. The IEIP analysis may also be problematic when measurement precision is strongly dependent on the ambient values.

<u>IEIP Example: CO instrument precision assessment</u>

Figure 1 shows an example of IEIP assessment for CO for both relative and absolute uncertainties. Note that the modes of the distributions (i.e., the location of the peaks)

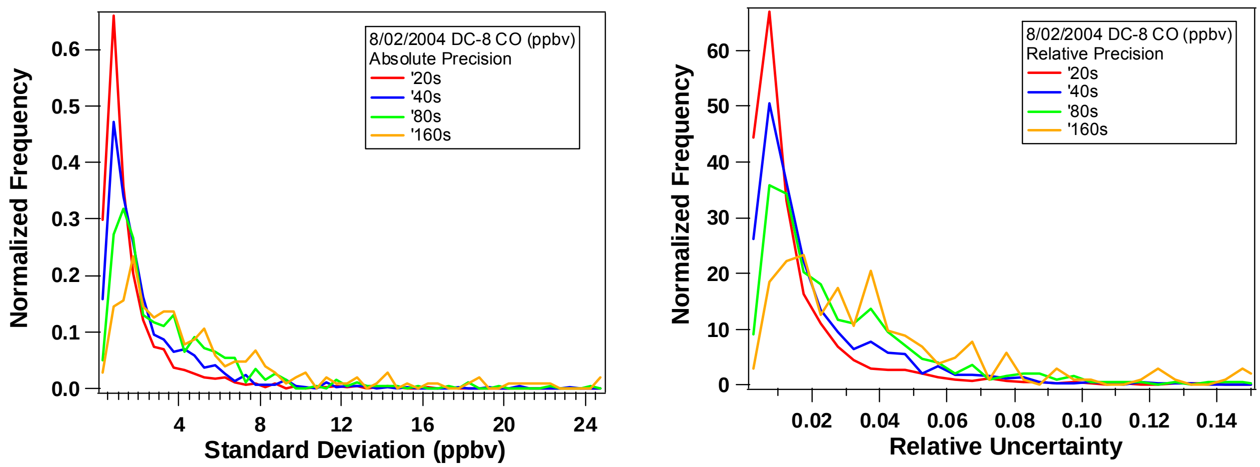

Figure 1. Example of IEIP analysis of NASA DC-8 CO observations during INTEX-A/ICARTT.

are relatively constant over the range of Δt from 20-40 seconds. The standard deviation increases with longer Δt times, which is likely due to the CO natural variability. The resulting relative uncertainty for this DC-8 flight is about 0.7% and absolute uncertainty is 0.7 ppbv.

II. Assessment of Measurement Consistency between two Instruments.

Several approaches were discussed during the first TAbMEP meeting to better quantify the difference between the measurements for the intercomparison time period:

- Statistics based on the absolute difference Δx (and/or relative difference, i.e., ratio Rx = x1/x2) between two measurements, alternatively these quantities may be averaged over a small time interval, Δt, which are noted here as Δx_{avg} and Rx_{avg} (i.e., $x1_{avg}/x2_{avg}$)
    - ODR fit between x1 and x2 or $x1_{\_avg}$ and $x2_{\_avg}$ either weighted or not weighted by the random uncertainties.
    - distribution of Δx and/or Rx or Δx_{avg} and/or Rx_{avg}
    - examine the dependence of Δx and/or Rx or Δx_{avg} and/or Rx_{avg}.
    - examine the dependence of Δx and/or Rx or Δx_{avg} and/or Rx_{avg} on other atmospheric conditions, e.g., alt, temperature, etc.
- ANOVA analysis to examine if there is a significant difference in comparison results from flight-to-flight.

III. Approach to Unify Data Sets from Multiple Measurements.

The panel recommended approach for unifying data will not adjust the PI data, but the uncertainties associated with the data will be expanded so that at least 95% of points from the multiple measurements will overlap each other within 2σ uncertainties. The panel also noted that the unified data sets include only those rated by TAbMEP panel as suitable for model assessment. A equitable, objective, and data-driven approach was developed by P. Parker to assign panel assessed uncertainty to each of measurements. The uncertainty adjustment is based on intercomparison results and overlapping data during the intercomparison period. The procedure involves three steps:

1. Calculate average $x_{avg}(j)$ of the measurements for each of the overlapping data point for all instruments involved.
   a. arithmetic average or
   b. weighted average

2. Estimate intercomparison uncertainty.
   The total intercomparison uncertainty associated with the measurement of species/parameter x by the i^{th} instrument can be derived from intercomparison data with n overlapping points. The panel recommends the following set of equations to determine the magnitude of the total intercomparison uncertainty:

$$Int\_Comp\_unc_i = |\bar{d_\iota}| + \sigma_{di} \cdot t_{0.025,n-1}$$

$$\bar{d_\iota} = \frac{1}{n}\sum_{j=1}^{n}(x_i(j) - x_{avg}(j))$$

$$\sigma_{di} = \sqrt{\frac{\sum_{j=1}^{n}\left(x_i(j) - x_{avg}(j)\right)^2}{n-1}}$$

   where $Int\_Comp\_unc_i$ is the estimated total intercomparison uncertainty; $\bar{d_\iota}$ and σ_{di} are the average and standard deviation of the difference between individual points from the i^{th} instrument and x_{avg}; $t_{0.025,\ n-1}$ is the student-t value for 2.5 percentile and n - 1 degrees of freedom (
   n = total number of overlapping points involved in the analysis). The value of $t_{0.05,\ n-1}$ is approximately 2 when n is larger than 30.

   Determine the overall uncertainty for each of the panel validated measurements such that:
   a. If the PI reported total uncertainty is larger than the Int_Comp_Unc, then no adjustment needed.
   b. If quadratic sum of Int_Comp_Unc and IEIP is larger than the PI reported total uncertainty, then the quadratic sum would be assigned as over all measurement uncertainty.

Using the above approach, the overall uncertainty values will be assessed for each of the measurements involved, based on the intercomparison data. These uncertainty values will be applied throughout the datasets so that the all measurements will be consistent with the combined 2σ uncertainties.