

The Package for Analysis and Visualization of Environmental Data

Steve Thorpe,¹ John Ambrosiano,¹ Rajini Balay,² Carlie Coats,¹ Alison Eyth,¹ Steve Fine,¹ Dan Hils,¹ Ted Smith,¹ Atanas Trayanov,¹ Tim Turner,³ and Mladen Vouk²

¹ MCNC Environmental Programs, Research Triangle Park, NC, USA

² North Carolina State University, Computer Science Department, Raleigh, NC, USA

³ Turner Engineering, Chapel Hill, NC, USA

ABSTRACT: *We have developed the Package for Analysis and Visualization of Environmental data (PAVE), a flexible and distributed application to visualize multivariate gridded environmental datasets. Design goals included (1) baseline graphics with the option to export data to high-end commercial packages, (2) access and manipulation of datasets located on remote machines, (3) support for multiple simultaneous visualizations, (4) an architecture that allows PAVE to be controlled by external processes, (5) low computational overhead, and (6) no software distribution cost. Supported platforms include DEC, HP, IBM, SGI, and Sun workstations. Remote data can also be read from CRAY supercomputers.*

1 Background

The *Environmental Decision Support System* (EDSS) is a large software project of the Environmental Programs Group at MCNC. EDSS originated as part of a cooperative agreement with the U.S. Environmental Protection Agency's (EPA's) Office of Research and Development in order to help the agency develop the next-generation air quality modeling system [Ambrosiano et al., 1995; Ambrosiano, 1996]. Although models run within EDSS have typically been air quality applications, the system is extensible to other problem domains such as water quality simulations.

Previous MCNC visualizations have often used high-end data flow environments. Advanced Visual System Inc.'s *Application Visualization System* (AVS) [Upson et al., 1989] has been extremely useful for creating high-end visualizations such as 3-D animating isosurfaces and volume renderings, and for easily creating custom visualizations for unique data types. For example, MCNC's successful *UAMGUIDES* system [Pua, 1994] was built on top of AVS and the X/Motif Toolkits to assist modelers in the complex task of running and visualizing the results of EPA's *Urban Airshed Model* (UAM).

Today's popular data flow environments have several limitations, however. Their user interfaces do not lend themselves to

easily creating and managing more than three or four separate visualizations simultaneously. These packages tend to require very large amounts of memory for each visualization created, as data can be duplicated several times as they flow through the system. And finally there is the issue of distribution cost for these and most commercial systems—each user must purchase a license in order to run an application built on top of it. This restricts our ability to get such software out to the environmental community.

We were unable to find an off-the-shelf solution to meet all of our diverse requirements for analysis and visualization. We need to analyze large multivariate datasets that are often located on different machines, geographically register map backgrounds with modeled data, and handle numerous data subsetting and manipulation tasks. We also want to animate many windows simultaneously on machines with small memories and no specialized graphics hardware.

The EDSS project was developed under EPA's High Performance Computing and Communications initiative, whose requirements analysis [Coats et al., 1992] concluded that modelers often need immediate response with simple interactive visualizations (tile plots, line plots, etc.). The delays associated with more complex and high-end visualizations (3-D animating

isosurfaces, volume renderings, etc.) can distract the user during data analysis tasks.

MCNC decided to address some of these issues by developing the EDSS software component called the *Package for Analysis and Visualization of Environmental data* (PAVE), a flexible and distributed application used to visualize a variety of multivariate gridded environmental datasets [Thorpe, 1996]. Design goals included the following:

- Baseline visualization capabilities with the option of exporting data to higher end commercial packages such as AVS
- Easy access and manipulation of datasets located on remote machines
- A user interface that easily supports multiple simultaneous visualizations
- An architecture that allows PAVE to be controlled by external processes
- Low computational overhead
- No software distribution cost

2 PAVE Design

2.1 Low Computational Overhead

PAVE's Motif-based user interface and internal data structures were built using custom-developed C and C++ code. When compared to off-the-shelf packages, this code development and maintenance adds considerably to the effort required to create a visualization system. However, the data flow paradigm's high computational overhead is avoided. PAVE has simultaneously animated eight plots on a Sun Sparcstation with only 16 MB of memory, for example. The system response time was adequate to get useful information from the data, a feat that would be impossible using typical commercial data flow environments on a machine with so little memory.

