A MAP INTERFACE
FOR EXPLORING MULTIVARIATE PALEOCLIMATE DATA

David DiBlase, Catherine Reeves, Alan M. MacEachren, John B. Krygier, Martin von Wyss, James L. Sloan II* and Mark C. Detweiler**

Department of Geography
The Pennsylvania State University
302 Walker Building
University Park PA 16802
Email: dibiase@essc.psu.edu

*Earth System Science Center
Penn State

**Department of Psychology
Penn State

ABSTRACT

We begin with an abbreviated review of recent approaches to the display of three or more geographically-referenced data variables from the literatures of statistics, computer graphics and cartography. In comparison we describe a method we have developed for exploring relationships among multivariate paleoclimate data produced by a global circulation model at Penn State's Earth System Science Center. Key features of the interface include 1) display of multivariate data in two-dimensional, planimetrically-correct map formats; 2) optional display of a single map on which up to three variables are superimposed or four maps among which up to four variables are juxtaposed; and 3) the ability to quickly revise data variable selections, to focus on subsets of the data and to experiment with different combinations of point, line and area symbols.

CONTEXT

This paper discusses methods for displaying three or more geographically-referenced data variables. In particular we are concerned with methods appropriate for exploratory data analysis in Earth system science. Earth system science is an integrative, multidisciplinary approach to understanding “the entire Earth system on a global scale by describing how its component parts and interactions have evolved, how they function, and how they may be expected to continue to evolve on all time scales” (Earth System Sciences Committee 1988). Numerical modeling is an important method of formalizing knowledge of the behavior of the Earth system and for predicting natural and human-induced changes in its behavior. The volume of quantitative data produced by model simulations and earth observing satellites imparts a high value on effective graphical techniques for identifying potentially meaningful patterns and anomalies. Because the Earth system comprises many interrelated phenomena, the ability to display multiple data variables in a format that fosters comparison is particularly desirable.

In the following we will first discuss several recent approaches to the display of three or more variables from the literatures of statistics, computer graphics and cartography. Then we describe an interface we have developed for displaying up to four paleoclimate data sets produced by a global climate model at Penn State's Earth System Science Center. Key features of the interface include 1) display of multivariate data in two-dimensional, planimetrically-correct map formats; 2) op-
tional display of a single map on which up to three variables are superimposed or four maps among which up to four variables are juxtaposed; and 3) the ability to quickly revise data variable selections, to focus on subsets of the data and to experiment with different combinations of point, line and area symbols. In this approach we have attempted to incorporate the three functional characteristics that differentiate cartographic visualization from cartographic communication: the interface is tailored to the needs of a specific set of users, its intended use is more to foster discovery rather than to present conclusions, and it is an interactive environment that encourages users to experiment with different combinations of data and graphic symbols (MacEachren in press). We conclude with a brief discussion of the problem of evaluating the effectiveness of this and other multivariate exploratory methods.

EXPLORATORY MULTIVARIATE DATA ANALYSIS

Most display media for computer-based analysis available to geoscientists and analysts in other disciplines provide two-dimensional images. Primary among these are raster CRTs and paper or film imaging devices. Three-dimensional illusions can be created by employing psychological or physiological depth cues (Kraak 1988). Though 3-D offers higher information-carrying capacity than 2-D, adding a dimension to the depiction also introduces additional cognitive and computational costs. The relative advantages of 2-D versus 3-D depend on the dimensionality of the phenomena to be visualized (and their surrogates, the data variables), and whether relations among different variables are assumed to be orthogonal or parallel. Also pertinent are the hardware and software available to the analyst, the cognitive tasks to be performed and his or her aesthetic preferences.

Two-dimensional approaches One class of 2-D approaches exploits "glyphs"—compound point symbols such as faces, trees, castles, and the "icons" used in the Exvis system developed by Grinstein and colleagues at the University of Lowell (Smith and others 1991). An Exvis display consists of a dense 2-D matrix of strings of short line segments. The orientation of each line segment relative to those connected to it represents a quantity for one of several variables at a given location. Remarkably differentiable patterns can emerge as textural differences in properly "focused" Exvis displays. The system provides user control for the size of each icon, line segment length, line segment orientation limits, and displacement of icons to reduce artifactual patterns.

