ICARTT measurement comparisons

Final implementation plan

(5/27/2004)

<u>Introduction</u>

Goal. The goal of comparison exercises planned for the 2004 ICARTT campaign is to create a unified observational data set from measurements acquired from multiple aircraft, ground, and ship platforms. To achieve this goal, comparisons are planned to help establish data comparability between the various platforms, and to verify that different analytical approaches are mutually consistent within quantifiable uncertainties. Planned measurements include a wide variety of *in situ* and remotely sensed gas-phase chemical species, aerosol chemical and physical data, radiative effects, and meteorological parameters. These data will be acquired using a variety of techniques, each with specific instrumental accuracy and precision. Quantifying data uncertainty establishes an objective basis upon which subsequent scientific interpretation can be founded.

Scope. This effort requires coordination between the multiple participating organizations of ICARTT, and will primarily involve side-by-side measurement opportunities between combinations of aircraft, ship, and ground stations located in and between North America and Europe. In particular, comparison opportunities are planned that will link the platforms participating in the ITCT-Lagrangian-2k4 task of IGAC. While further comparisons of these data sets to satellite retrievals and model output are equally important, such analyses will involve the entire 2004 data set and will be carried out primarily by other ICARTT working groups. This document describes the protocol for acquiring, evaluating, and disseminating the results of side-by-side data comparison activities for all participating platforms exclusive of satellite and model data.

Organization and formality. A small working group, with one representative from each major participating organization, has been identified; a list of delegates is available here. This group will be responsible for developing comparison strategies, will act as referees, and will attend to the logistical details required for the comparisons. However, this group will solicit input, suggestions, and guidance from all participants in the 2004 field campaign, and the active participation of interested parties is greatly encouraged. Close cooperation with the Aircraft and Ship working group is also planned to best integrate the field comparison exercises into other research goals of ICARTT and with the science plans of individual participating organizations.

The comparisons are envisioned as semi-formal exercises, which can be used in part to help identify any recoverable errors in time to correct them during the field campaign. For the field comparisons, “field-quality” data accompanied by estimated uncertainties will be submitted independently to the working group. A goal for data turnaround of 24 hours after the comparison exercise has been set; this goal can be relaxed to accommodate the exigencies of field operations. Data from instruments utilizing a post-flight analysis step, e.g., GC measurements of whole-air canister samples, are not typically readily available on these timescales in the field. These data will be compared in the same fashion, but paced by the normal data turnaround rate for these instruments.

After all readily available data for a given comparison are submitted (ideally within 24 hours) the flight data will be made accessible to all study participants. This provides for an “informal, but blind” comparison process, agreed upon by ICARTT study participants. The following day, the ICARTT comparison data manager (Gao Chen, from NASA Langley Research Center) will also post the comparison data in graphical form – a time series and an x-y plot for each measurement – to facilitate their comparison. (Details of the comparison data exchange procedure are outlined in **2. Field Campaign**: *Data Exchange and Availability* section, later in this document).

These comparison plots will be updated post-campaign as the data sets and their associated uncertainties are refined. Finally, the working group, in conjunction with the measurement PIs, will draw consensus conclusions from the final data sets regarding the comparability of the measurements. These final side-by-side data, plots, and conclusions will be posted on a public area of the ICARTT measurement comparison web site.

Outline. Three phases are loosely defined for planning purposes: pre-campaign (Fall 2003 through Spring 2004), field campaign (Summer 2004), and post-campaign (Fall 2004 through Spring 2005). Working group tasks during the pre-campaign phase include exchange of standards and coordination of ground comparisons of instrumentation, where possible. During the summer 2004 field campaign, multiple comparisons between the platforms will be carried out, preliminary data exchanged and evaluated, and the comparison results posted (password-protected, but accessible to all study participants) on the ICARTT web site. Post-campaign tasks will include analysis of the final data sets and assessment of the comparability of data from the different platforms. Dissemination of results of these comparison exercises will include posting of the final comparison data and analyses in a public area of the web site, as well as a presentation of the summarized results at the data workshop planned for April 2005. Details of the planned tasks for each of these three phases are given below.

Measurement comparison tasks.

