J3.13 IMPROVEMENTS TO AND STATUS OF ARM'S DATA QUALITY HEALTH AND STATUS SYSTEM

Randy A. Peppler *, Kenneth E. Kehoe, Karen L. Sonntag
Cooperative Institute for Mesoscale Meteorological Studies, The University of Oklahoma, Norman, OK

Sean T. Moore
Mission Research Corporation, Santa Barbara, CA

Kathy J. Doty
Brookhaven National Laboratory, Upton, NY

1. INTRODUCTION

The U.S. Department of Energy Atmospheric Radiation Measurement Program Data Quality Office (DQO) was formed in July 2000 in response to a program review in 1999 identifying a need for formal coordination of ARM's data quality efforts.

As the DQO has evolved and matured during the ensuing five years, it has taken on the role of coordinating the continued evolution and implementation of efforts to assure the quality of the data collected by the Program's field instrumentation. The DQO has the specific responsibility for ensuring that quality assurance results are communicated to (1) data users so that they may make informed decisions when using the data, and (2) ARM's Site Operations and Engineers to facilitate improved instrument performance, thereby minimizing the amount of unacceptable data collected. Toward these goals, DQO analysts, instrument mentors, site scientists, and others in ARM help review and assess ARM's datastreams and write and submit reports documenting their findings. Such reporting includes interactive tools for documenting the problem resolution process, as the DQO works closely with instrument mentors and site operators to solve problems.

To facilitate this process, the DQO developed a web-based tool called the Data Quality Health and Status (DQ HandS) system (http://dq.arm.gov/). DQ HandS reads and processes ARM data files, displays flag information in the form of hourly color tables that provide a mouse-over capability for determining the nature of flags violated, produces diagnostic plots of primary measurements, including key measurement intercomparisons when possible, in both individual plot format and within a plot browser, allows for interactive plotting of datastream variables using a tool called NCVweb, and hosts or links to various assessment and problem reporting mechanisms.

Since the 14th Conference on Applied Climatology (Peppler et al. 2004), a number of improvements have been made to DQ HandS, and are reported on here and in our presentation.

* *Corresponding author address:* Randy A. Peppler, CIMMS/The University of Oklahoma, 100 E. Boyd St., Rm. 1110, Norman, OK 73019; rpeppler@ou.edu.

These improvements include:

- Inclusion of ARM Mobile Facility data
- Incorporation of more ARM datastreams and datastream intercomparisons
- Development of a new thumbnail plot browser to facilitate the viewing of trends
- More functionality for NCVweb, an interactive plot viewer
- Development of an improved method for writing and databasing data quality assessment reports and an improved method for searching for ARM reports to better gauge problem context

We also discuss some new ideas for improving the amount and quality of automated flagging provided within ARM data files and a method for tying what we learn quality-wise during the creation of value-added products back to the lower-level datastreams used as input. Challenges are also described.

2. IMPROVEMENTS TO DQ HandS

2.1 Inclusion of ARM Mobile Facility Data

The ARM Mobile Facility (AMF) began data collection at Point Reyes, CA, in March 2005 in its first official deployment as part of a study to observe marine stratus and drizzle processes. Data are fed continuously into DQ HandS just like data from ARM's fixed Climate Research Facilities in the U.S. Southern Great Plains (SGP), North Slope of Alaska (NSA), and Tropical Western Pacific (TWP). DQO analysts look at DQ HandS output on a daily basis and file problem reports as needed. Assessment reports are issued weekly. Fig. 1 shows a schematic of the AMF and an example data quality assessment report issued for sonde data during the week of 28 April-4 May 2005.

2.2 Incorporation of More ARM Datastreams and Datastream Intercomparisons

Key additions to the DQ HandS suite of instrument datastream analyses and datastream intercomparison include the Microwave Radiometer Profiler at the AMF Point Reyes deployment and at the NSA Barrow facility

PYE SONDE Data Quality Assessment Report for ...

File Edit View Go Message Tools Help

Subject: PYE SONDE Data Quality Assessment Report for 20050428-20050504
From: webformadmin@arm.gov
Date: 5/9/2005 3:16 PM
To: kkehoe@qcn.ou.edu

Data evaluated based on visual inspection of DQ Health and Status tables found at the DQHandS website.
These reports are produced by the ARM Data Quality Office at the University of Oklahoma.

