CLIMATE DATA ANALYSIS TOOL: AN OPEN SOFTWARE SYSTEM APPROACH.

J2.1

Dean N. Williams*, Robert S. Drach, Paul F. Dubois, Charles Doutriaux,
Charles J. O'Connor, and Krishna M. AchutaRao, Michael Fiorino
Program for Climate Model Diagnosis and Intercomparison (PCMDI)
LLNL, Livermore, CA

1. INTRODUCTION

This document reviews and updates the current status and future development plans of the PCMDI software system commonly known as the Climate Data Analysis Tools (CDAT).

We focus on hardware, network, software infrastructure, tool development, and support needed for PCMDI research, for LLNL's Atmospheric Science Division (ASD) chemistry research, and for other national and international research institutions.

2. CDAT

The Climate Data Analysis Tools (CDAT) is a software infrastructure that uses an object-oriented scripting language to link together separate software subsystems and packages, and thus forms an integrated environment for solving model data analysis problems. The power of the system comes from the ability of the Python system to seamlessly interconnect disparate and specialized data access and display software within a uniform user interface. The Python system provides a general purpose and full-featured scripting language with a variety of user interfaces including command line, stand-alone scripts (e.g., applications) and graphical user interfaces (GUI). The CDAT subsystems, implemented as Python modules, provide access to and management of geo-referenced gridded data (Climate Data Management System or CDMS); large-array numerical operations (Numeric Mask Arrays or MA); and visualization (Visualization and Control System or VCS).

3. PYTHON

Python is an open-source object-oriented scripting language and users find it to be simple to use, dynamic, powerful, and flexible. Serving as an excellent "glue" or "steering" language, Python has the power to integrate unrelated software packages under a single system. This modular structure will be important as we extend the framework to provide a grid-enabled data analysis system. We are leveraging our infrastructure efforts with a national and international community of contributors.

4. PURPOSE OF CDAT

One of the greatest challenges faced by climate researchers is the cataloging and analysis of very large multi-dimensional global atmospheric and oceanic model data sets. To reduce the labor-intensive and time-consuming process of data management, retrieval, and analysis, PCMDI, other DOE sites and DOE-supported programs have come together to develop an intelligent data management system for linking of storage devices located throughout the United States and the international research community. This effort, headed by PCMDI, NCAR, and ANL will allow users anywhere to remotely access this distributed multi-petabyte archive and perform analysis.

PCMDI's CDAT is an innovative system that supports exploration and visualization of climate scientific datasets. As an "open system", the software sub-systems (i.e., modules) are independent and freely available to the global climate community. CDAT is easily extended to include new modules, and as a result of its flexibility, PCMDI has been able to integrate CDAT with other popular software systems such as the Live Access Server (LAS) and the Distributed Oceanographic Data System (DODS). Together with ANL's Globus middleware, CDAT's focus is to allow climate researchers the ability to access and analyze multi-dimensional climate datasets located at various DOE sites.

* *Corresponding author address*: Dean N. Williams,
PCMDI, Lawrence Livermore National Lab,
P.O. Box 808, Mail Stop-264, Livermore, CA 94550
Williams13@llnl.gov, www-pcmdi.llnl.gov/software
UCRL-ID145189

5. CURRENT STATUS OF CDAT

CDAT consists of tools that simplify the storage, diagnosis, and visualization of climate data. These tools play a crucial role in the development, testing, validation, and intercomparison of global climate models. The current version of CDAT (i.e., CDAT version 3.0) includes seven such tools:

- The Climate Data Management System (CDMS) module – automatically locates and extracts meta-data (for example, variables, dimensions, grids, attributes, etc.) from PCMDI's collection of model runs and analysis files and provides a uniform data access interface to datasets in all conventional formats used in the climate community (e.g., netCDF, and GRIB/GrADS);
- Contrib – Collection of modules useful to the climate community, such as: statistical routines, NCAR's SpherePak, NCAR's RegridPak, averaging routines, etc.
- The Numeric Mask Array (MA) module – manipulation/analysis of large arrays that include missing data;
- Pyfort – generates Python modules from FORTRAN sub-routines;
- Python – the object-oriented interpreter and scripting and programming language;
- The Visual Climate Data Analysis Tools (VCDAT) –GUI tool that interacts with Python and CDAT sub-systems; and
- The Visualization and Control System (VCS) module – displays, animates, and manipulates generated graphics.

In mid-May 2001, PCMDI released version 3.0 of CDAT. Since its release, there have been 500+ downloads of the system and related documentation. In September 2001, PCMDI will release version 3.1 of CDAT. This version will include: 1) additional scientific modules (e.g., Mudpak and regridders); 2) new VCS graphics methods for non-rectilinear grids; 3) bug fixes and new features for CDMS, MA and VCS; 4) new VCDAT capabilities; and 5) upgrade to Python 2.2.