1. Pre-campaign

A. *Standards exchange*. Exchange of standards is planned to aid in harmonizing instrument calibrations across the study platforms. These are offered as aids to help put instrument calibration on a common basis; we encourage participants to take advantage of these if it would be useful to you. If the timing, logistics, or other factors make sampling from these standards a burden, however, there is no requirement to participate in this standards exchange, and there is certainly no penalty for not doing so.

Several kinds of standards are available and their uses are described below.

• Shippable standards: NOAA-AL is providing certified, high-pressure, low-ppmv-level standard compressed gas mixtures of NO, SO_2 + CO, and CO_2 (each with an associated regulator), to participating investigators. Eric Apel of NCAR-ACD has donated a VOC transfer standard containing low-ppmv levels of the following compounds: methane, ethane, ethene, acetylene, propane, propene, butane, benzene, toluene, acetone, acetonitrile, isopropyl nitrate, HFC-134a, CFC-113, CCl_4, and CO. Interested parties should contact Eric Williams (eric.j.williams@noaa.gov) at NOAA-AL to arrange scheduling of these compressed gas cylinder shipments.

For lower-level VOC standards exchange, including whole-air samples, Elliot Atlas (eatlas@rsmas.miami.edu) and Don Blake (drblake@uci.edu) have offered to prepare and circulate exchange cylinders; please contact them directly to arrange shipping of these VOC standard cylinders.

A short turnaround period, ca. 1 week, is requested of each investigator to permit all groups to have an opportunity to compare these transfer standards to their own in-house calibration standards. Currently it is planned to have a single set of standards serve for both the North American and European contingents. If international shipping time and cost is prohibitive, a separate set of tanks might be circulated between the European groups. Trish Quinn of NOAA-PMEL has volunteered to provide liquid standards for detector calibration of PILS-IC and filter measurements of soluble inorganic ions on aerosol particles. Groups interested in obtaining liquid standards of, e.g., Na, NH_4, K, Mg, Ca, MSA, Cl, Br, NO_3, and/or SO_4 should contact Trish directly at Patricia.K.Quinn@noaa.gov.

• Non-shippable or developmental standards: Creating and delivering known amounts of other chemical and aerosol species has been demonstrated, but these typically remain research-grade devices requiring an experienced operator. Some advance coordination and planning will likely be required to successfully and meaningfully interface these new calibration devices with different instruments. We provide a partial list below of calibration devices that have been offered to be made available to other interested participants. In many cases, the easiest opportunity for sampling from these devices may come during the field campaign phase. Please contact the PIs listed below for more details and to organize an opportunity to sample from these standards.

Species	Contact	Email
Aerosol number, size, and chemical composition	C. Brock	charles.a.brock@noaa.gov
HNO_3	A. Neuman	john.a.neuman@noaa.gov
NH_3	J. Nowak	john.b.nowak@noaa.gov
HO_x	W. Brune, C.Cantrell	brune@essc.psu.edu cantrell@ucar.edu

• Centralized national and international calibration facilities: Accepted central facilities exist to calibrate or evaluate measurements of, e.g., CO_2, O_3, and actinic flux. Some research groups already reference their CO_2 standards to the NOAA-CMDL scale. Many O_3 measurements are based on UV absorption, which as a primary measurement cannot be calibrated; however, national facilities often provide a reference measurement against which the output of field instruments can be compared. While O_3 reference instruments and standard-output lamps are potentially transportable, we refer the individual investigators to the existing national and international calibration facilities for these reference standards.

B. *Direct comparison of measurements*. Running instruments from different groups side-by-side in the laboratory or in a field setting is an excellent way to test instrument performance before the 2004 summer field campaign. Because of the logistics and time involved, this is more easily done for some instruments than for others; this sort of comparison will be left up to the various investigators to arrange as possible.

