SGP Site Scientist Team Data Quality Assessment Activities

R. A. Peppler and M. E. Splitt
Cooperative Institute for Mesoscale Meteorological Studies
University of Oklahoma
Norman, Oklahoma

Introduction

A primary task of the Site Scientist Team (SST) of the Southern Great Plains (SGP) CART site is to analyze and assess the quality of the various SGP data streams. This task is performed in conjunction with other groups within ARM such as Instrument Mentors (responsible for individual instruments), various Data and Science Integration Team (DSIT) members (responsible for multiple data stream comparisons such as (Quality Measurement Experiments [QMEs] and Value-Added Procedures [VAPs]), and Site Operations (perform daily existence and instrument performance checks on all incoming data streams).

The SST currently uses two tools in its efforts to routinely assess data quality—data quality graphical displays and performance metrics. These tools, plus information provided by other data quality assessment participants, are primarily used to produce periodic reports on data quality and issue them to key persons within the ARM infrastructure. Reports were initially limited to the Solar and Infrared Observing System (SIROS) platform (though many other data streams are routinely analyzed by the SST), but were expanded in April 1997 to include all data streams the SST is routinely assessing. These can be obtained via e-mail by contacting the lead author.

The SST also periodically reviews and discusses calibration and maintenance issues related to SGP instrumentation, representing a view of instrument performance.

The SST uses the tools they have developed and the information provided by others to them to help assess the quality of the data produced by the SGP sensors, through communication of what is observed back to Instrument Mentors and Site Operations, to effect corrective actions if necessary.

This paper briefly reviews the SST's data quality assessment activities and where they are headed.

Data Quality Assessment Tools

Following is a discussion of data quality performance metrics, data quality graphical displays, and calibration and maintenance information summaries.

Data Quality Performance Metrics

In an effort to develop a systematic, integrated view of the quality of the SGP data streams over the long-term, the SST in February 1996 began routine monitoring of SGP instruments through the creation and analysis of data quality performance metrics. These metrics determine the percentage of data values which fall within specified (by Instrument Mentors, factory recommendations, SST, etc.) quality tolerances. The integration of these metrics will provide an estimate of the overall health of SGP instrumentation, and will help determine to what degree the data and instrumentation are satisfying the scientific goals of the ARM Program.

The basic form of a metric is:

F(x) = Fraction of data delivered by an instrument that falls within specified quality tolerance(s)

where x is a given data stream for a given platform, and F is made up of various sub-components (if necessary), or sub-metrics, based on the factors that are known to govern the quality of data from a particular instrument.

The main task here of the SST is to best determine F(x) for each data stream it analyzes. The sub-metrics of F(x) include, but may not be limited to, data existence, range (min/max), consistency (delta), and intercomparison checks. Data quality itself can be affected by many things, such as intrinsic instrument precision, calibration, operation and maintenance, placement, data sampling strategy, communications medium, and data ingest method.

Metrics are currently divided into those categorizing existence (were data actually collected?) and quality (are the data collected any good?). A data existence metric will establish the upper baseline for any data quality metric. The data existence metric is simply the fraction of data actually collected when compared to what was expected. These are being computed now.

The development of the data quality metrics is an evolutionary process and will be geared to never-ending quality improvement. It is expected that each F(x) for a given platform will be built over time with the refinement and addition of sub-metrics. The most basic single data stream sub-metrics being calculated include range and consistency checks (for many instruments). In addition, more sophisticated sub-metrics involving multiple instrument intercomparisons and objective analyses of multiple-site instrumentation are also being computed for some data streams (e.g., Energy Balance Bowen Ratio [EBBR] and SIROS net radiation; SIROS/Baseline Surface Radiation Network (BSRN) comparisons; multiple-site Surface Meteorological Observing Station (SMOS) and EBBR)

As F(x) increases, tolerances may subsequently be tightened to reduce F(x) and reinitiate the process toward further improvement of quality (increase F[x]).