2.2 Access to Remote Data

EDSS model executions and datasets may reside on a variety of platforms, including CRAY C90, T90, and Y-MP supercomputers and DEC, HP, IBM, SGI, and Sun workstations. A typical EDSS model result might consist of about 40 variables, 15 layers, 35 rows, 32 columns, and 72 time steps, for a total dataset size of approximately 185 MB. Work is currently under way on models using larger grids (30 layers by 75 rows by 75 columns) that will create datasets approaching 2 GB in size. It is often not feasible to move such large datasets to a visualization workstation, so PAVE was designed to access datasets from any of the above platforms remotely (see Figure 1).

Remote "daemon" processes are launched when necessary to allow PAVE users to browse and select remote datasets to be visualized, and to read the raw data. Interprocess communication is handled by the EDSS software bus [Balay and Vouk, 1996], a socket-based library that enables messages and data to be passed among distributed processes.

The PAVE data reader daemons extract only the information needed to create the visualization requested. For example, to visualize ground-level ozone for day one of the simulation described above, only 107 KB (1 variable \times 1 layer \times 35 rows \times 32 columns \times 24 time steps \times 4 bytes/float) would be extracted and shipped over the network. This subselection allows networks to be used efficiently and provides faster system response times.

PAVE reads files in the EDSS/Models-3 IO/API [Coats, 1995] formats. The IO/API library is built on top of the netCDF standard to provide cross-platform portability and the selective access described above. UAM-IV and UAM-V file formats are also supported to allow users of these models to utilize PAVE's feature set.

2.3 Flexible Data Manipulation

PAVE was designed to enable easy operations on and among variables stored in the datasets. Each variable in a dataset can be used as an operand in formulas entered by the user, and derived

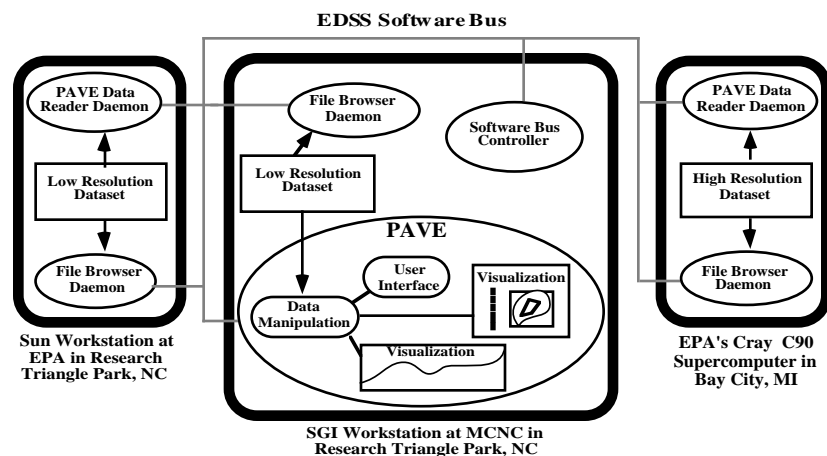


Figure 1: Schematic of a PAVE session accessing remote and local datasets. The software bus handles the data and message passing between processes.

variables will be calculated. For example, a user can calculate the sum of the variables NO and NO₂ from dataset “a” by simply entering the formula “NO_a+NO_{2a}”. Calculating the ratio between ozone from two different datasets (denoted by “a” and “b”) is as simple as entering “O_{3a}/O_{3b}”. The datasets in a formula need not reside on the same machine. There are a number of mathematical operators in PAVE’s formula parser, such as sin, cos, abs, and tan.