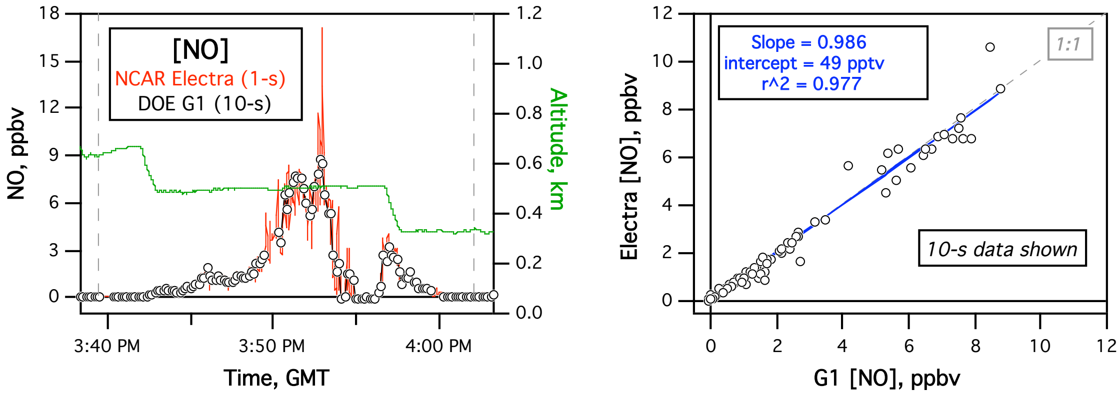
C. *Sampling coordination and planning*. Coordinating the sampling details, where possible, of study instrumentation may substantially improve the comparability of ambient data from different platforms. For example, small differences in inlet transmission as a function of aerosol size, and especially of relative humidity at the sampling point, can potentially affect data from otherwise identical instruments. Further, tabulating instrumental sampling conditions can help to understand potential differences between *in situ* and remotely-sensed aerosol properties. Knowledge of instrumental time response may also be useful in comparing gas-phase chemical data between platforms. For non-continuous gas chromatographic (GC) measurements or whole-air canister sampling, synchronizing sample times (at least for the duration of the comparison periods) will substantially improve data overlap.

The primary constraint on sampling details and timing will certainly be the science goals determined by each participating investigator and organization. However, prior coordination of, e.g., aerosol size cuts, between ICARTT platforms may substantially enhance the utility of the combined data set while still fulfilling individual science goals. For example, the various groups measuring aerosol number, size, chemical composition, and optical properties aboard the NOAA WP-3D and the NOAA ship *Ronald H. Brown* have agreed on a 1.0-micron aerodynamic cut-off diameter to separate accumulation and coarse mode particles to facilitate combining data sets from several of their instruments. To facilitate knowledgeable instrument comparisons, and coordinate sampling details where possible, we'll draft and circulate *brief* instrument questionnaires to all instrument Pis before the summer field phase of the joint missions.

2. Field campaign

• *Generating comparable data.* Data taken during wingtip-to-wingtip aircraft flight legs, or low-level aircraft overflight of ground or ship locations, can permit a direct comparison of instrument performance. Ideally, ambient levels are encountered that test each instrument over a wide range of parameters, e.g., mixing ratio, altitude, water vapor, and potential interferences. Prior experience in comparing continuous, fast-response gas-phase instrumentation suggests these criteria can often be met by spending between 15-30 minutes in level flight at different altitudes, e.g., one in the clean free troposphere and one in the more polluted continental boundary layer.

An example below shows quantitative agreement for NO data taken by two aircraft flying in formation in the Houston metropolitan area in September 2000, sampling over a wide range of ambient parameters. While altitude changes were small, this comparison flight leg (data between the vertical dashed lines) sampled the clean free troposphere, a polluted urban and industrial plume, and the clean marine boundary layer, all within 20 minutes. Despite very high spatial variability of ambient NO mixing ratios, both aircraft were clearly sampling the same air masses at the same time, suggesting that quantitative comparison of these and other data was warranted.



Data that overlap but are taken at different time resolutions will be need to be averaged over comparable periods before a comparison can be meaningfully made. Data from instruments with widely varying time resolution may still be comparable if the two platforms can be shown to have sampled from the same air mass(es) for the duration of the comparison datum. However, for data generated from an aircraft overflight of a ship or ground site, usefully comparing measurements of vastly different time resolution (e.g., seconds vs. hours) may not be possible.

