



U.S. DEPARTMENT OF
ENERGY | Office of
Science

DOE/SC-ARM-12-014

Data Quality Assessment and Control for the ARM Climate Research Facility

June 2012



DISCLAIMER

This report was prepared as an account of work sponsored by the U.S. Government. Neither the United States nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

Data Quality Assessment and Control for the ARM Climate Research Facility

June 2012

Work supported by the U.S. Department of Energy,
Office of Science, Office of Biological and Environmental Research

Contents

1.0	Introduction	1
2.0	Data Quality Program Organization and Activities.....	1
2.1	The Data Quality Office.....	1
2.2	Instrument Mentors	2
2.3	Translators.....	2
2.4	Site Scientists	2
2.5	Site Operators.....	3
3.0	Data Quality Program Focus Areas: Radars and Value-Added Products.....	3
3.1	Radars.....	3
3.2	Value-Added Products	3
4.0	Resources for Assessing Data Quality.....	4
4.1	Instrument Web Pages: Specifications and Expectations.....	4
4.2	Data Quality Health and Status: Display of Current and Recent Conditions.....	4
4.3	Data Quality Reporting: Notification to Problem Solvers and Data Users	5
4.4	Value-Added Products: Second-Generation (and Higher) Datastreams	5

1.0 Introduction

The mission of the Atmospheric Radiation Measurement (ARM) Climate Research Facility is to provide observations of the earth climate system to the climate research community for the purpose of improving the understanding and representation, in climate and earth system models, of clouds and aerosols as well as their coupling with the Earth's surface. In order for ARM measurements to be useful toward this goal, it is important that the measurements are of a known and reasonable quality. The ARM data quality program includes several components designed to identify quality issues in near-real-time, track problems to solutions, assess more subtle long-term issues, and communicate problems to the user community.

2.0 Data Quality Program Organization and Activities

2.1 The Data Quality Office

The core of the ARM data quality program is the Data Quality (DQ) Office that was established in July 2000 to help coordinate the continued evolution and implementation of efforts to ensure the quality of the data collected by field instrumentation. The DQ Office has the responsibility to ensure that quality assurance results are communicated to

- data users, so that they may make informed decisions when using the data
- ARM's Site Operations and engineers, to facilitate improved instrument performance and thereby minimize the amount of unacceptable data collected.

The DQ Office is based at the University of Oklahoma. In addition to reviewing data, this group develops tools to facilitate DQ review and supervises a group of students who review data plots on a daily basis. The DQ staff members include a lead DQ Manager as well as one person focused on assessing radar data and one focused on advanced data products.

The DQ Manager provides overall guidance and management for the DQ program to ensure that the data collected at the ARM Climate Research Facility sites meet the data quality objectives and tolerances as defined by the ARM Instrument Team and the ASR Science Team. The DQ Manager reports to the ARM Technical Director and works with various parts of the Atmospheric System Research (ASR) Science Team and ARM Infrastructure to develop and monitor an end-to-end data quality system that provides both continuous, consistent quantitative assessment and continual improvement of ARM datastreams.

The DQ Manager must be able to work with and help monitor the activities of those involved in the data quality effort, in conjunction with the Engineering and Operations components of the ARM Facility infrastructure. The DQ Manager is responsible for ensuring consistency within instrument suites and across research sites regarding data-checking algorithms, metadata, and maintenance prescribed by quality checks, calibration, and reporting. In particular, the DQ Manager works closely with the Technical Director, Operations Manager, Instrument Team Coordinator, Data System Coordinator, and Archive Manager as part of the end-to-end process, from instrument engineering and deployment to data collections and maintenance. The DQ Manager is an ex officio member of the ARM Science and Infrastructure Steering Committee (SISC).

While the DQ Office plays a central role in assessing and controlling ARM data quality, it works closely with a number of other groups. Additional groups that play important data quality roles are instrument mentors, translators, site scientists, and site operators. These groups span a range of technical responsibilities and represent both the ARM infrastructure and the science user community.

2.2 Instrument Mentors

Instrument mentors are the technical leads for ARM instruments. Each ARM instrument is assigned one or more mentors. Mentors are charged with developing the technical specifications for instruments procured for the ARM Climate Research Facility, developing procedures for instrument operations (e.g., daily rounds, maintenance, and calibration), assessing instrument status and data quality, and managing instrument repairs. In addition, the mentors work with data system personnel on data product requirements. These requirements include the specification of appropriate operating ranges and associated flags when data fall outside of that range. This data product flagging is the first level of data quality assurance. Data flags provide an indication to the DQ Office, mentors, and science data users regarding obvious data quality problems.

