# EXPLORING ENVIRONMENTAL DATA IN A HIGHLY IMMERSIVE VIRTUAL REALITY ENVIRONMENT

DIANNE COOK[1], CAROLINA CRUZ-NEIRA[2], BRADLEY D. KOHLMEYER[2], ULI LECHNER[3], NICHOLAS LEWIN[4], LAURA NELSON[2], ANTHONY OLSEN[5], SUE PIERSON[6], and JÜRGEN SYMANZIK[1]

*[1]Department of Statistics, Iowa State University, Ames, IA 50011, [2]Iowa Center for Emerging Manufacturing Technology, Iowa State University, Ames, IA 50011, [3]German National Research Center for Information Technology, Institute for Media Communications, Schloß-Birlinghoven, 53574 Sankt Augustin, Germany, [4]Geographic Information Systems Support and Research Facility, Iowa State University, Ames, IA 50011, [5]US EPA National Health and Environmental Effects Research Laboratory, Western Ecology Division, Corvallis, OR 97333, [6]OAO, c/o USEPA NHEERL Western Ecology Division, Corvallis, OR 97333.*

**Abstract.** Geography inherently fills a 3D space and yet we struggle with displaying geography using, primarily, 2D display devices. Virtual environments offer a more realistically-dimensioned display space and this is being realized in the expanding area of research on 3D Geographic Information Systems (GISs). Traditionally, a GIS has only limited tools for statistical analysis, and 3D GIS research has concentrated on the visualization of the geographical terrain. Here we discuss linking multivariate statistical graphics to geography in the highly immersive C2 virtual reality environment at Iowa State University using mid-Atlantic streams data.

## 1. Introduction

In this paper we discuss exploring multivariate spatial data in a highly immersive virtual reality environment. There is considerable excitement about rendering geography in 3D, but there (as yet) are few efforts to build in analytical tools. Entwined to make good data analysis are visualization tools and analytical tools. This work explores developing these tools in the highly immersive C2 virtual reality environment.

We have data on 501 sampling sites on streams in the mid-Atlantic states of Delaware, Maryland, New York, Pennsylvania, Virginia, and West Virginia. It is a probability sample of wadeable streams in the period 1993-1995 by the U.S. E.P.A.'s Environmental Monitoring and Assessment Program (Klemm and Lazorchak, 1995). At each site the subset of measurements we use are Closed System pH, Calcium (μeq/L), Sodium (ueq/L), Ammonium (μeq/L), Chloride (μeq/L), Nitrate (μeq/L), Sulfate (μeq/L), Dissolved Organic Carbon (mg/L), Total Suspended Solids (mg/L), and Total Phosphorous (μg/L). In addition to these measurements we have the latitude and longitude, elevation, and the Aggregated Omernik Level 4 ecoregion membership of each sampling site.

In order, the sections of the paper discuss geographic information systems, dynamic statistical graphics for multivariate data, immersive virtual reality technology, linking the geography with the multivariate graphics in the virtual environment and applied to the streams data.

*Environmental Monitoring and Assessment* **51:** 441–450, 1998.

## 2. Geographic Information Systems

Geographic Information Systems (GISs) have played and do play an enormous role in the analysis of environmental data. They perform the task of storing and displaying spatial data and concomitant geographic variables. Critical to any good GIS are database storage and retrieval and solid map drawing capabilities.

The existing software for GIS concentrates on 2D display. Figure 1 shows the mid-Atlantic streams data (described in the Introduction) displayed in ArcView 3.0™[1]. There is growing interest in constructing and designing appropriate data structures for 3D GIS, which is quite natural given that geography is inherently 3D. GRASS is an example of software that is developing tools for 3D GIS (Brown *et al.*, 1995; Mitasova *et al.*, 1995). Also, see http://www.esri.com/base/products/arcinfo/3dtin/tin.html for ESRI developments. Raper (1989) provides more introduction into this field.