Other evaluations are also planned using data from side-by-side flights and overflights. Examples of these might include evaluating the NO_y budget by comparing the sum of measured constituent species ($NO+NO_2$+PANs+$HNO_3+NO_3+N_2O_5$+aerosol nitrate) to the NO_y measurements, comparing measured aerosol optical depth to that inferred from a vertical profile of *in situ* aerosol optical data, and comparing ozone profile measurements from LIDARs or balloonsondes to a vertical profile generated from *in situ* ozone instruments. Certain assumptions need to be satisfied for these kinds of these comparisons to be valid; these assumptions will be taken into account in designing the comparison flight legs and in the subsequent interpretation of the data.

• *Comparison flight planning*. Comparison flight planning requires consideration of a complex function of individual program requirements, aircraft flight envelopes, air traffic control restrictions, weather, instrument readiness, and scientist and flight crew coordination. As these are constantly changing parameters during any field campaign, some details and actual comparison flight dates will best be decided in the field, in conjunction with the Aircraft and Ship Coordination group, and with the individual mission scientists from each organization.

Some comparisons should be conducted as soon as practical in the mission, so that any recoverable problems can be identified and addressed early on. The main requirement for these early comparisons is that the instruments be tested previously in flight and be working properly. Comparisons throughout the rest of the mission are useful for confirming instrument calibration stability and for comparing in a wider range of environmental conditions.

Comparison flights will take proportionally more or less of an individual science flight depending on individual aircraft endurance. In the past it has often been possible to include comparison legs as an organic part of flight plans addressing other science issues. For example, for a coordinated East Coast regional survey jointly involving the NOAA WP-3D flying from Portsmouth, NH and the DOE G1 from Latrobe, PA, a comparison might easily take place by the aircraft joining up on the westernmost leg of the WP-3D flight and the easternmost leg of the G1 flight.

Ultimately the comparisons are limited to overlapping deployment periods (see the ICARTT deployment schedule), so the scheduling of some pairings may be more flexible than for others. Past experience has shown that longer-endurance aircraft may execute more than one side-by-side comparison exercise during a given flight, but that comes with the additional planning requirements for smooth execution by more than two platforms.

• *Proposed comparisons*. A matrix of comparison flights is proposed to best link measurements between the various aircraft, ship, and ground-based sites and groups participating in the 2004 campaign. Particular importance is given to linking the measurements between the heavy aircraft, ship, and ground sites participating in the

ITCT-Lagrangian-2k4 task. While it will be advantageous to repeat any given comparison, time and logistical constraints may dictate only the most important linking comparisons can be repeated. Extra consideration may be given to repeating a comparison flight if, on the first try, any substantial disagreements are noted that can be effectively addressed in the interim by the investigators. The proposed comparisons include the following pairs, which are also presented graphically in Appendix 1:

aircraft/aircraft:

Navy Twin Otter and MSC Convair	DOE G1 and Navy Twin Otter
COBRA King Air and NOAA WP-3D	NOAA WP-3D and DOE G1
DOE G1 and UMD Duchess	NASA Jetstream-31 and MSC Convair
NOAA DC-3 and NASA DC-8	NASA DC-8 and FAAM BAe-146
NASA DC-8 and NOAA WP-3D	NOAA WP-3D and MSC Convair
FAAM BAe-146 and DLR Falcon	DLR Falcon and CNRS Falcon

aircraft/ship:

NOAA WP-3D and NOAA *Ron Brown*	NOAA DC-3 and NOAA *Ron Brown*
NASA Jetstream-31 and NOAA *Ron Brown*	

aircraft/ground site:

FAAM BAe-146 and Pico, Azores	COBRA King Air and flux tower
NOAA WP-3D and Castle Springs, NH	(NOAA *Ron Brown* and Chebogue Pt.)
NOAA WP-3D and Harvard Forest, MA	

• *Quantifying the comparisons*. Putting the comparisons on an objective, quantitative basis will require the data be accompanied by uncertainty estimates. For the 24-hour data turnaround planned for the comparison exercises, it is recognized that the data will not have been subjected to the full quality checking that characterizes a final data set. Estimated uncertainties will be correspondingly larger for many, if not all, of these quick-look, “field-grade” data. Nonetheless, to quantify the degree of data agreement, uncertainty estimates are required to determine if any observed departures from fitted slopes of 1.0 and intercepts of 0.0 are consistent within the known errors, or lie outside the known errors and are indicative of one or more instrumental issues. This will facilitate one goal of the comparison exercise, to use the comparisons to identify potentially recoverable problems (leaks, calibration offsets, electrical noise issues) in time to address them during the field campaign.