The user interface allows data cropping along any dimensions—*x*, *y*, *z*, and time. Slider widgets handle the *z* and time cropping, while a domain selection box allows a user to click-drag over their geographic region of interest in the *x-y* plane (see Figure 2). After the data have been suitably “sliced and diced,” they are then passed to the visualization components described in Section 3.

2.4 Integration with Other Systems

PAVE was not designed as a complete stand-alone solution for all users. Therefore our goal was not only to make all of its features accessible through its Motif user interface, but also to make them available to external processes that can talk to PAVE with command line arguments, commands sent to PAVE’s standard input stream, or messages passed using the EDSS software bus.

Data can be exported as tabbed ASCII data suitable for use by spreadsheet applications, as AVS fields, and as the IO/API format mentioned above.

3 The Visualizations

Visualization renderings are done by a combination of custom-developed X/Motif library calls and several public domain tools. PAVE accesses the following public domain software as needed: Tcl/Tk, BLT, and PLplot. Tcl/Tk provides a way to easily develop graphical user interfaces. BLT provides graphing capabilities and is built on top of Tcl/Tk. PLplot is a basic scientific plotting package that also uses Tcl/Tk. Because we utilize only noncommercial software, end users do not have to pay any licensing fees to use PAVE.

Visualization types available include tiled and smoothed horizontal and vertical cross sections (see Figures 3 and 4), wind vector plots (see Figure 5), time series (see Figures 6 and 7), scatter plots (see Figure 8), and data values projected into a time-animated 3-D mesh (see Figure 9). Any number of cross-section plots can be animated synchronously to compare various datasets and variables. MPEG, Gif, RGB, and other image formats can be exported.

4 Results and Lessons Learned

There is a growing PAVE user base. Users have come from MCNC, EPA, a number of state air quality agencies, and a variety of contractors in the environmental community.

During the October 1995-September 1996 period, there were over 4000 logged runs (developer sessions not included) by more than 50 users on 48 different machines. The average duration of a PAVE session has been just shy of 44 minutes. Usage

by workstation platform has been 36% DEC, 2% HP, 30% SGI, and 32% Sun. The platforms chosen have been due mostly to the machine types available on the individual users’ desks, rather than to one architecture performing better than another, as PAVE runs well on all of the above platforms.

Although PAVE does not have built-in high-end capabilities such as animated isosurfaces and volume renderings, it has been useful in the community. We have found that users frequently take advantage of basic easy-to-use visualization techniques even with high-end 3-D packages. Therefore, most of a typical user’s visualization needs can be satisfied by these home-grown and public-domain tools that have no distribution cost.

Our work has shown the importance of decoupling the data manipulation (i.e., “slicing and dicing”) from the rendering portions of a visualization system. In this way we are free to easily replace the public domain rendering tools as new ones become available.

There has been much positive feedback and many suggestions for future development from the user community. As a direct result, there are several main goals that we hope to address:

- Comparisons between scattered observations and gridded models
- More efficient memory management and CPU usage
- A more intuitive user interface, possibly including a World Wide Web interface

5 Conclusions

We have developed a flexible, powerful, and easy-to-use analysis and visualization system for multivariate gridded environmental datasets. Our system meets a large portion of our scientific modelers’ visualization needs without resorting to commercial software. Initial response from users has been positive and we expect to continue development based on this feedback.

6 Acknowledgments

Development of this software and paper was supported in part through the EPA-MCNC cooperative agreement number CR822066.

Special thanks are due to John Ousterhout for Tcl/Tk (©1987-1993, the Regents of the University of California), to Maurice LeBrun and Geoffrey Furnish of the University of Texas for PLplot (portions ©1987-88 by Digital Equipment Corporation and the Massachusetts Institute of Technology), to the University Corporation for Atmospheric Research/Unidata for netCDF (©1993), to AT&T Bell Laboratories for BLT (©1993-4), to Todd Plessel and Mark Bolstad for the MapUtilities library (©1996 Lockheed/Martin Technical Services), to the United States Geological Survey for the PROJ library (©1995), and to Doug Young for several example Motif Widgets (©1992, Prentice Hall).