Questions can be sent to Karen Sonntag (ksonntag@ou.edu) or Randy Peppler (rpeppler@ou.edu)

*** **new developments are highlighted in bold** ***

PYE SONDE Data Quality Assessment Report for 20050428 - 20050504

M1 (Point Reyes, CA): Data quality OK except for the following:
4/28 23Z launch -no data above 925mb.
4/30 17Z launch -incorrect T, Td at surface.
4/30 23Z launch -incorrect T, Td at surface.
5/1 18Z launch -incorrect T at surface, no data above 350mb.
5/1 23Z launch -incorrect T at surface.
5/3 23Z launch -missing wind above 400mb.

To view this assessment and related open DQPRs,open PIFs,DQRs,BCRs and ECRs please visit the DQA View web page.

Fig. 1. ARM Mobile Facility (top panel) and data quality assessment report for sonde data (bottom panel) for the period 28 April-4 May 2005.

and the Eddy Correlation system at Point Reyes and at various SGP facilities. Quality assurance for the microwave radiometer profiler includes comparison of key measurements such as precipitable water vapor, liquid water path, and brightness temperature with those from co-located microwave radiometers, and of temperature with that from co-located sonde launches. Examples of these diagnostic plot comparisons are shown in Figs. 2 and 3. For the eddy correlation system, comparison of sensible and latent heat and other variables are being made with those from co-located energy balance Bowen ratio systems (Fig. 4).

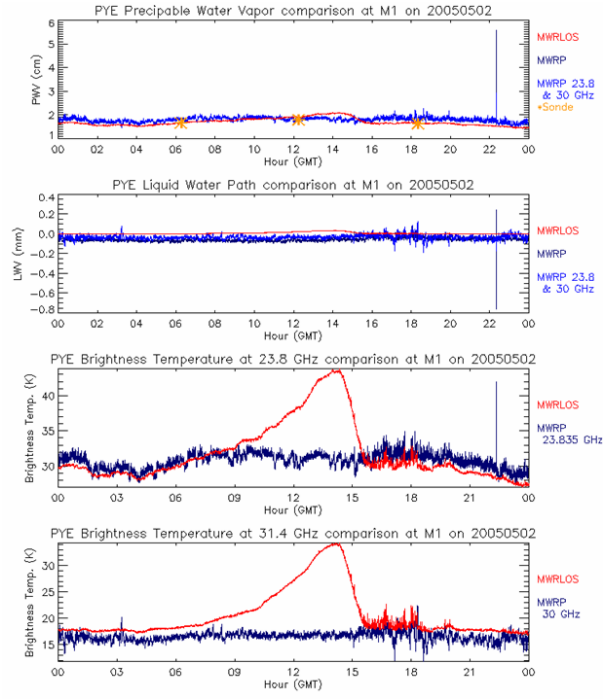

Fig. 2. Comparisons of precipitable water vapor, liquid water path, and brightness temperature at 23.8 and 31.4 GHz at Point Reyes from a microwave radiometer profiler (red) and a microwave radiometer (blue).

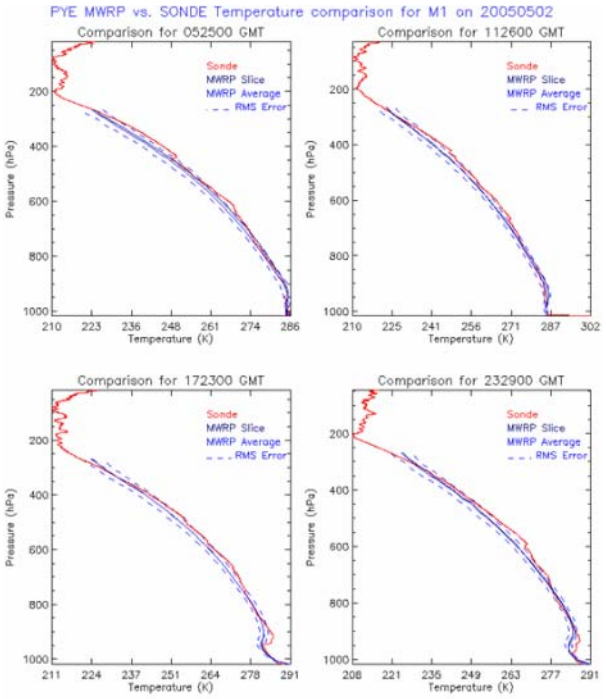

Fig. 3. Comparisons of temperature at Point Reyes from a microwave radiometer profiler (blue) and a sonde launches (red).