The DQ Office team provides weekly assessments of data quality status and submits problem reports on shorter time scales where necessary. Instrument mentors use the results of the operational data quality assessment to solve problems that ultimately improve instrument performance and functionality, including re-engineering if necessary. In addition, the mentors will perform a consulting role by providing tools and advice on how to best assess the status of their instruments and answering questions when unusual instrument behavior is observed. The mentors also will provide a monthly summary of data quality status for their instruments and submit Data Quality Reports (DQRs) that are associated with specific data files and delivered to users when those files are ordered (see below).

2.3 Translators

Translators are a group of scientists who serve as liaisons between ASR working groups and the ARM infrastructure team. These individuals have expertise ranging from radiometric measurements to cloud modeling. One of the main roles of the translators is to facilitate the development of advanced data products by translating the scientific needs of the research community to the software developers who actually build the data products. However, through the development of these data products, this group develops a special familiarity with many datastreams over extended periods. Though this group is not specifically responsible for instruments, this group also gains unique knowledge about instrument issues, which they feed back to instrument mentors or the DQ Office.

2.4 Site Scientists

ARM sites are typically assigned a site scientist who, like the translators, serves as a liaison between the ARM infrastructure and the science community. Unlike the translators, site scientists are part of the ASR user community (versus the ARM infrastructure). These individuals represent their sites (e.g., the North Slope of Alaska or the Tropical Western Pacific) by advocating for their site, promoting and carrying out research, supporting field campaigns, and assessing data quality. Site scientist data quality assessment efforts involve more in-depth evaluation of both individual and multiple sets of datastreams, as needed, to

address data quality issues specific to the site. These efforts both complement and augment those of instrument mentors and the DQ Office. Site scientists may also perform research on topics related to site data quality issues.

2.5 Site Operators

Site Operators provide the day-to-day support in the field for instruments, along with other site operations activities. When DQ Office staff and instrument mentors identify a problem with an instrument, they will work with site operators in the field and the instrument mentors to help diagnose the problem and carry out corrective maintenance. DQ Office staff, with the respective site scientists, shall be responsible for overseeing the documentation of the problem and related corrective action.

3.0 Data Quality Program Focus Areas: Radars and Value-Added Products

As noted above, the DQ Office includes staff members who particularly focus on the radars and on value-added products (VAPs). The focus on radars followed the significant expansion of the ARM Facility's radar network through the American Recovery and Reinvestment Act of 2009 (Recovery Act), while the focus on VAPs was in recognition of a need to formally assess these advanced products, which are becoming increasingly important for advancing research.

3.1 Radars

The DQ Office team plays a special role with respect to the quality assurance of radars and radar profilers obtained through the Recovery Act. The DQ Office team monitors the near-real-time radar datastreams on a daily to weekly basis in order to identify short-term problems, and it provides weekly data quality assessment reports. Radar data that are not delivered in near-real-time are reviewed and assessed in a similar method when data become available. Once identified, problems are communicated to Site Operations, instrument mentors, and radar technicians. The Data Quality team also aids the instrument mentor in performing mid- to long-term monitoring of radar data in order to determine data quality issues that may affect the data and aids the instrument mentor in communicating these issues to the end users.

3.2 Value-Added Products

The DQ Office has developed (and continues to develop) procedures for assessing ARM VAPs in coordination with the translators. The analysis of VAPs includes the following:

- production-level datastreams that have been fully developed and released for general use
- evaluation products that are still under review and/or development but that are available to the user community in a limited way.

Currently the DQ Office supports VAPs primarily by evaluating that the quality checks built into the VAPs are functioning as intended. As such, there is an emphasis on products under development. Typically, VAPs incorporate significantly greater internal error checking than first-generation

datastreams, and the DQ Office has been working to ensure that the information provided by these checks in the form of quality flags in the VAP data files are accurate. Related to this work, the DQ Office works with the VAP translators to review implementation plans for new evaluation products under proposal. Evaluation products that are not on track for release as production VAPs are not typically reviewed unless deemed necessary by the Technical Director, VAP translators, and DQ Office staff.