The authors gratefully acknowledge the advice and assistance of the following individuals: Kiran Alapaty, Ed Bilicki, Christine Bullock, Mike Clark, Jeanne Eichinger, Ken Galluppi, Adel Hanna, Elizabeth Hayes, Marc Houyoux, Dongming Hwang, Clint Ingram, Carey Jang, Hassan Karimi, Prasad Kasibhatla, Rohit Mathur, John McHenry, Talat Odman, Don Olerud, Kathy Pearson, Eng Pua, Uma Shankar, Jeff Vukovich, Dick Watkins, and Aijun Xiu.

7 References

Ambrosiano, J., R. Balay, C. Coats, A. Eyth, S. Fine, D. Hils, T. Smith, S. Thorpe, T. Turner, and M. Vouk, "The Environmental Decision Support System: Air Quality Modeling and Beyond," *Proceedings of the US EPA Next Generation Environmental Modeling Computational Methods (NGEM-COM) Workshop*, Bay City, MI, August 7-9, 1995.

Ambrosiano, J., "The Environmental Decision Support System," <http://www.iceis.mnc.org/EDSS/EDSSPage.html>, MCNC Environmental Programs, RTP, NC, 1996.

Balay, R. and M. A. Vouk. "A Lightweight Software Bus for Prototyping Problem Solving Environments," Accepted for the Special Session on Networks and Distributed Systems in the *Eleventh International Conference on Systems Engineering*, Las Vegas, 1996.

Coats, C. J., Jr., J. Lear, M. Matthews, "System Requirements Specification for EPA's Third Generation Modeling System (Models-3)," internal report produced by Computer Sciences Corporation under Contract No. 68-W0-0043-462, U. S. Environmental Protection Agency, RTP, NC, 1992.

Coats, C. J., Jr., "The EDSS/Models-3 I/O Applications Programming Interface," <http://www.iceis.mnc.org/EDSS/iaoapi/H.AA.html>, MCNC Environmental Programs, RTP, NC, 1995.

Pua, E., "UAMGUIDES: Urban Airshed Model With Graphical User Interface and Decision Support," International AVS Users Group Conference Proceedings, Boston, MA, pp. 199-212, 1994.

Thorpe, S., "PAVE User Guide," http://www.iceis.mnc.org/EDSS/pave_doc/Pave.html, MCNC Environmental Programs, RTP, NC, 1996.

Upson, C., T. Faulhaber, D. Kamins, D. Laidlaw, D. Schlegel, J. Vroom, R. Gurwitz, and A. vanDam, "The Application Visualization System: A Computational Environment for Scientific Visualization," *IEEE Computer Graphics and Applications*, vol. 9, no. 4, pp. 30-42, 1989.

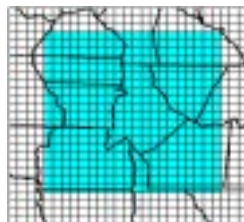


Figure 2: Domain selection dialog allows users to select their geographic region of interest