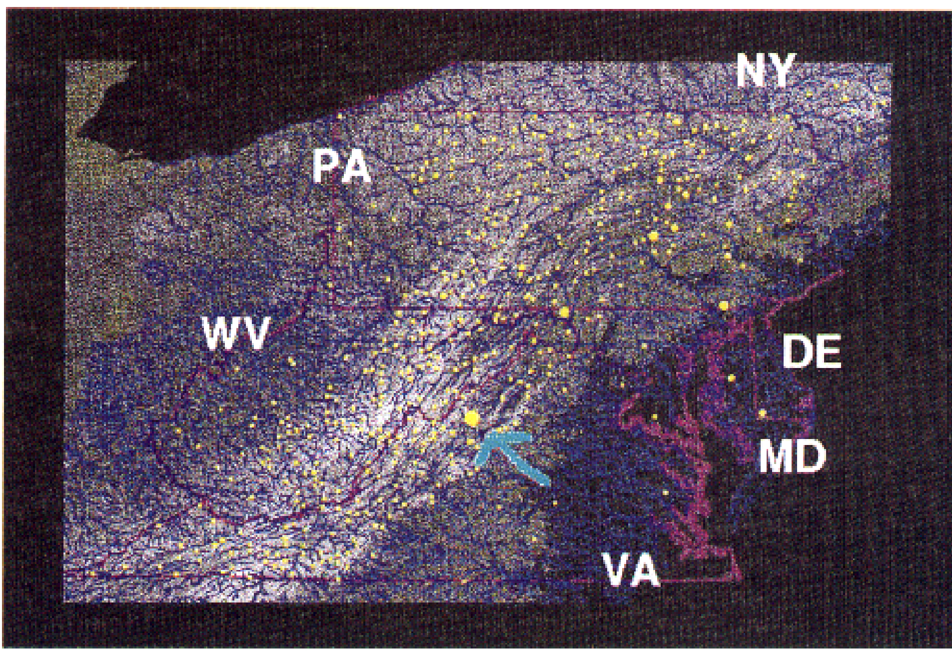


*Fig. 1.* ArcView 3.0™ view of the elevation and sampling sites in the mid-Atlantic states with Nitrate value coded proportional to circle diameter.

In addition, traditionally, GISs lack tools for visualizing multivariate spatial data. In Figure 1, the diameter of the spot at each sampling site is a function of the nitrate concentration. One site has a very high nitrate concentration (highlighted in Figure 1). Does this site also have high ammonium concentration? To answer this, we could code ammonium similarly. But keeping track of the relationships between these variables, and more, becomes increasingly difficult. A nicer approach to exploring the multivariate nature of the measurements is to use dynamic statistical graphics linked to the geographic location. An example of this approach can be found in Cook *et al.* (1996) and Cook *et al.* (1997) (shown in Figures 2,3).

In short, a GIS provides the geographic context but lacks the tools for multivariate statistical visualization and analysis.

## 3. Dynamic Statistical Graphics

Dynamic statistical graphics enables data analysts in all fields to carry out visual investigations leading to insights into relationships in complex data. Graphics for displaying one or two variables simultaneously are well-known and familiar but plotting three, four, or five variables simultaneously and in a manner that allows easy insight into the data is a tricky proposition. This has been the domain and challenge of statistical graphics research for over 30 years. Through an evolution of ideas, we have scatterplot matrices, parallel coordinate plots, icons (for example, Chernoff faces), and animated plots for viewing many variables. (Videos demonstrating these methods can be borrowed from the American Statistical Association Statistical Graphics Section Video Lending Library. See http://orion.oac.uci.edu/~rnewcomb/statistics/graphics/graphics.html.)

Using the animation approach to build directly from the familiar 2D scatterplot, we can animate plots by choosing different linear combinations of several variables. The most familiar technique is to rotate three variables by looking at a continuous sequence of 2D linear combinations of the three variables. This method extends naturally to linear combinations of four, five, or more variables, and the extension is called a grand tour (Asimov, 1985; Buja *et al.*, 1997). With a grand tour we can see multivariate structure such as clusters, multicollinearity, or outliers, that may not be visible in plots of one or two variables.

Using multivariate graphical methods we might take two windows to view the data and in one look at the measured variables in a grand tour and in the other plot the geographic location of the sampling sites. As a simple example, Figure 2 shows two views of the mid-Atlantic streams data. The software used is XGobi (Swayne *et al.*, 1991) and the geographical locations are shown in ArcView 3.0™. In the XGobi window (top), several observations that are extreme in comparison to the majority of points are identified. These