For analysts willing to suspend the orthogonality of longitude and latitude, relations among many geographic variables can be viewed at once using parallel coordinate diagrams (Bolorforoush and Wegman 1988). In this method each variable is assigned a parallel linear scale. Corresponding cases on each scale are connected by straight lines. Positive and negative correlations, clusters, and modal cases among many variables can be revealed in patterns of parallelism, convergence, and central tendency among the connecting lines. When the number of cases is large, the density of overlapping connecting lines can be interpolated and displayed as isolines (Miller and Wegman 1991). Coordinate scales must often be normalized or transformed in other ways to produce meaningful patterns.
Cartographers and geographers also have demonstrated display methods for multivariate quantitative data on 2-D maps. Most efforts by cartographers to date have been directed to maps intended for presentation rather than exploratory use. Exceptions include Bertin's (1981) “trichromatic procedure" for superimposing three point symbol maps of quantitative data colored with the three printers' primaries (cyan, magenta and yellow) whose combination results in an informative range of hues and lightnesses. Dorling (1992) produced a time series of population cartograms of England superimposed with arrows symbolizing the mixture of alliances to three political parties (by hue and saturation) and the magnitude and direction of voting swings (by length and orientation) in more than 600 constituencies in Parliamentary elections.

**Three-dimensional approaches** Three-dimensional representations are invaluable for portraying volumetric phenomena. Describing the animated stereoscopic display terminals developed for use with the McIDAS weather display system, Hibbard (1986) remarks that "[t]hey create an illusion of a moving three-dimensional model of the atmosphere so vivid that you feel you can reach into the display and touch it." Interactive 3-D displays are also desirable for investigating forms of relationships among non-spatial variables in 3-D scatterplots (Becker and others 1988). For map-based displays, however, 3-D does not necessarily simplify the problem of effectively superimposing three or more data distributions.

In a review article on meteorological visualization, Papathomas and colleagues (1988) note that two approaches to multivariate 3-D visualization have been attempted: “One is to portray each variable by a different attribute [graphic variable] ... another is to assign different transparency indices to the various surfaces that represent the variables.” The latter approach has proven limited for multivariate display. As Hibbard (1986) observed, “When viewing a transparent surface and an underlying opaque surface simultaneously, the eye is quite good at separating the shade variations of each surface and deducing their shapes. However, this breaks down when there are several transparent layers.” The former method has been exploited successfully in a revealing simulation of the cyclogenesis of a severe storm by Wilhelmson and his colleagues at the National Center for Supercomputing Applications at the University of Illinois (Wilhelmson and others 1990). Tracers, streamers, buoyant balls and translucent shading are artfully combined to reveal storm dynamics. Display of 3-D imagery can be costly, however: the seven and one-half minute simulation required the equivalent of one person’s full-time effort for more than a year and 200 hours rendering time on a Silicon Graphics Iris workstation.

**Realism vs. abstraction** Papathomas and colleagues (1988) remark that “[b]oth of the above methods result in images that are highly "unrealistic," illustrating that there may be instances in scientific computing in which the visualization technique may have to transcend "realism."” More recently, several authors have argued that increasing abstraction rather than realism may often be an effective visualization strategy (Muehrcke 1990, MacEachren and Ganter 1990). Scientists seem skeptical about this argument, however. We hypothesize that desire for realistic 3-D imagery in earth science visualization is often motivated by the aesthetic appeal of
ray-traced 3-D imagery and the intuitive metaphor of the terrain surface as a vehicle for expressing statistical surfaces. Robertson (1990) has elaborated this metaphor as a "natural scene paradigm." In our view, the uncritical acceptance of this paradigm will sometimes lead to both ineffective data representations and unnecessarily expensive visualization hardware and software.

**Multiple views** Exploratory analyses involve a variety of tasks. Carr and colleagues (1986) observe that "in general, we do not know how to produce plots that are optimal for any particular perceptual task." Since a data display may be called upon to serve several tasks, a search for a single optimal display method is likely to be futile. Some (for example, Monmonier 1991) have argued that single displays created to communicate data relationships to the public may even be unethical.

Windowing standards such as X-Windows now make it easy to produce multiple data displays for exploratory analysis. Two strategies are constant formats (for example, a sequence of maps of the same extent and scale showing changes in time series data) and complementary formats (such as a scatterplot linked to a map).

Tufte (1990) refers to ordered series of information graphics displayed in a constant format as "small multiples," and asserts that "for a wide range of problems in data presentation, small multiples are the best solution." Bertin (1981), who generalizes the variety of visualization tasks as two types of questions (What is at a given place? and Where is a given characteristic?), declares that "it is not possible, with complex information, to represent several characteristics in a comprehensive way on a single map while simultaneously providing a visual answer to our two types of question." Only comparison of several juxtaposed univariate maps can answer both types, he argues.