Establishment of F(x) for each instrument platform is the important first step in determining the overall health of the SGP instrumentation. Trends in F(x) will be an important tool in observing changes in the performance of a given instrument and may lead to problem recognition and resolution. Critical components in establishing the metrics will include specification of acceptable tolerances and then creation of appropriate tests (metrics) to evaluate whether the data fall within those bounds. The Instrument Mentors and DSIT will help guide the SST in specifying the tests to be administered, the display of the results, and the prioritization of the order in which instruments are treated. The latter will presumably follow the prioritization of SGP issues based on ARM's overall goals/needs.

At present, performance metrics are being automatically produced on a daily basis for the following individual and multiple instrument platforms:

Individual Instrument Platforms: Microwave Water Radiometer (MWR), Balloon-Borne Sounding System (BBSS), EBBR, SIROS, 915-MHZ Radar Wind Profiler (RWP915).

Multiple Instrument Platforms: EBBR/SIROS net radiation, SIROS/BSRN component comparisons at the Central Facility, surface objective analysis of EBBR and SMOS components, shortwave radiation ensemble at the Central Facility.

A prototype world wide web (WWW) page, developed by the SST for graphically displaying the metrics, is being replaced by a more real-time WWW area spearheaded by the DSIT.

Data Quality Graphical Displays ("Modules")

Data quality graphical displays, or "modules," are intended to provide a tool for conducting "first line of defense" quality control. The modules will aid in identifying problems in instrument performance via daily analysis of the data streams as they arrive from the instruments.

The modules act as a key supplement to the performance metrics. The metrics may not be able to identify all problems since some undesirable instrument performance may translate to data which fit within a particular metric's tolerance. The human eye, when looking at a plot of actual data, can identify problems that are not easily caught by a computer algorithm. A graphical display will also allow identification of an individual data stream that is of concern, which may or may not be the case for a performance metric.

The intent of the modules is to build automated graphical displays that help condense and compare a large amount of information so as to provide a near real-time assessment of instrument performance, and in a comparison sense, key geophysical parameters. Module development is being focused into economical displays that will make it easy for a data analyst to identify problems, given proper guidance.

The modules will also take advantage of model output which can serve as a guide or baseline as to how field-collected data should look. A good example of this is the use of clear-sky models to estimate the hemispheric flux for comparison to ARM unshaded precision spectral pyranometer (PSP) values and the summed direct and diffuse fluxes.

At present, the following graphical display modules are currently being produced:

Clear-sky (all Extended Facilities with SIROS), **Albedo** (same as Clear-sky), **Temperature** (Central Facility EBBR, SMOS, BBSS, 60-m tower), **Relative Humidity** (Central Facility EBBR, SMOS, BBSS, 60-m tower), **Surface Mixing Ratio** (Central Facility EBBR, SMOS, BBSS, 60-m tower), **Vapor Pressure** (Central Facility EBBR, SMOS, BBSS,

60-m tower), **MWR** (Central Facility and Boundary Facilities), **Mixing Ratio** (Central Facility and Boundary Facility BBSS profiles).

These displays, formerly housed on an SST-managed WWW page, are being transitioned to algorithm scripts which can be run on an as-needed basis on the Central Facility's "research one" computer system.

Calibration and Maintenance Information Summaries

In late 1995, the SST began to assist Site Operations in tracking calibration and maintenance information produced by Instrument Mentors and others within the ARM Program. This was done to try to produce more readily available, up-to-date information on calibration and maintenance procedures and their results. The way in which this kind of information is presented depends upon the user of the information. Site Operations needs a comprehensive documentation of calibration and maintenance procedures so that it can fulfill its function of performing calibration checks and maintenance activities to the best of its ability. Data users likely need less detailed information about the exact procedures, but do need good information on what procedure was performed and when it was performed, and when the next procedure is scheduled to be performed. ARM data system managers need to know when calibration coefficients for particular instruments change so that they can make appropriate changes in data ingest procedural files.

Specific questions needed to be answered before these needs could be met. These include the following questions:

"what are the calibration and maintenance procedures?"