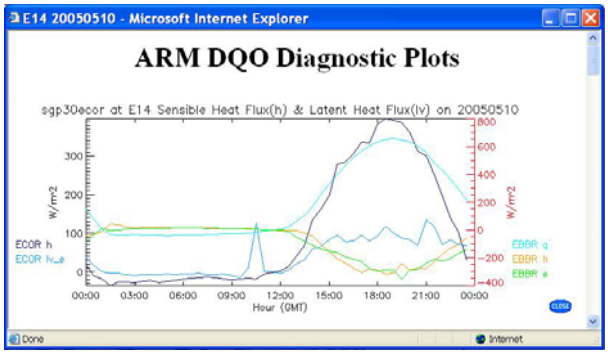

Fig. 4. Comparisons of sensible heat flux and latent heat flux from an eddy correlation system and an energy balance Bowen ratio system located at SGP Lamont.

Also, a new set of comparisons have been established between SGP atmospheric state measurements of temperature, relative humidity, precipitation amount, and wind speed and direction and those from nearby Oklahoma Mesonet sites. An example comparison is shown in Fig. 5 for the SGP Extended Facility 20 (Meeker, OK) and three nearby Mesonet sites.

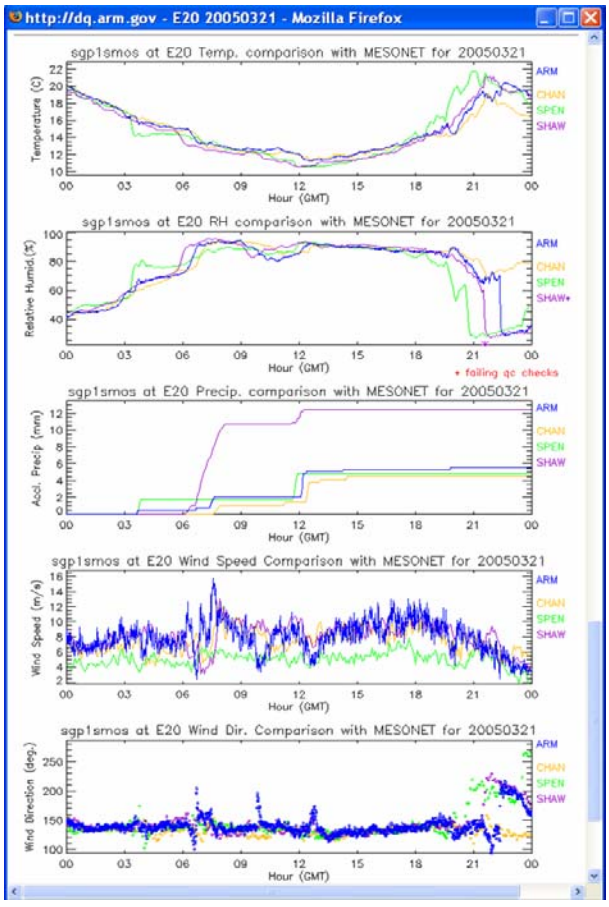

Fig. 5. Comparisons of temperature, relative humidity, precipitation, and winds from SGP Meeker and three nearby Oklahoma Mesonet sites.

2.3 Development of a New Thumbnail Plot Browser

To facilitate the viewing of multiple diagnostic plots in one computer window, a new browser was developed to show multiple thumbnail plots in chronological sequence (http://dq.arm.gov/plotbrowser/). This tool also allows for list view for more selective viewing of variables and for the expansion of thumbnails into full view plots. The number of days desired for viewing can be specified, as can be the number of variables and facilities to be shown. Fig 6. shows an example thumbnail view of surface meteorological data from several SGP facilities and a corresponding list view displaying a full plot from one facility.