Monmonier has adapted several multiple view, complementary format display techniques for geographic analysis. In 1989 he described the concept of "geographic brushing" as an extension of scatterplot brushing (Becker and others 1988). Geographic brushing of trivariate distributions involves linked display of scatterplot matrices or 3-D scatterplots (whose axes represent thematic attributes) and a map. Extending the work of Gabriel (1971), Monmonier (1991) also describes the "geographic biplot" as "a two-dimensional graphic on which a set of place points represents enumeration areas or sample locations and a set of measurement points represents different indicators or different years for a single measure." The biplot's axes are usually derived from principal components analysis. Attributes are plotted by eigenvector coefficients, places by component scores. "Relative proximity on the biplot reflects relative similarity among places and variables." The biplot is most effective when two principal components account for a large proportion of the variation in a multivariate data set. When three axes are required to explain a satisfactory fraction of variation, 3-D biplots can be constructed (Gabriel and others 1986).

**Single superimposed views** Superimpositions (single displays on which symbols representing three or more variables are overlaid) are preferred for some exploratory analyses. Synoptic climatologists, for example, rely on multivariate
maps to characterize interactions of atmospheric phenomena that give rise to severe weather events. One example is a series of seven-variable superimpositions published by Lanicci and Warner (1991). Their maps employ a strategy for multivariate superimposition suggested by Bertin (1981), in which the discriminability of overlaid distributions is maximized by expressing different data variables with symbols of different dimensions (point, line and area). This strategy has also been employed by Crawfis and Allison (1991) for their "scientific visualization synthesizer.

**Dynamic graphics** A recurring theme in the literatures we have consulted is the advantage of dynamic graphical methods over their static counterparts. For most analysts, the term "dynamic graphics" connotes interaction: the "direct manipulation of graphical elements on a computer screen and virtually instantaneous change of elements" (Becker and others 1987). Both fast hardware and an interface that promotes efficient performance of intended tasks are required for fruitful interaction. Highly interactive systems allow users to easily select among lists of data variables to be displayed, different display formats and symbolization options. In object-oriented displays, individual elements or groups of elements may also be selected. Several operations on selected elements may be provided. For superimposed map-based displays, an analyst may want to be informed of the values of the multiple variables at a selected location. Analysts using linked, complementary format displays (as in brushing) need selections on one display to be automatically highlighted on the other.

Another important interaction task is rotation. User-controlled coordinate transformations that appear to rotate viewed objects are usually required to discover structure in multivariate distributions displayed as surfaces, point clouds and translucent volumes. Some systems allow users to specify real-time coordinate transformations by which analysts can guide a 2-D viewing plane through n-dimensional data spaces. Young and Rheingans (1990), for instance, demonstrated a real-time "guided tour" through a six-dimensional data space derived from principal components analysis of rates of seven types of crime for the United States. Cubes representing the states were "rocked" back and forth between two 3-D biplots formed by the first three and last three principal component coefficients, respectively. Clusters in the six-dimensional space were revealed by similar patterns of movement among the cubes as the biplots were rocked.

The abbreviated review which space permits us here cannot do justice to the variety and ingenuity of methods devised for the display of three or more geographically-referenced data variables. We hope that it provides at least a backdrop for the interface development project we describe next.

**CASE STUDY:**

**EXPLORATORY MULTIVARIATE ANALYSIS OF PALEOClimATE DATA**
The project we describe here is being carried out in the Deasy GeoGraphics Laboratory, a cartographic design studio affiliated with the Department of Geography that serves the Department, the College of Earth and Mineral Sciences and The Pennsylvania State University. One of the leading clients of the laboratory over the past five years has been Penn State's Earth System Science Center (ESSC), a consortium
of 23 geoscientists, meteorologists and geographers who study the global water cycle, biogeochemical cycles, earth's history and human impacts on the earth system. ESSC is one of 29 interdisciplinary teams participating in NASA's Earth Observing System (EOS) research initiative.

Recently, in response to emerging demand for animated and interactive multimedia presentations, Deasy GeoGraphics (like similar publication-oriented labs at other universities) has adopted a new emphasis on software development. When asked how the lab might apply its cartographic expertise in a development project that would benefit ESSC analysts, ESSC's Director replied that he had long wished to be able to simultaneously display multiple spatial "fields" on a single map base.