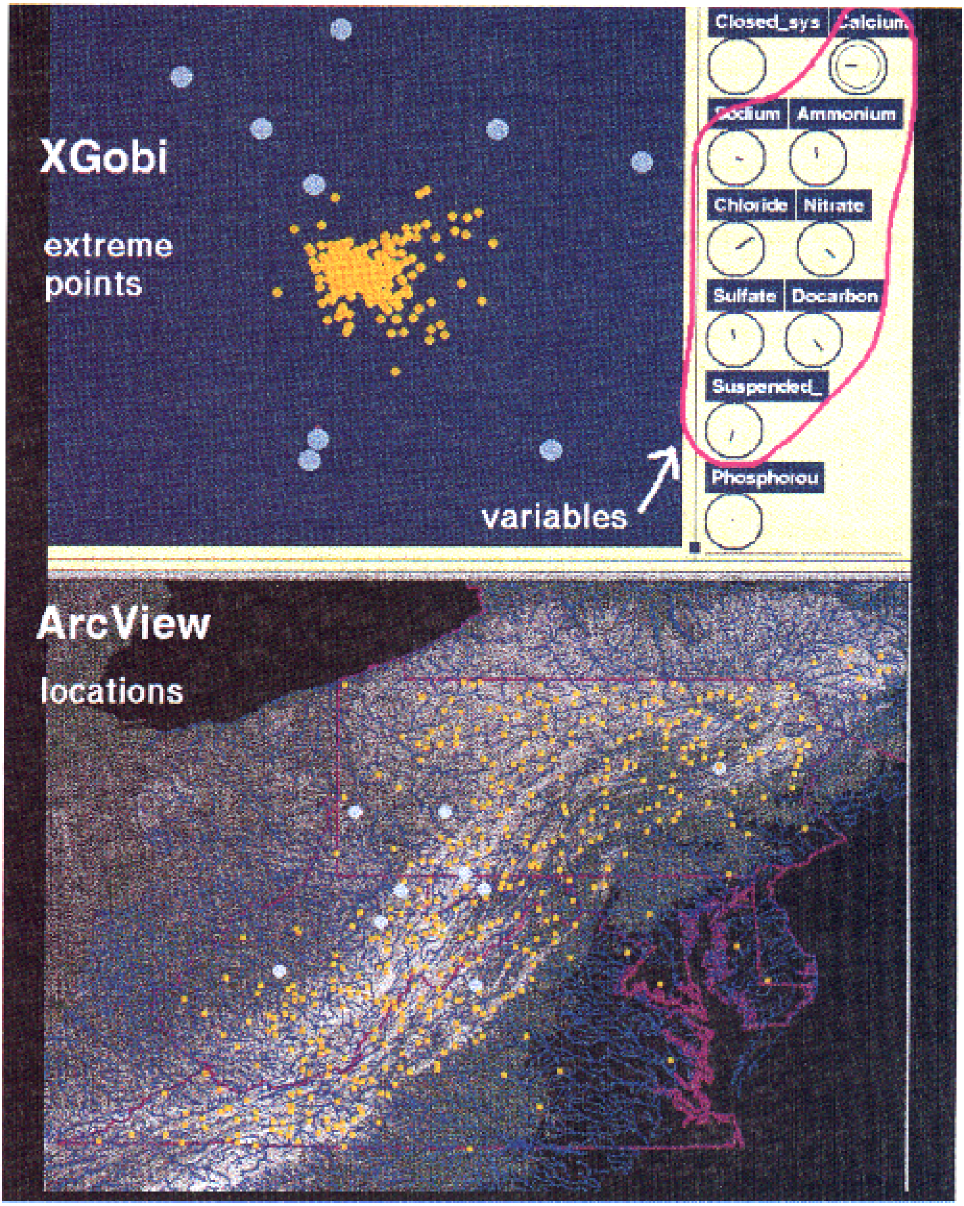


*Fig. 2.* Using XGobi linked to ArcView 3.0™ to explore the streams data. Several observations that are extreme in comparison to the majority of points are identified (top). These have outlying values on one or more of several variables: Calcium, Sodium, Ammonium, Chloride, Nitrate, Sulfate, Dissolved Organic Carbon, Total Suspended Solids. These extreme values are spread throughout the geographic region (bottom).

have outlying values on one or more of several variables: Calcium, Sodium, Ammonium, Chloride, Nitrate, Sulfate, Dissolved Organic Carbon, Total Suspended Solids. These extreme values are spread throughout the geographic region as the ArcView window (bottom) shows. In another view, a dotplot of Closed System pH has the more acidic samples brushed (Figure 3, left) and the locations appear mostly at higher altitudes (Figure 3, right).