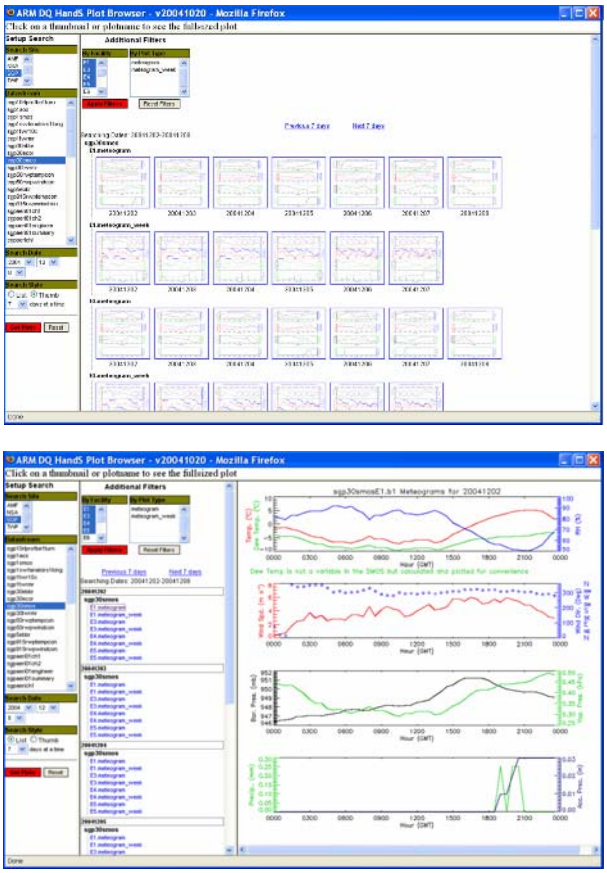

Fig. 6. Example view of the thumbnail plot browser (top) and list view (bottom) with a full view of one thumbnail.

2.4 More Functionality for NCVweb

NCVweb provides a capability within DQ HandS to interactively play with data to zoom in on a feature of particular interest or interactively string together a series of data to facilitate looking at trends. It also provides the capability to overlay variables from multiple facilities and look at three-dimensional variables. It provides a number of data file features as well, including listings of variable details, file headers, flagging statistics, and creation of file dumps. Fig. 7 shows examples of multi-day, multi-facility and three-dimensional views.

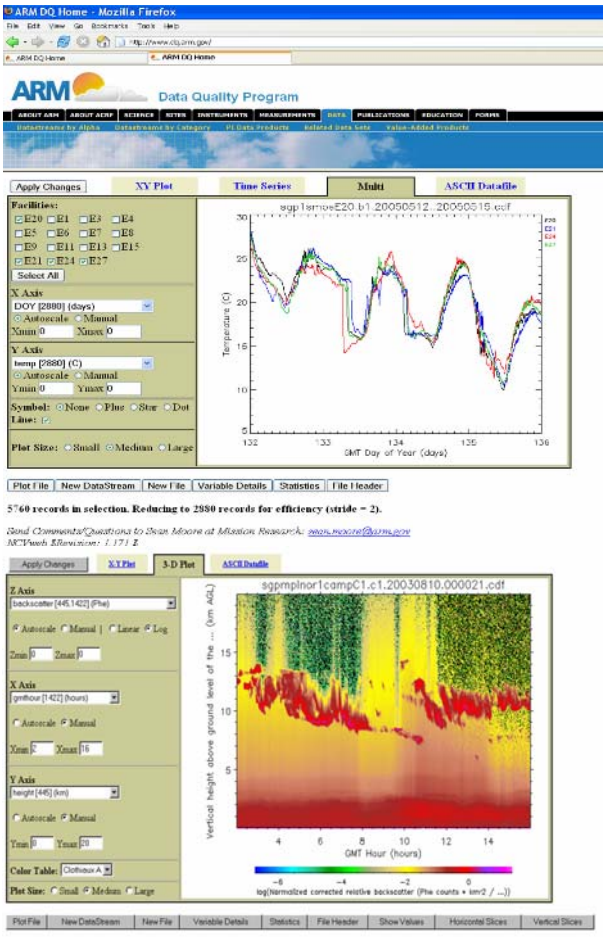

Fig. 7. NCVweb display of surface temperature from four SGP Extended Facilities over a four-day period (top) and a visual of micropulse lidar cloud spectra from SGP Lamont (bottom).