"where are they documented?"

"are they vendor or Instrument Mentor prescribed?"

"who implements the various procedures?"

"how are calibration and maintenance metadata made available for analysis?"

"what are the prescribed schedules for performing procedures?"

"are unscheduled procedures performed, and who initiates them?"

"how are competing interests prioritized?"

The SST attempted to begin answering these questions by issuing to the Instrument Mentors an "Assessment of Instrument Calibration and Maintenance Procedures" questionnaire to fill out for each instrument platform. The questionnaire form was made to be comprehensive to cover both calibration and maintenance issues, and how that type of information is made available to others.

The results of returned questionnaires were compiled into WWW tables which can be viewed by contacting the lead author. They were the subject of a session on SGP calibration issues held at a meeting in New Mexico in February 1997. Although no final decisions have been made concerning the long term monitoring of such information, it became clear that it would be a good idea for the SST to develop a recommended calibration schedule for use by Site Operations staff based on information provided by Instrument Mentors (Site Operations already conducts a comprehensive set of maintenance procedures on a bi-weekly basis at each ARM facility), and that some sort of a "Calibration Handbook" should be produced by ARM for data users. The form of the latter has not been decided upon.

It is clear that Site Operations needs solid documentation for the calibration procedures they are asked to perform (an Instrument Mentor function) and firm work schedules (aided by the SST). Site Operations staff must also be properly trained to perform the procedures prescribed by both Instrument Mentors and vendors. Data users, for a particular data stream they are analyzing, need to know when and how the instrument producing the data was calibrated and maintained, and where they can go to find such information.

In response to the WWW questionnaire summary tables, Instrument Mentors have agreed to re-evaluate their current procedures and the frequency with which they are performed. The SST will continue its efforts to update and occasionally review calibration and maintenance information.

Special Shortwave Radiation Tracking Exercise for the Central Facility

With the DSIT and others in the ARM Program, the SST is helping to facilitate the tracking of solar radiation measures at the Central Facility in anticipation of the launch of the Cloud and Earth Radiant Energy System (CERES) remote sensing capability in August 1997. Ten tracking mechanism

measurements were suggested through Science Team input to comprehensively look at solar radiation, with a primary and a secondary method for obtaining the measurement identified. These include (DHF refers to downward hemispheric flux):

Suggested Tracking Mechanisms			Present SST QC Effort
1. DHF-	Primary:	SIROS NIP+PSP (shaded)	YES vs. PSP (unshaded)
	Secondary:	BSRN	YES vs SIROS
		SIROS PSP (unshaded)	YES; and vs. NIP+PSP (shaded)
		BSRN PSP (unshaded)	YES vs. SIROS
2. Direct -	Primary:	SIROS NIP	YES
	Secondary:	BSRN NIP	YES vs. SIROS
		MFRSR (calculated)	NO
3. Diffuse (total) -	Primary:	SIROS PSP (shaded)	YES
	Secondary:	BSRN PSP (shaded)	YES vs. SIROS
4. Direct (spectral) -	Primary:	MFRSR (Cimel future)	NO
	Secondary:	Spectral MFRSR (future)	NO
		ASD; ASTI (future)	NO

The SST has been charged to produce the algorithms to compute the metric values while the DSIT is creating the interfaces to view the results as an ensemble. Work is currently in progress with an August 1997 target date for completion.

Summary

Even though the SST can be viewed as the final arbiter of ARM data quality, data quality issues are addressed at several levels within the ARM Program. This division of activities especially aids the SST in making quality assessments, allowing them to draw upon the expertise of other groups, such as the Instrument Mentors, DSIT, Site Operations, and ARM scientists (particularly those involved in the creation of QMEs and VAPs). One of the goals of the ARM Program is to provide data streams of known and reasonable quality. Achieving this for a program of such size and complexity is a significant challenge. Chapter 4 of the January-June 1997 issue of the *Site Scientific Mission Plan* discusses each of these group's activities and responsibilities in greater detail.