Connecting multivariate graphics with a geographic information system allows for a more comprehensive analysis because available concomitant information, such as elevation, ecoregion, or census information, can be overlaid onto the map.

## 4. Immersive Virtual Reality Technology

Virtual reality (VR) is a rapidly developing technology that involves many aspects of computer-augmented visualization. The technology dates back to 1965 when Sutherland (1965) proposed the Ultimate Display and built the first head-mounted display, the Sword of Damocles, with cathode ray tubes and a ceiling suspension system in 1968. A force-feedback system was developed in 1971 by Frederick Brooks, and the Data Glove which measures finger angle was developed in 1985 by Thomas Zimmerman. A brief chronology of events that influenced the development of VR can be found in Cruz-Neira (1993), and a more complete overview can be found in Pimentel and Teixeira (1995).

The C2 at Iowa State University is a highly immersive VR environment in which images are projected onto the walls and floor of a small "room" to create the illusion of 3D when CrystalEyes Stereographics' LCD shutter glasses are worn. It uses position tracking and auditory feedback through multiple speakers to immerse users in a 3D environment. The position and orientation of the user's hands and head are determined through the use of a magnetic based tracker, a cyberglove, and a hand-held wand. It is possible for multiple viewers to enter the C2 and view the same scene with minimal equipment. Technical details and applications of a VR environment similar to the C2 can be found in Cruz-Neira (1995) and Roy *et al.* (1995).

## 5. Linking Multivariate Graphics and Geography in the C2

In the C2 environment, we have two viewing areas: a viewing box which contains the scatterplot view of the variables allowing us to conduct a grand tour and an elevation surface for the region of the study on which locations of the sampling sites are shown. In addition, we have several controls/interaction tools: a speed pole for changing the speed of the tour, a color palette for selecting a brush color, a symbol type, and a way to resize/reshape the brush from a sphere to a rectangle (Figure 4). There is sound feedback on interaction activities which makes the interaction more efficient. More details on dynamic statistical graphics in the C2 (at earlier stages) can be found in Symanzik *et al.*

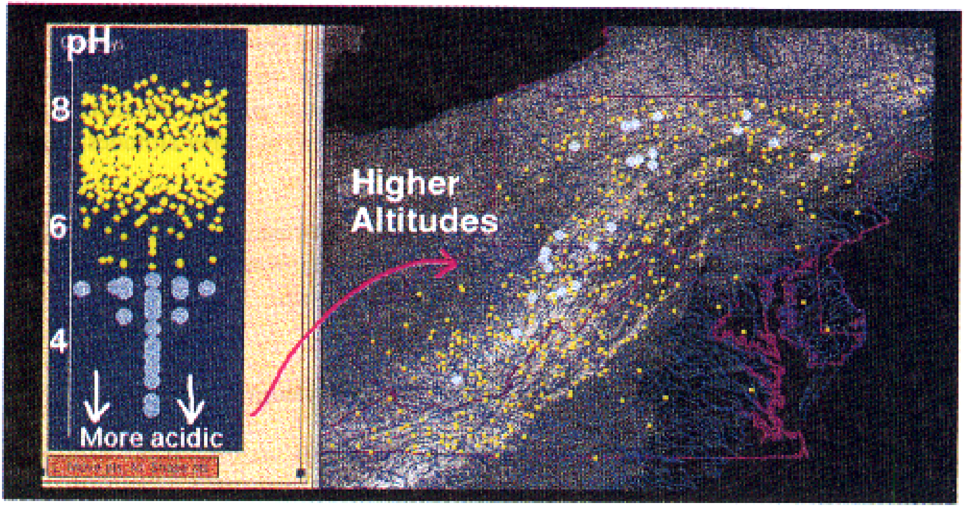


*Fig. 3.* Using XGobi linked to ArcView 3.0™ to explore the streams data. A dotplot of Closed System pH has the more acidic samples brushed (left) and the locations appear all to be at the higher altitudes (right). The dotplot is read like a vertical histogram. Closed System pH can be seen to be skewed with most measurements at the top around the slightly alkaline values 7-8, and less in the acidic range at the bottom of the dotplot. Values less than 5 have been brushed light blue.