CDAT-3.1 will be a source distribution with standard and robust configuration and installation procedures. The system will be built and fully tested on the following Unix platforms: Linux (SuSE 7.0/7.1, RedHat (6.2,7.1) and Mandrake (7.2)); SGI (IRIX 6.5); DEC alpha (OSF 4.0); IBM AIX (3.4); SUN (Solaris 5.6). (Jean-Yves Peterschmitt from Laboratoire des Sciences du Climat et de l'Environnement (LSCE) recently ported CDAT-3.0 to the DEC Alpha platform and has been an important contributor to the improved installation system.)

We have built CDAT with DODS so that users can access data over the Internet via a URL web address. In addition, PCMDI's developers have developed an alpha version of the Live Access Server (LAS) interfacing to CDAT. LAS was developed at the Pacific Marine Environmental Laboratory (PMEL) and will provide a web browser which interacts with CDAT to display the AMIP and CMIP datasets. The user will be able to select datasets, regions, plots, output, and animation from the web interface.

6. OTHER PCMDI SOFTWARE ACTIVITIES

Processing model intercomparison data (e.g., the Coupled Model Intercomparison Project (CMIP), and the Atmospheric Model Intercomparison Project (AMIP)) and reanalysis data from ECMWF and NCEP is an important PCMDI activity. All members of the PCMDI software team are directly or indirectly involved in this endeavor.

PCMDI software activities focus on the future through involvement with external software research activities, which include the Earth Systems Grid (ESG) and the National Transparent Optical Network (NTON). ESG efforts move PCMDI towards an increased collaboration between groups of climate researchers, in that ESG is designed to support data replication among climate research centers over high-speed transport, and flexible remote access of distributed data analysis. This project addresses coordinated development and integration of advanced applications, middleware, and network services. This effort is impossible without the ability to access high-speed networks, which will enable audio and visual communication, white board interaction, visualization, and animation. PCMDI and the climate research community expect to use NTON when accessing remote databases, formulating queries, transferring data, and visualizing analyzed results.

CDAT is also being extended to support the Atmospheric Chemistry Group in LLNL's Atmospheric Science Division (ASD). Special features and dimensions have already been added to CDAT to accommodate atmospheric chemistry datasets. More work is needed to address the use and access of these datasets, as they are not typical of numerical model data. In addition, PCMDI is involved with the installation and setup of a Power Wall at ASD. The Power Wall will display PCMDI's graphics when VCS is ported over to OpenGL.

7. MAJOR SOFTWARE/CDAT COLLABORATORS

Argonne National Laboratory
Energy and Environmental Sciences at Lawrence Livermore National Laboratory
Lawrence Berkeley National Laboratory
Geophysical Fluid Dynamics Laboratory
Laboratoire des Sciences du Climat et de l'Environnement
National Center for Atmospheric Research
Information Sciences Institute, University of Southern California
University of Reading, UK
Hadley Centre for Climate Prediction and Research
Max-Planck Institute for Meteorology

8. HARDWARE FOR CDAT DEVELOPMENT

Hardware acquired in the past month (i.e., August 2001) for CDAT development included four Dell processor PC workstations.

The following hardware is still needed for development:

- Dell PC workstation for PCMDI's CDAT developer; and
- ½ TB disk space for ESG-II development, to be placed outside firewall

9. NEAR-TERM PLANS FOR CDAT

PCMDI will continue to develop the new ESG infrastructure that in turn will require development and exploration of new computational tools, development tools, diagnostic libraries and visualization to advance the concept of distributed climate centers. Through this infrastructure, climate scientific collaboration will become independent of geographic and organizational boundaries and will improve the ability to solve complex scientific climate problems.

Planned extensions to the CDAT components applications include:

- CDMS – non-rectilinear grids; CDSCAN/XML interface, enhancements; bug fixes;
- MA – additional functions; upgrade to the latest version of Numeric;
- VCS – new graphics methods; enhancements; port to OpenGL;
- VCDAT – new features; bug fixes; additional GUI front-end to modules (i.e., template, graphics methods, page description panel, etc.);
- INSTALLATION – source-based distribution; simplified/robust installation; port to more Unix platforms;
- Diagnostics – develop more basic diagnostics capabilities; port compall to CDAT;
- CDAT/LAS – add more graphics; choose colormap; and select product output (i.e., gif netCDF, postscript, cgm, animation, metadata); size of window; comparisons, and selection.
- LAS server-side analysis;
- LAS service of AMIP I and PCMDI observational datasets; and
- CDAT/ESG-II – continue to incorporate ESG development into CDAT.

10. AKNOWLEDGEMENTS

This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.