2.5 Development of Improved Methods for Writing and Databasing Data Quality Assessment Reports and Problem Report Searching

The DQO needed a way to better write and database its weekly assessment reports on the quality of ARM data. This had previously been done through e-mail. A new method was developed, within DQ HandS, to write these assessments, database them, and recall past ones for context as needed. In addition, a capability was created to e-mail these reports to appropriate personnel once completed. Shown on these reports is the assessment of the analyst, any past related ARM problem reports pertaining to this measurement and facility, and any past change requests relating to the same. This work produced an ARM-wide search capability and relational reporting database that allow one to put into historical context any current problems. Fig. 8 shows an example assessment report, and Fig. 9 shows the output of an ARM report search.

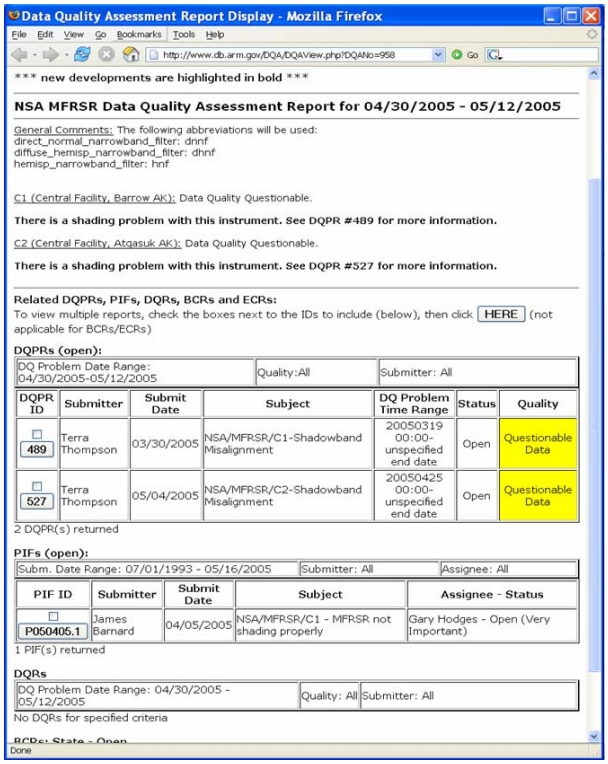

Fig. 8. An example of a data quality assessment report.

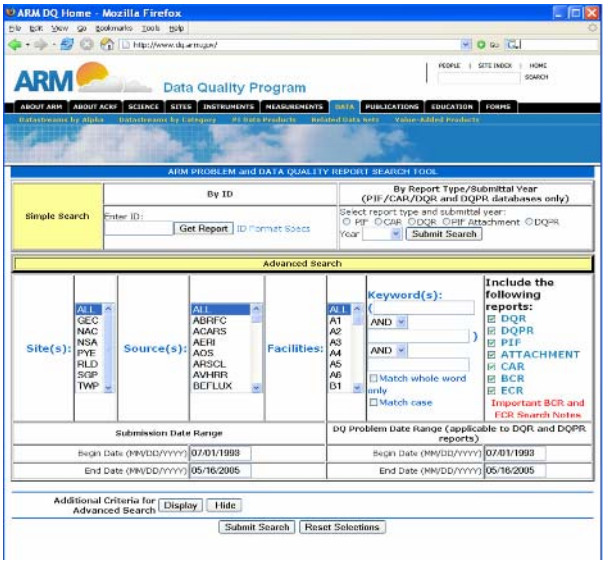

Fig. 9. The problem report search menu.

