CONF-960138--1

BNL-62153

TOOL TO ASSESS CONTENTS OF ARM SURFACE METEOROLOGY NETWORK NETCDF FILES

Amanda Staudt

Harvard University
Cambridge, MA

Tammy Kwan and Joyce Tichler[1]

Brookhaven National Laboratory
Upton, NY

1. BACKGROUND

The Atmospheric Radiation Measurement (ARM) Program (Stokes and Schwartz, 1994), supported by the U.S. Department of Energy, is a major program of atmospheric measurement and modeling designed to improve the understanding of processes and properties that affect atmospheric radiation, with a particular focus on the influence of clouds and the role of cloud radiative feedback in the climate system. The ARM Program will use three highly instrumented primary measurement sites. Deployment of instrumentation at the first site, located in the Southern Great Plains of the United States, began in May of 1992. The first phase of deployment at the second site in the Tropical Western Pacific is scheduled for late in 1995. The third site will be in the North Slope of Alaska and adjacent Arctic Ocean.

To meet the scientific objectives of ARM, observations from the ARM sites are combined with data from other sources; these are called external data. Among these external data sets are surface meteorological observations from the Oklahoma Mesonet (Brock, et al, 1994), a Kansas automated weather network[2], the Wind Profiler Demonstration Network (WPDN) (WPDN Report, 1994), and the National Weather Service (NWS) surface stations. Before combining these data with the Surface Meteorological Observations Station (SMOS) ARM data, it was necessary to assess the contents and quality of both the ARM and the external data sets. Since these data sets had previously been converted to netCDF (Jenter and Signell, 1992) format for use by the ARM Science Team, a tool was written to assess the contents of the netCDF files.

2. DESCRIPTION OF TOOL

The tool, netCDF Quick Look (ncql), has three goals: to retrieve basic descriptive information about the data in the file; to find and document time intervals in which data is missing or out of the expected range; and to determine the actual range of data while noting data points that fall outside of the expected maxima and minima and the time periods over which they occur.

An internet search of netCDF utilities officially maintained by Unidata, user contributed netCDF utilities, and on-line netCDF user support groups revealed the existence of two other command line tools that give a quick overview of netCDF files. Ncstat (Rew and Emmerson, 1992), officially released by Unidata as a Unidata Utility, determines the minima, maxima, means, and number of fill values (numeric codes used to indicate missing data) for variables dependent on a set of user input dimensions. NetCDF Kitchen Sink (ncks)[3], is a user contributed program that returns summary information about the variables and dimensions in a file. Both utilities give the user more control over which vari-

---

1. *Corresponding author address:* Joyce Tichler, Brookhaven National Laboratory, Department of Applied Science, Building 490D, Upton, NY 11973-5000. E-mail: tichler@bnl.gov.
2. These data are made available to ARM by Mary Knapp, the Kansas State Climatologist.
3. Ncks was created by Charlie Zender and can be obtained by anonymous ftp from the pub/zender directory of ftp.cgd.ucar.edu.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.

ables are analyzed and which output statistics are retrieved than ncql allows. However, ncql lists the number of values out of range and flagged as missing, and the time periods when such invalid data occur, thus fulfilling objectives not met by the other routines.

Ncql is written in C, takes a netCDF file as input, and currently outputs to STDOUT. No changes are made to the input data file. The coding aims to optimize the time and memory needed to assess the file. The data values are scanned once and only those of interest (maxima, minima, and missing value ranges) are stored.

Input netCDF files that conform to ARM formats permit retrieval of missing value flags and expected maxima and minima from the metadata of the input file rather than from defaults. Ideally, each variable should be described by the following attributes: long_name, units, missing_value, valid_min, and valid_max. However, since ARM has not established rigid netCDF file format conventions, variations occur among data sets and ncql must be more flexible as to the data file format.

Currently, an ascii file containing default information for each variable can be included as an optional command line argument to ncql. In order to create the default file in the correct format an additional utility is under development. Ncql Get Defaults (ncql_gd) will prompt the user for variable names, expected extrema, and missing value codes to use for comparisons with the actual data. The ability to incorporate defaults will make ncql general enough to accept virtually any netCDF file.

Ncql produces a series of statistics about the netCDF file including:

- expected start and end times of data collection retrieved from the ARM filenaming convention and from the metadata
- actual start and end times of data collection
- expected reporting time interval retrieved from the file's global attributes
- inconsistent deltas between time entries (indicating data retrieval was not constant - for example, gaps in time when all instruments were inoperable)
- expected minimum and maximum values for each variable, as given in the metadata or in the defaults
- actual minimum and maximum values for each variable
- coded values for missing data, as given in the metadata or in the defaults
- number of values that are out of range or flagged as missing for each variable
- intervals of time when values are out of range or flagged as missing for each variable

An example of the output for a subset of variables in an ARM WPDN netCDF file can be found in table 1.

3. APPLICATIONS AND USE OF NCQL

Ncql gives end users of netCDF files the ability to obtain a quick overview of lengthy files and highlight any irregularities within the data. It can be utilized to locate instrument and transcription problems, as well as to determine ranges of invalid data to ignore during analysis and ranges of unusual data to study further.