Typically, ESSC researchers rely on juxtaposed comparison of laser-printed univariate isoline maps to assess the spatial correspondence of model-produced atmospheric and oceanographic distributions. ESSC staff describe how the Director had not long ago performed a synoptic analysis of Cretaceous precipitation, upwelling and winter storms for an award-winning paper (Barron and others 1988) by overlaying paper maps on his office window.

Overlay analysis is a generic GIS problem. More specifically, however, the problem we have addressed is how to effectively display multiple spatial distributions for ESSC researchers' exploratory analyses. This problem is not susceptible to generic solutions.

Paleoclimate data Our interface displays data produced by numerical models such as the National Center for Atmospheric Research Community Climate Model (CCM). The CCM is a three-dimensional climate model whose resolution is 7.5° in longitude by approximately 4.5° latitude, yielding a regular grid of 1,920 predictions. At ESSC, the Genesis version of the CCM runs on a Cray Y-MP2 supercomputer. A typical application of the model is to simulate three to five climatic phenomena for nine to twelve atmospheric levels for the Cretaceous (100 million years before present). The product of the simulation is a digital "history volume" which consumes about two gigabytes of disk space or tape in binary form. Analysts subsequently use a separate software module called the CCM Processor to produce "history save tapes" in which the model-produced data may be aggregated to monthly, seasonal or yearly averages. One kind of history save tape is a "horizontal slice tape," an IEEE-format binary file containing aggregate univariate data predicted for a plane that intersects the atmosphere parallel to Earth's surface. The data are quantitative, interval/ratio scaled, gridded, 2-D representations of continuous atmospheric phenomena.

We have christened our map interface "SLCViewer" to denote its role in the ESSC software toolbox as an interactive, mouse-driven utility for viewing multiple horizontal slice tapes. We developed SLCViewer using the Interactive Data Language (IDL) on a Sun Sparc2 workstation under Unix. The project is currently in the "proof of concept" stage. Procedures developed in IDL may be implemented later in other visualization or GIS applications if warranted.

The SLCViewer interface(s) SLCViewer actually offers analysts a choice of two interfaces: a single map or four juxtaposed maps. After launching IDL, then entering "slcview" to compile a main call program, the SLCViewer base widget
The base widget provides two buttons named "one" and "four" by which users select an interface. The base widget also provides pop-up menus for selecting data variables, geologic time periods, and for specifying isoline levels and colors. Current selections are displayed textually. An "exclusion" button calls up another widget for focusing on subsets of selected data. We'll discuss focusing a little later.

The analyst moves his or her mouse pointer from the base widget to the map window to symbolize the selected data. SLCViewer can display data with point, line and area symbols on a cylindrical equidistant map projection (which despite its shortcomings is used commonly in the global change community). Figure 1 shows three variables superimposed in the single map window. Symbol designation controls are identical for both the one map and four map windows.

**Focusing on subsets of the data** Our work has been influenced by two ESSC scientists who have met with us on several occasions to observe our progress and offer suggestions. Early on they outlined the sequence of exploratory tasks they expected to perform with the interface. The sequence involves initial inspection of univariate maps of complete distributions, followed by focused observation of selected subsets of distributions (a process they referred to as "exclusion"), then superimposition of several complete or subsetted variables. SLCViewer supports the focusing process by providing an exclusions widget by which users can restrict displays to values greater than or less than a specified threshold value. Excluded

![Figure 1](image.png)

**Figure 1** Single-map window of the SLCViewer interface in which three paleoclimate variables are superimposed (38 percent of original size). Average June, July and August temperature data for the Cretaceous (100 million years before present) have been mapped onto a range of 100 gray levels and are displayed as a matrix of gray areas that correspond to the spatial resolution of the model. Lighter areas represent warmer temperatures. Average precipitation data for the same period have been interpolated and are depicted by weighted isolines that increase in width for higher values. Average evaporation is represented by four categories of point symbols classified by nested means. Light and dark points (red and blue in the original) represent data above and below the mean. Large and small points designate the outer and inner two categories.
values in data sets symbolized with area (gray scale) symbols are represented with
dark blue grid cells. Isolines can be suppressed via the exclusions widget or simply
by specifying the desired isoline levels in the lines widget. The exclusion widget
does not withhold data from interpolation but merely suppresses display of speci-
fied isolines. Exclusions do not cause any point symbols to disappear, but a speci-
fied value (such as zero) replaces the mean of the data set in the classification of
symbol sizes and colors.

