Climate Research Data Tools (CRDTools) and COADS

Donald R. Mock
Climate Monitoring and Diagnostics Laboratory
NOAA Environmental Research Laboratories
Boulder, Colorado 80303

Anne R. Messenger
Cooperative Institute for Research in Environmental Sciences
University of Colorado
Boulder, Colorado 80309

CRDtools is a software environment for scientific workstations that aids the scientist in performing many common data access, analysis, and visualization tasks. Originally developed within NOAA's Climate Research Division to help local scientists with their exploratory examination of the COADS 2°x2° monthly mean data set, it has now been expanded to work with many other gridded data sets and can be downloaded across the Internet to serve the needs of climate researchers across the country and around the world. In addition, many of the data sets accessible by CRDtools are also available across the network in an anonymous-FTP archive on noaacrd.colorado.edu (128.138.218.1).

1. Introduction

The Climate Research Data Center (CRDC) project is a NOAA Climate and Global Change (C&GC) initiative charged with assimilating new technologies to better serve the data access and analysis needs of the C&GC research community. This goal is being approached along two avenues: 1) a software data extraction, visualization, and preliminary-analysis package is being developed that will work across a broad spectrum of high-powered, but inexpensive, UNIX computer workstations; and 2) an archive of commonly used data sets for climate research is being made available through the Internet, a TCP/IP-based international computer network that is now available at most educational and research institutions throughout the United States.

2. Software

CRDtools (Climate Research Data tools), the software package under development by the CRDC, is designed to operate primarily on the end-user's computer system, not on the CRDC's central host. The software conforms to the OPEN LOOK Graphical User Interface (GUI) and uses the XView toolkit, developed by Sun Microsystems, Inc. and available in source form as part of the MIT X 11 distribution. As an X 11 application, CRDtools can run as an X 11 client process on a Sun or DEC workstation, while the X 11 display server can be a PC-compatible, Macintosh, X-terminal, or another brand of Unix workstation. Adequate response time has been demonstrated where the client and display workstations were separated by 1500 miles.

CRDtools currently includes tools for calculating, displaying, and printing long-term and seasonal means, anomalies, correlations, power spectra, and time-series from a number of gridded data

sets. Both black and white contour plots and color raster images can be used to visualize planar fields. Hovmöller diagrams provide an alternative means of visualizing the time series of a variable at a specified latitude or longitude. Point values and cross-sectional views of any raster image can be obtained interactively with click-and-drag techniques. Data fields can be arithmetically combined to create new fields, and multiple fields can be visualized simultaneously. A user who is unfamiliar with a given data set can undertake a considerable preliminary analysis of the data without resort to traditional programming techniques.

The extraction capabilities of CRDtools allow for multiple simultaneous extractions of subsets from the compressed data archive. A user merely specifies all of the desired spatial and temporal characteristics for each desired variable and the software makes a single pass through the data archive to perform the extraction. A data dictionary simplifies the process of identifying the available data sets in the archive and can make use of both local and remote nfs-mounted file systems. Once extracted, the resulting subsets are in a form that can be directly used by other CRDtools modules or by the end-user's FORTRAN and C programs.

The CRDtools interface relies heavily on pop-up menus and flexible edit fields to reduce user memorization and to speed the data extraction and display process. On-line information windows, context-sensitive help, and automatic defaults assist the novice user to become quickly proficient in the use of CRDtools, without the need of a traditional hardcopy version of a user's guide. Program feedback to the user, in the form of status messages and percent-completion sliders, during cpu-intensive and I/O-intensive tasks, give assurance that something is really happening. Confirmation windows and error message windows steer the user in the right directions and prevent the loss of valuable work.

Figure 1 illustrates some of the features of CRDtools. In this screen dump from a Sun 4/280 console, a user has called up the Autoex window from the main CRDtools window. With this tool the user has extracted the monthly COADS sea surface temperatures for the period 1980-1989 for the entire globe. Then, using the Image window, only the average January sea surface temperatures over the tropical Pacific have been plotted. The point value at 15°N, 155°E was determined by clicking the mouse button while the cursor was over the image. In reality, of course, this image is in color. A more detailed description of the history, structure, and philosophy of CRDtools is contained in Messenger and Mock (1992).