3. NEW IDEAS

3.1 Better Flagging within Files

At present, in most cases, automatic flagging within ARM data files during official processing is limited to checking for violation of minimum and maximum limits, and simple delta (abrupt change) checks. The

DQO for many instruments either provides and applies more stringent limits on these checks or applies more sophisticated checking algorithms as prescribed by instrument mentors. At present, the results of these additional checks are not provided to data users. Thus, a program is underway within ARM to provide more sophisticated automated checking of data within the ARM data files that are made available at its data archive. This processing would be conducted not during hour-by-hour data ingest and processing, but once a complete 24-hour set of data have been collected. A module will be developed for each instrument/primary measurement that will reprocess a day's worth of data and rewrite that file. This file will then be sent to the DQO for display of its data quality information and for the construction of diagnostic plots, and the file will also be sent to the data archive. Fig. 10 shows the current perspective and proposed revision.

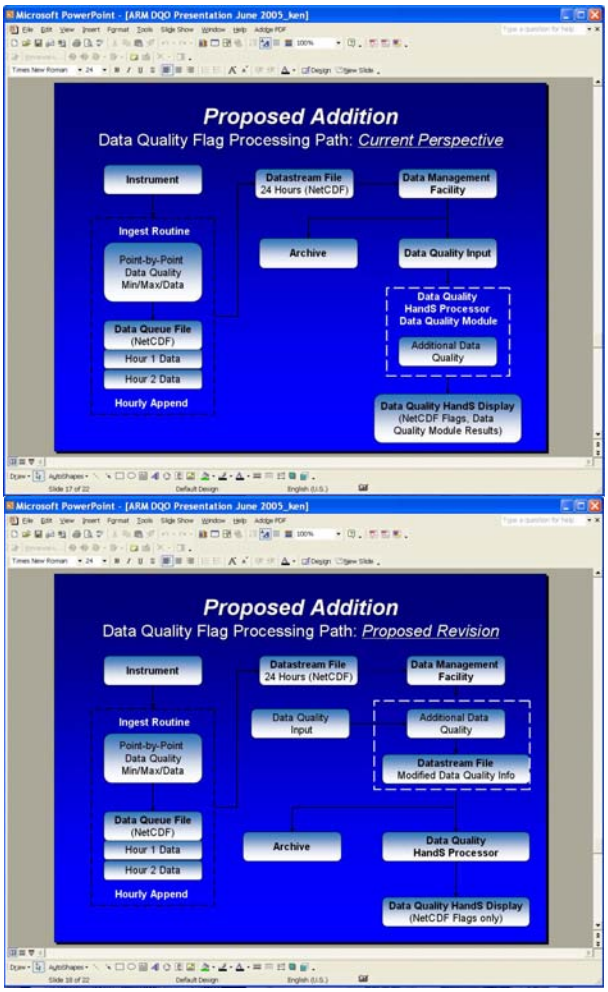

Fig. 10. Current perspective (top) and proposed revision (bottom) to ARM data file automated quality processing.

3.2 Value Added Processing Quality Control

At present, there is no easy way to comprehensively and consistently extract information compiled about the quality of input data streams during the creation of second-generation, value added products, such as best estimate datastreams and cloud products. The creation of such products often subjects the input data to new stringencies that may well reveal issues with them that were previously unknown. Thus, an automated check and alert system is being developed to extract such information identified during the creation of value added products, determine whether the issue identified was previously found and documented during the quality processing of the input datastream, and if not, produce and issue a warning report about a potential previously unseen problem. This report will then activate a problem resolution process that may well result in the reprocessing of a datastream.

4. CHALLENGES

The DQO and the ARM Program in general faces two main challenges in addition to the two now being addressed and described above in section 3. These include providing detail about the quality of ***all*** individual datastreams produced by ARM, and applying new and improved automated checking algorithms to past ARM data (particularly those collected before the formation of the DQO). Both are important yet slow tasks to complete, but will be given priority in the coming years.

5. REFERENCES

Peppler, R. A., K. L. Sonntag, A. R. Dean, and C. M. Shafer, 2004: Quality of Atmospheric Radiation Measurement (ARM) Program data. *Preprints, 14th Conference on Applied Climatology*, Seattle, WA, January 11-15, 2004, American Meteorological Society, CD-ROM – P4.3.