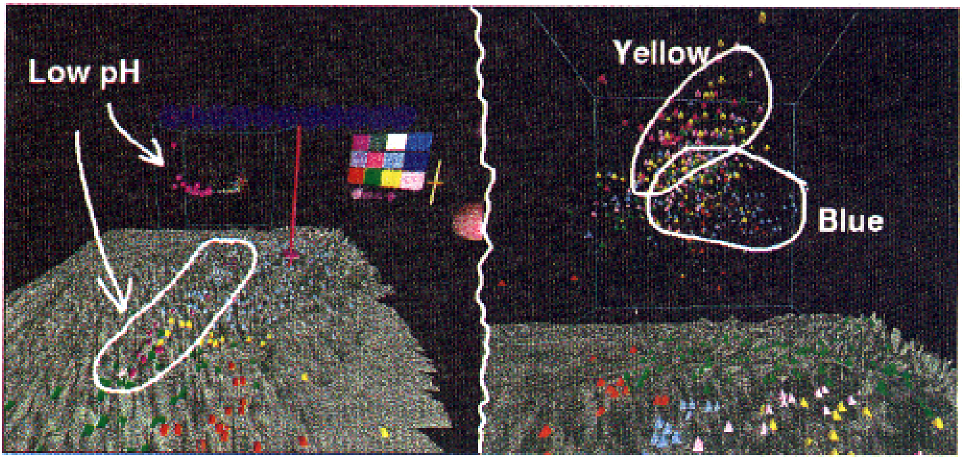


*Fig. 4.* In simulator mode: (left) Low pH values are brushed purple in the scatterplot, and these sampling sites which are more acidic fall higher up in the mountains. This corroborates what was seen in Figure 3 but in the C2 it is much easier to see the sites are at high altitude because the mountains are rendered in 3D. (right) Grand tour view over the elevation indicating differences in combination of chemical contaminants over ecoregion.

(1997). (Other examples of work on building a general multivariate data visualization system in a virtual reality environment are discussed in van Teylingen *et al.* (1997) and Carr *et al.* (1996).

Sampling sites are linked one-to-one between the scatterplot and the elevation map. So painting a point (or group of points) in the scatterplot highlights it (them) on the elevation map. In this way, locations of interesting features can be discovered and assessed in relation to other interesting features or elevation. And most importantly, the multivariate nature of the data can be explored more extensively than is possible with a univariate variable-by-variable analysis.

## 6. Mid-Atlantic Streams Application

We have displayed the chemical information in the viewing box of the C2, and pulled out geographic location of the sampling site and drawn these with regard to the elevation of the site. In addition we have colored the observations according to their ecoregion identity. The elevation map is drawn on the floor that we walk on (Figure 5). The effect is stunning: acting like Gulliver we can walk on the mountaintops and look along valleys.

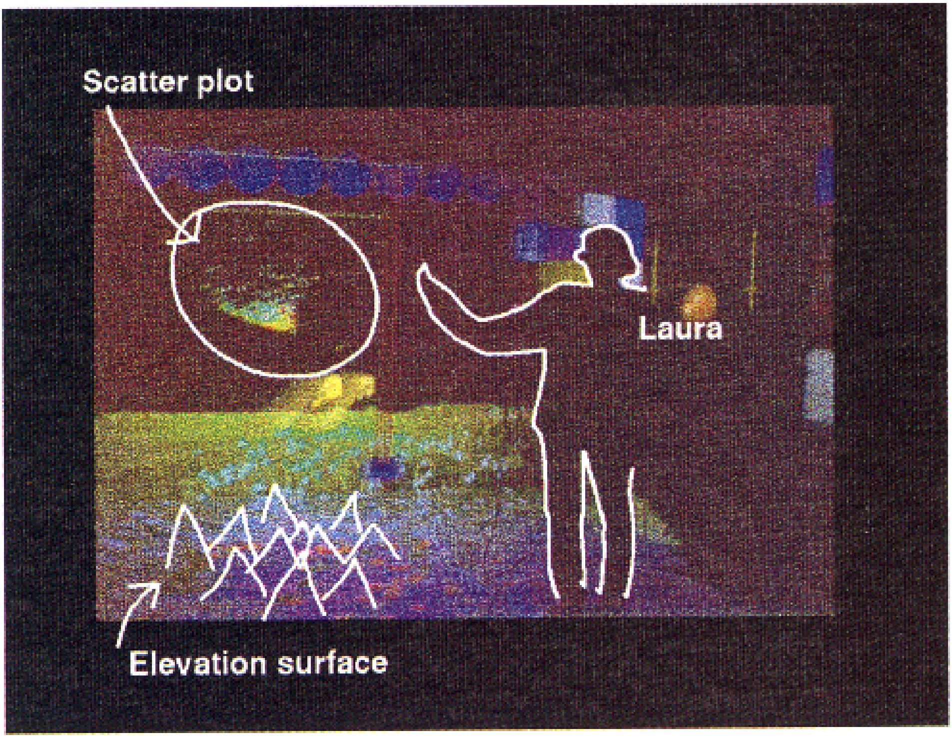


*Fig. 5.* The measurements on samples in streams of the mid-Atlantic states in the C2.