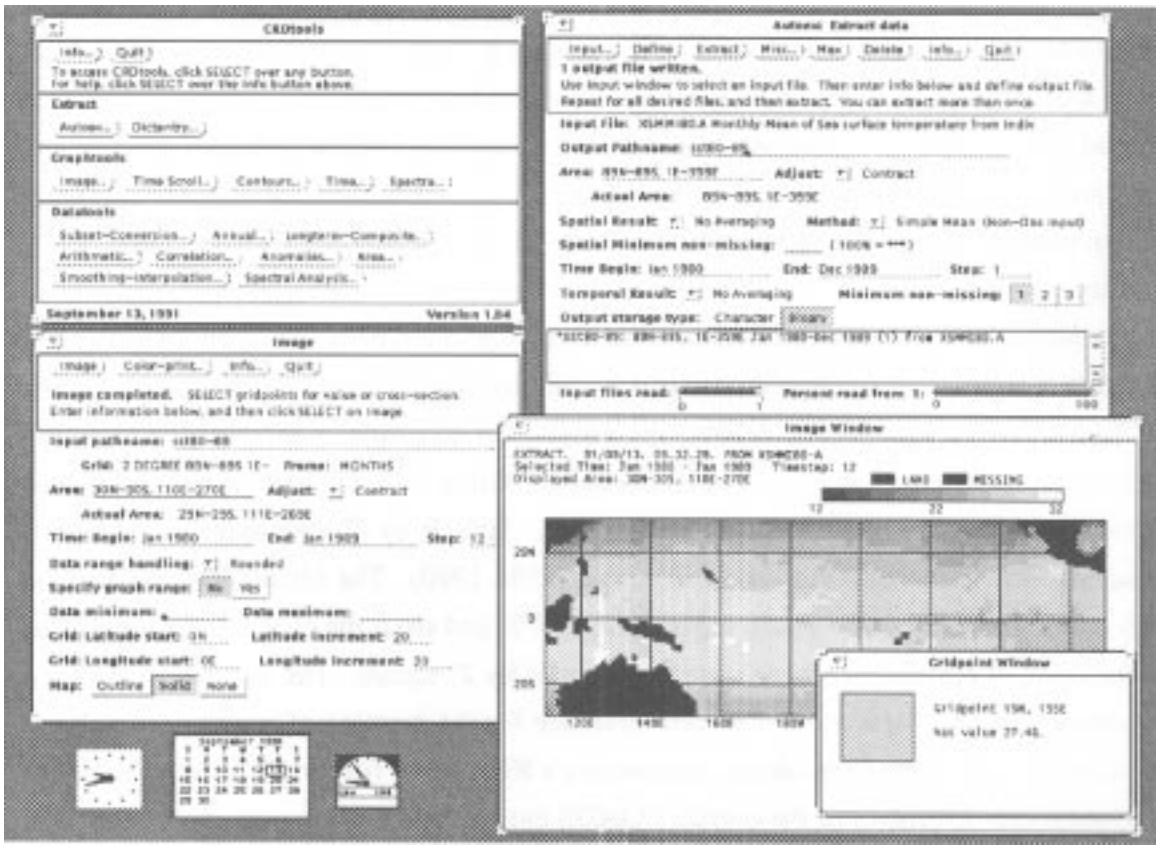

Figure 1. Screen dump from a Sun 4/280 console illustrating some major features of CRDtools, including the main application window (upper left), a data extraction window (upper right), an image window (lower left), and an example raster image of mean January sea surface temperatures for 1980-89 over the tropical Pacific (lower right).

3. Dataset Access

Complementary to the CRDtools software development effort, an on-line archive of climate data sets is being assembled that can fully take advantage of CRDtools' capabilities and be readily downloaded to remote users across the Internet. End users may gain access through anonymous-FTP (File Transfer Protocol) to both the data sets and to the binary and source versions of the CRDtools software. One of the primary purposes of this archive is to significantly increase the accessibility of commonly used climate data sets to the research community.

Access to the data set and software archive is founded on the principle that things should be kept simple. Therefore, the archive is straightforward and self-describing as much as possible. Each directory and subdirectory of the archive has a **README** file that can be retrieved by the remote FTP user for additional information or instructions before pursuing the more time-consuming transfer of data or software. The **Software** subdirectory contains binary and source versions of CRDtools for all supported computer architectures. These are in a compressed tar format to save time during transmission. After unpacking at the end user's site, the resulting set of files include instructions for completing the installation. While a binary installation is most desirable,

differences in various operating system, windowing system, and compiler releases make a source installation necessary in a majority of cases.

The **Datasets** subdirectory contains a number of gridded data sets commonly used in climate research. Each of these data sets is broken up into a number of files, typically one for each primary variable or related quantity. For example, beneath the **Datasets** subdirectory is another subdirectory called **coads**. Within this subdirectory are thirty-nine related files. Ile file **sst.mean** contains the entire 2° latitude by 2° longitude monthly mean sea surface temperature for the period of record (1854-1990). The file **air.ltm** contains the 2° by 2° air temperature long term mean (1950-1979) and **slp.nobs** contains the number of observations of sea level pressure per month per 2° by 2° square. The purpose of the long term mean files is to serve as a standard reference for the creation of anomalies. Table. 1 illustrates the contents of the **coads** subdirectory's **README** file, which briefly describes the general characteristics of the overall COADS data set and the contents of specific files. In addition, each data file is self-describing, i.e., each file has an internal header which contains metadata such as the type of parameter, level of observation, spatial and temporal range and resolution, and original source of the data. The CRDtools software can be used to display the contents of the internal header for any data file.