The tool may also be incorporated in an automated quality control procedure to routinely evaluate data sets. For example, an algorithm has been developed by the WPDN which assigns a quality control code to each data point in their surface meteorological files. The code indicates the existence of missing, uncertain, and out of range data thereby giving end-users additional information to use when determining the validity of the data. The ARM SMOS netCDF files also contain quality control fields that utilize a more complicated, condensed algorithm to flag fields with values above or below valid extrema and those that exceed the maximum allowed difference between values in consecutive time periods.

In a similar manner, ncql may be further developed to create quality control fields in a new type of ARM data file which is under development. The integrated surface meteorological (ISM) data set will combine the ARM surface observations with the four external data sources (OK mesonet, KS mesonet, WPDN surface observations, and the NWS surface observations). The combined data set will contain hourly and 30 minute averages of meteorological quantities and will improve the accessibility of the data to the end user. By evaluating the data with ncql before including it in the integrated file, invalid data

Table 1: Sample output for file containing one day's WPDN surface hourly observations

Filename	=	Dsgp60wpdnsurfX1.a1.950219.000000.cdf
Date of ncql analysis	=	Wed Aug 16 20:35:31 1995
Attribute Base Time	=	19-Feb-95,0:00:00 GMT
Filename Exp. Start	=	19-feb-95, 00:00:00 GMT
Actual Start	=	19-feb-95, 00:00:00 GMT
Expected End	=	19-feb-95, 23:00:00 GMT
Actual End	=	19-feb-95, 23:00:00 GMT
Expected Interval (sec)	=	3600
First Difference (sec)	=	3600
Number of Gaps	=	0

Variable Names

Short Name	Units	Long Name
lat_stn	Degrees	Station latitude
lon_stn	Degrees	Station longitude
alt_stn	meters	Station altitude
dp	Degress Kelvin	Surface dew point temperature
qc_dp	qc_code	quality control byte for dew point temperature
rh	Percent	Surface relative humidity
qc_rh	qc_code	quality control byte for relative humidity

Minimum Values for Each Variable

Variable	Valid Min	Actual Min	Time	Station ID
dp	230.00	261.70	00:00:00	5
qc_dp	0.00	0.00	00:00:00	6
rh	0.00	12.00	00:00:00	5
qc_rh	0.00	0.00	00:00:00	6

Maximum Values for Each Variable

Variable	Valid Max	Actual Max	Time	Station ID
dp	330.00	289.70	22:00:00	9
qc_dp	6.00	4.00	19:00:00	12
rh	100.00	100.00	15:00:00	0
qc_rh	6.00	4.00	19:00:00	12

Number of Values Out of Valid Range

Variable	Above Max	Below Min	Missing Value Codes	
			[-9999]	[-99999]
dp	0	0	0	23
qc_dp	0	0	0	0
rh	0	0	0	23
qc_rh	0	0	0	0

Missing Gaps

Variable	Plat. Num.	Time Start	Time End	Missing Type
dp	0	11:00:00	14:00:00	-99999
dp	9	01:00:00	19:00:00	-99999
rh	0	11:00:00	14:00:00	-99999
rh	9	01:00:00	19:00:00	-99999

can be immediately flagged and the resultant data will have a higher level of quality assurance.

4. SUMMARY

Ncql fills a niche in the ARM program by enabling users to quickly appraise the contents of a file. Moreover, the utility is general enough to accept a variety of netCDF files making it useful to a wider audience. Please contact the corresponding author to obtain the source code and executable files.

5. ACKNOWLEDGEMENTS

This research was supported by the Environmental Sciences Division of the United States Department of Energy (under Contract No. DE-AC02-76CH00016) under the auspices of the Atmospheric Radiation Measurement Program.

Amanda Staudt performed this work while a participant in the Department of Energy's Office of Science Education and Technical Information, Office of Energy Research, Science and Engineering Research Semester (SERS) Program.

6. REFERENCES

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson and M. D. Eilts. "The Oklahoma Mesonet: a Technical Overview" *J. Ocean & Atmos. Tech.*, 1994.

Jenter, H. L. and R. P. Signell. "NetCDF: A Public-Domain-Software Solution to Data-Access Problems for Numerical Modelers" Proceedings of the American Society of Civil Engineers Conference on Estuarine and Coastal Modeling. Tampa, Florida, 1992

National Weather Service and the Office of Oceanic and Atmospheric Research, August 1994: Wind Profiler Assessment Report and Recommendations for Future Use 1987-1994, Silver Spring, MD, 141 pp.

Rew, R. and S. Emmerson. "NetCDF Operators and Utilities" Jan. 30, 1992.

Stokes, G. M. and S. E. Schwartz. "The Atmospheric Radiation Measurement (ARM) Program: Programmatic Background and Design of the Cloud and Radiation Test Bed" *Bull. Amer. Meteor. Soc.*, 75, July 7, 1994: 1201-1221.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government 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 United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.