Recently, data quality reporting tools that had previously been accessible only to first-generation data products have been made available also for VAPs. This has allowed translators to more easily communicate VAP data quality issues to the user community and the ARM infrastructure. At present VAPs are not reviewed on a regular basis by the DQ Office in the same way that first-generation products are, in part because of the varied timescales on which VAPs are produced; however, discussions are underway to develop an effective process for continually assessing VAP data quality in a formal way beyond the automated checks that are built into the products themselves.

4.0 Resources for Assessing Data Quality

4.1 Instrument Web Pages: Specifications and Expectations

The ARM Climate Research Facility has created a series of detailed web pages that describe the specifications of its instruments. A fundamental of quality control is a "statement of expectations." Quality is the measure of how closely something conforms to an expectation. Without an expectation, a quality assessment is not possible. Thus, the instrument pages represent ARM's statement of expectations for its instruments, the baseline against which the observations can be compared. These web pages include our current understanding of the measurement systems and their quirks and deficiencies, including common problems encountered or inherent to the measurement.

These pages are also the place for you to find specific information about the data quality for each instrument. For a quick summary of ARM instruments at each field site, view the Instrument Location Table under the "Instruments" tab at <http://www.arm.gov>.

4.2 Data Quality Health and Status: Display of Current and Recent Conditions

The DQ Office has produced a web area dedicated to displaying data quality "health and status" results in near-real-time so that those involved in quality assurance activities can monitor data quality. This web area provides access to a suite of data quality assessment tools as well as links to supporting information, such as instrument and maintenance logs and various reports related to the data quality effort. The tool suite includes the following tools:

- DQ Explorer: daily QC metrics with a color representation of quality flags on an hourly basis
- DQ PlotBrowser: a specialized diagnostic plots browser that allows quick navigation of multiple plot thumbnails at once
- NCVweb: an interactive plotting capability that allows for easy plot manipulation plus the display of file headers, variable details, and statistics; listings of the data values themselves; and a mechanism to create an ASCII copy of the netCDF data file.

4.3 Data Quality Reporting: Notification to Problem Solvers and Data Users

A [Data Quality Problem Report](#) (DQPR) allows the DQ Office, instrument mentors, site operators, and site scientists to submit problems to site operations and track the problem resolution process. A DQPR is closed when the problem is solved, with the final act being the writing of a DQR. If a problem cannot be solved within a reasonable amount of time (usually 30–45 days) or is catastrophic in nature, it is escalated in importance and is put before the Problem Review Board (PRB). The PRB consists of representatives from across the ARM infrastructure and meets on a weekly to bi-weekly basis to ensure that problems are being resolved and reported in a timely manner.

A [Data Quality Report \(DQR\)](#) is a written statement about the quality of data in a particular datastream. The information could be quite simple (e.g., stating that an instrument system was turned off and the data do not exist) or quite complex (e.g., providing detailed analyses and equations that should be used to adjust the instrument's data). At present, when a person orders ARM data from the [ARM Data Archive](#), all DQRs written on the desired datastreams are attached to the order. DQRs are typically written by the appropriate instrument mentor, who has the final say on data quality, though they can be written by other members of the ARM infrastructure or suggested by the user community. These DQRs, as described earlier, provide a quality designation beyond that of automated flagging. While DQRs are currently delivered as text files, work is underway to make the high-level quality information contained in DQRs (assessment of whether the data are good, bad, uncertain, or missing) available in more useful ways. These include using this information as a filter during the data ordering process and the availability of a web service that would allow data users to filter data within their own programs based on information within DQRs.

Although data users may submit DQRs directly, they are also encouraged to alert ARM infrastructure staff if they observe an undocumented data quality issue. Alerting the appropriate mentor, translator, or site scientist or any member of the DQ Office will initiate the process of documenting and addressing the problem if it is ongoing. Typically, any ARM web page has one or more links through which such comments can be submitted if it is not clear who is the appropriate contact.

4.4 Value-Added Products: Second-Generation (and Higher) Datastreams

Many of the scientific needs of the ARM Climate Research Facility are met through the analysis and processing of existing data products into "value-added" products or [VAPs](#). Despite extensive instrumentation deployed at the ARM sites, there will always be quantities of interest that are either impractical or impossible to measure directly or routinely. Physical models that use ARM instrument data as inputs are implemented as VAPs and can help fill some of the unmet measurement needs of the Facility. Additionally, ARM produces some VAPs not in order to fill unmet measurement needs, but instead to improve the quality of existing measurements. For several types of measurements (e.g., broadband radiation), when more than one measurement is available, ARM produces "best-estimate" VAPs.



U.S. DEPARTMENT OF
ENERGY

Office of Science