**Two-dimensional displays** SLCViewer displays multivariate paleoclimate
data in 2-D, planimetrically-correct map formats. Although the CCM is a 3-D model
of the atmosphere, the horizontal slice tapes produced by the CCM Processor are 2-
D matrices representing predicted values for discrete horizontal planes in the atmo-
sphere. Following Tufte's (1983) maxim that the dimensionality of a graphic repre-
sentation should not exceed the dimensionality of the data, 2-D data ought to be
represented by 2-D images. His assertion can be extended for display of multiple
horizontal slice tapes. Since each horizontal slice is parallel to Earth's surface, all
horizontal slices are parallel to one another. When the slices chosen for combined
display represent phenomena at the same level of the atmosphere (as they usually
do), they are coplanar. When these conditions are met, superimposed 2-D map
displays are faithful representations of multivariate paleoclimate data.

Given the good sense of Tufte's principle, and partial support for it in em-
pirical research (see Barfield and Robless 1989 and Kraak 1988), we have chosen to
portray multiple 2-D data fields on 2-D, planimetric map bases. A secondary practi-
cal benefit of the 2-D approach is that substantially less computing power is re-
quired for interactive performance, meaning that analysts need not have access to
very high performance workstations to use SLCViewer effectively.

**Multiple views** SLCViewer provides optional display of a single map on
which up to three variables are superimposed or four maps among which up to
four variables are juxtaposed. In an interactive exploratory environment the choice
between single and multiple displays need not be exclusive. SLCViewer allows
analysts to have it both ways. The single map interface facilitates superimposition
of three variables (or more, if the analyst is willing to deal with multiple isoline or
point displays). The four-map interface provides small multiples. Although we are
keen on the notion of multiple complementary formats, we have restricted our
current efforts to developing map-based strategies.

**Interactivity** Finally, SLCViewer is a dynamic environment in which ana-
lysts may select data variables to be displayed and express each variable with a
distinctive set of point, line or area symbols. The ability to quickly revise data
variable combinations, to focus on subsets of the data and to try different symbol-
ization strategies could increase the efficiency of exploratory analyses of model data
at ESSC. By easing the exploratory process we hope that SLCViewer may even
foster new insights. The ESSC scientists who have consulted with us in SLCViewer's
development expect that the interface will be useful also as a teaching tool in
geoscience classes.

Symbol dimensions and graphic variables The trivariate map strategy we
adopted for the single map interface involves contrasting the dimensionality of
symbols (point, line and area) and enhancing each symbol type with the graphic variables that are most potent for expressing quantitative data (size for lines and points, lightness for area symbols). This approach contrasts with the iconographic approach used for Exvis, another 2-D method in which all variables are represented as compound line symbols that vary in size, orientation and color. Though Exvis displays excel in revealing overall patterns among multiple variables, interpretation of patterns in the distribution of individual variables seems difficult. We have not yet begun to formally evaluate the effectiveness of our strategy, however.

EVALUATING EFFECTIVENESS
Two crucial tests of the effectiveness of an exploratory technique are whether it is used at all, and if so, whether its use profits the analyst. If comparable methods exist, these can be compared in controlled studies. Such studies are valuable since visualization is expensive both in terms of time and capital outlays. Benefits ought to be documented. The initial development of SLCViewer continues even as this article takes shape—the interface has not yet been used by ESSC scientists for actual research. Our plan is to promote SLCViewer within ESSC and, assuming it is adopted, closely monitor the results.

CONCLUSION
With long experience in the visual display of quantitative information, cartographers have much to contribute to the practice of scientific visualization. We only occasionally encountered references to cartographic research in the literatures we consulted, however. In an earlier publication, some of us called on cartographers to engage their experience and creativity in applied visualization projects and to observe carefully what scientists find useful (DiBiase and others 1992). In this project we have aimed to practice what we preached. Our efforts have been guided by the needs and expertise of scientists among whom visualization is a routine but inefficient activity. Our contribution is a prototype exploratory multivariate data analysis tool for geographically-referenced data that maintains the planimetric integrity and computational efficiency of 2-D maps, offers analysts a choice of single or multiple views, and provides user control over data variable and graphic symbol selections. We cannot yet gauge the value of this contribution. We’ll see.

ACKNOWLEDGMENTS
Thanks to Bill Peterson, Jim Leous and Jeff Beuchler of the ESSC Computing Facility who provided us the workstation, advice on IDL and access to the slice tape data. We appreciate also the interest of ESSC researchers Karen Bice and Peter Pavcett, and Eric Barron’s unflagging support.

REFERENCES