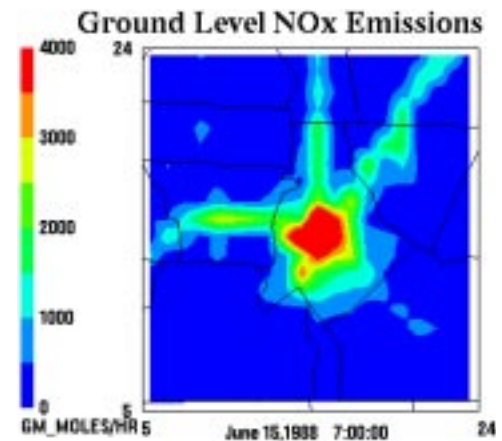


Figure 3: Bilinearly interpolated plot

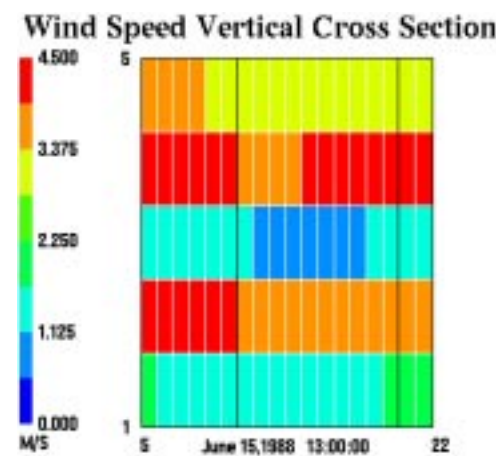


Figure 4: Tile plot

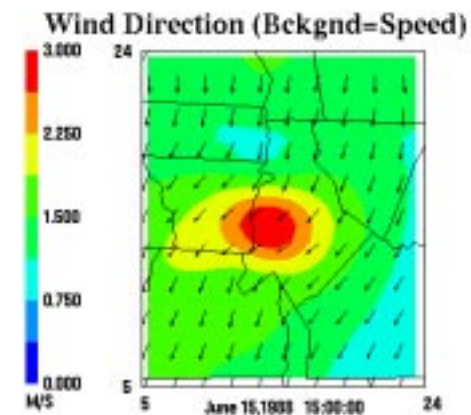


Figure 5: Vector plot

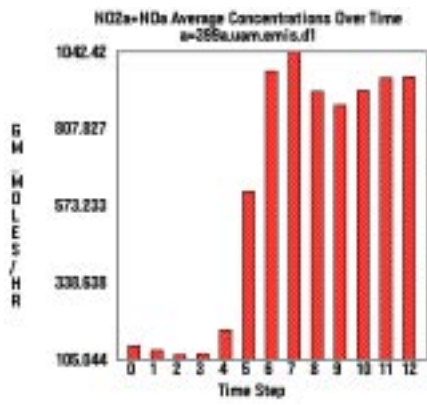


Figure 6: Time series bar chart

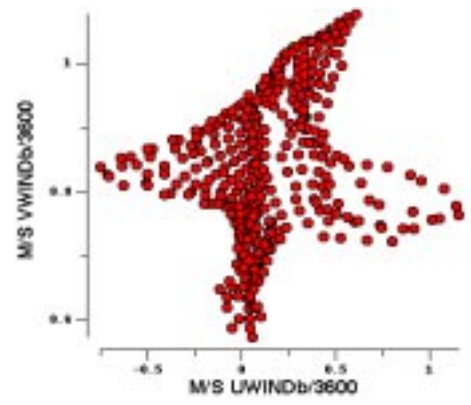


Figure 8: Scatter plot (from BLT)

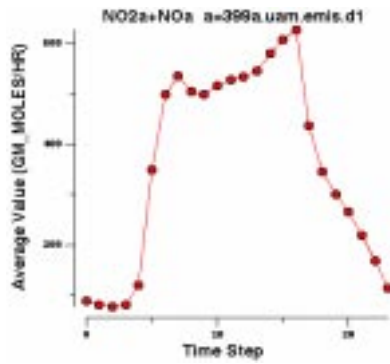


Figure 7: Time series line graph (from BLT)

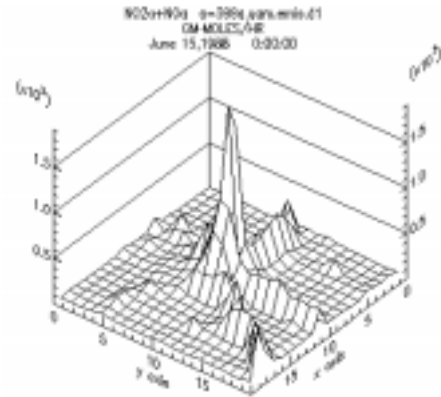


Figure 9: Mesh plot (from PLplot)