Table 1. Contents of the COADS subdirectory README file.

```
COADS (Comprehensive Ocean-Atmosphere Data Set)  11/21/91

2x2 monthly means January 1854 - December 1990
2x2 monthly numbers of observations January 1854 - December 1990
2x2 monthly long term means 1950 - 1979

All datasets 89N-89S, 1E-359E

Observed variables:
  Air temperature (A)                        air.mean,  air.nobs,  air.ltm
  Sea level pressure (P)                     slp.mean,  slp.nobs,  slp.ltm
  Sea surface temperature (S)                sst.mean,  sst.nobs,  sst.ltm
  Vector wind eastward component (U)        uwnd.mean, uwnd.nobs, uwnd.ltm
  Vector wind northward component (V)       vwnd.mean, vwnd.nobs, vwnd.ltm
  Scalar wind (W)                           wspd.mean, wspd.nobs, wspd.ltm
  Cloudiness (C)                            cldc.mean, cldc.nobs, cldc.ltm
  Specific humidity (Q)                     shum.mean, shum.nobs, shum.ltm

Derived variables:
  Relative humidity (R)                     rhum.mean, rhum.nobs, rhum.ltm
  Sensible heat parameter (E = (S - A)W)    sflx.mean, sflx.nobs, sflx.ltm
  Latent heat parameter (G = (QS - Q)W),
          where QS = saturation Q at S      lflx.mean, lflx.nobs, lflx.ltm
  U-wind stress (X = WU)                    ustr.mean, ustr.nobs, ustr.ltm
  V-wind stress (Y = WV)                    vstr.mean, vstr.nobs, vstr.ltm
```

The entire, COADS data set of thirteen variables in thirty-nine files, covering a period of 137 years, takes up approximately 275 megabytes of disk space. Other data sets of interest in the archive include monthly mean summaries of NMC analyses and ECMWF analyses. Additional data sets are being added at regular intervals as disk space and staffing permit. In actual practice, data files can be transmitted via FTP at the rate of one megabyte per three or four minutes of clock time, even across the country on a multi-hop route. While not appropriate for large data sets at this time, this performance is quite satisfactory for many small- and medium-sized data sets. In time, and with increasing network performance, network file transfers or even direct mounting of remote file systems may completely replace more traditional tape distribution methods.

The activity of “anonymous” FTP users is recorded, both as a security precaution and as a log of data set usage. A modified version of the BSD4.3 source code for the FTP daemon was acquired and tailored for CRDC needs. When a remote user accesses the “anonymous” FTP account, the originating IP (Internet Protocol) number and hostname is logged as well as the given password, which as a matter of courtesy is usually the user’s actual e-mail address. In addition, any “put” or “get” commands issued by the user are recorded. This allows notification of “anonymous” end users of any significant corrections or revisions to the software or data that they retrieved in previous sessions. The data host for this “anonymous” FTP account is currently located on noaacrd.colorado.edu (128.13 8.218. 1).

4. Conclusions

As the scientific research community quickly moves into an age where inexpensive desktop workstations approach the effective capabilities of yesterday’s supercomputers, data access, analysis, and visualization software will become an increasingly important facilitator of scientific research. Ever increasing data volumes will require sophisticated software tools to aid in the necessary data reduction tasks. COADS has served as a prototype of a large volume data set for the development of CRDtools and will continue to be an important focal point for the continued development of new capabilities within the Climate Research Data Center project.

5. Acknowledgements

The authors would like to express their appreciation to the staff of both the Climate Research Data Center and the Climate Monitoring and Diagnostics Laboratory of NOAA’s Environmental Research Laboratories in Boulder, Colorado. Funding for this project has been received from NOAA’s Climate and Global Change Program Office. Additional technical assistance has been received from the University of Colorado’s Cooperative Institute for Research in Environmental Sciences, the University of Washington’s Joint Institute for the Study of Atmosphere and Ocean, the National Weather Service’s Climate Analysis Center, the National Climatic Data Center, the National Geophysical Data Center, the National Center for Atmospheric Research, and the National Center for Supercomputer Applications. The modified BSD4.3 FTP source code is courtesy of Marty Ryba of Princeton University.

6. References

Messenger, A.R., and Mock, D.R., 1992: Climate Research Data Tools (CRDtools): Data Visualization and Analysis Software for Climate Researchers. Preprints, *Eighth International Conference for Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology,* Atlanta, AMS, 162-169.