To accomplish this, the working group will require that an estimate of data precision and accuracy (or of total combined uncertainty) to be submitted along with the data within 24 hours of the comparison exercise. As the data sets become finalized in the months after the summer 2004 campaign, it is expected that the data and the corresponding uncertainty estimates will change as well. The working group will ensure that the comparison data will be updated in a timely fashion to reflect these changes.

- *Comparison data exchange and availability*. The ICARTT Data Management working group has agreed on a common format, generally based on the NASA-Ames standard, for the final data. We will use this ICARTT format for the comparison data submission as well. While this may require some additional programming work up-front for first-time users, it will substantially streamline the data exchange process once the necessary procedures have been worked out. There is sufficient experience with this format amongst the ICARTT community that we can offer guidance on its use and in automating individual groups' data output to conform to this format. Please contact the Data Management working group for tools and software support for this new ICARTT format.

- *Comparison data flow*. The comparison exercise will not impede the normal and timely turnaround of aircraft data necessary for flight planning and Lagrangian forecasting. Normal field-grade data exchange and posting (data "flow") for a given measurement platform is prescribed by the Data Management working group as follows:

Normal data flow for a given platform:

1. PIs → Data manager → public data sites (web, ftp); ~ 24-hr turnaround

A slightly modified data flow will accommodate the informally blind comparison exercise:

Comparison data flow:

1. PIs → Data manager → platform-specific folders on Comparison ftp site
2. Gao Chen determines that relevant data are all submitted; emails managers
3. Data managers → public data sites (web, ftp); ~ 24-hr turnaround
4. Gao subsequently posts time-series and x-y plots of comparison data

Note that the 4-step comparison data flow still allows the release of aircraft data on the same 24-hour nominal schedule as the normal data flow, if all the comparison data are posted on schedule. If one or more data sets are delayed, Gao will have discretion to either continue to temporarily embargo all the data beyond 24 hours (to maintain the "informal, but blind" aspect of the comparison exercise), or to decide to release the available data and note the comparison of the delayed data set was not necessarily blind in this instance.

Exceptions to this comparison process may be necessary for optimal forecasting of transatlantic Lagrangian opportunities. In these cases, if data availability might otherwise be delayed, Gao may provide a forecast-critical subset of aircraft data (Time, aircraft position, ambient pressure, and [CO]) to the Lagrangian planning team.

Once all the field-quality data for a given comparison exercise have been submitted, these plots will be posted to a password-protected part of the ICARTT comparison web site (accessible here). All study participants will have access to this site; please contact the webmaster for user name and password information.

3. Post-campaign

After the mission, the working group will ensure that the comparisons plots are updated as final data become available. Following the Data Exchange WG suggestion, we will note whether a given comparison uses the initially submitted ("field-grade") data, or those from subsequent revisions ("preliminary"), up to and including the final data revisions. These final data will be similarly presented and posted, with consensus PI and working group conclusions on the degree of comparability, in a public area of the ICARTT measurement comparison page. Four possible conclusions are anticipated:

1. paired measurements agree quantitatively within stated uncertainties of xx%
2. paired measurements show significant differences, but were reconciled by the following means (sampling regimens differ, inlet effects identified, issues with calibration or data reduction for one or both instruments, etc. Note any adjusted uncertainty estimates for final data)
3. paired measurements show significant differences but are not reconciled. If possible, justify choice of one data set over the other, or provide consensus caveats on both, for final data usage.
4. comparison judged not to be a valid test (instrument malfunction, aircraft overflight did not sample surface layer, spatial inhomogeneity too great, etc.)

The co-chairs will present a short summary of comparison exercises at the data workshop scheduled for April 2005. Finally, if data that have been compared during the joint ICARTT 2004 joint campaign are used in publication, participants have agreed to note that a comparison was carried out and briefly state the results thereof.