Some initial observations of the data show that there are some isolated sampling sites that have extreme values on several chemistry-related variables. In the viewing box, these points are seen as outliers from the main clustering of points. These individual points are brushed to see their locations in the study region. You might expect that these sampling sites are located in a similar geographic region, but they appear to be spread throughout the entire sampling region.

Low pH values are concentrated at higher elevations (Figures 3,4 ) and the different level 4 ecoregions have different combinations of the measured variables (Figure 4). The relationship between pH and elevation has been noticed in other studies and is understood: acid deposition is neutralized as water flows downstream. The few sites which do not correspond to high elevation occur in the western part of Pennsylvania and are most likely associated with mine drainage.

The data are heavily skewed towards small quantities of contaminants. so we looked at both the raw data and power transformations of the variables. (Transformations of skewed data are commonly used in data analysis to make the data more symmetric. Once transformed it is easier to see other informative patterns in the data which may have been hidden by the skewness.) Amongst the transformed variables strong linear dependencies exist which are not immediately obvious from the pairwise plots.

## 7. Conclusions

Motivating this work is the question of "How this technology might enhance environmental monitoring and assessment?" The virtual environment does offer a lot of scope for visualizing environmental data. It is possible to represent geography to its full 3D extent. A virtual space allows for enormous flexibility in zooming or panning into global or local views. It has potential for communicating information to an uninformed audience by placing it in the context of a familiar physical environment, appropriate to the problem being addressed. But virtual reality is still far from being a desktop tool and the highly immersive tools that we have discussed are available at only a few (but expanding) locations worldwide.

The C2 is a highly immersive and very expensive environment. The benefit is that it provides a very realistic sense of being surrounded by the scene. It is a developing technology, especially at Iowa State University, and so it is constantly being enhanced by new software libraries and undergoing hardware updates and additions. Modeling to make a truly believable physical environment would require enormous efforts and probably considerably more computational power than even this state-of-the-art setup has, to track the user's movements and update the scene sufficiently quickly. The C2 uses the OpenGL graphics library. So to draw the elevation surface we had to resort to the first principles of computer graphics: start with a regular and ordered grid, and calculate the normals to the individual polygons to determine the appropriate light and shade.

It would be appropriate to collaborate with a 3D GIS research group who could model the terrain of the study region quickly and in a more sophisticated manner. Incorporating multivariate visualization tools with terrain visualization shifts 3D GIS "up a gear" into the domain of data analysis.

## Acknowledgements

Helpful comments and information about spatial data analysis methods were provided by Professor Noel Cressie.

The research reported in this article has been funded by National Science Foundation Grant DMS–96–32662 and the U.S. Environmental Protection Agency through Cooperative Agreement CR822919-01-0 with Iowa State University. This paper has not been subjected to the Agency's peer and administrative review. No endorsement of the contents by the Agency should be inferred.

## Additional Information

Further information and developments to the work can be found at:
http://www.public.iastate.edu/~dicook/research/C2/statistic.html.

## References

Asimov, D.. 1985, "SIAM" *J. of Sci. and Stat. Comp.*, **6** (1):128–143.

Brown, W. M., Astley, M., Baker, T., and Mitasova, H.: 1995, "Twelfth International Symposium on Computer-Assisted Cartography", pages 89--99, Charlotte, NC.

Buja, A., Cook, D., Asimov, D., and Hurley, C.: 1997, J. of Computational and Graphical Statistics. Submitted.

Carr, D. B., Wegman, E. J., and Luo, Q.: 1996, "Technical Report 129", Center for Computational Statistics, George Mason University.

Cook, D., Majure, J. J., Symanzik, J., and Cressie, N.: 1996, "Computational Statistics: Special Issue on Computer Aided Analyses of Spatial Data", **11** (4):467–480.

Cook, D., Symanzik, J., Majure, J. J., and Cressie, N.: 1997, Computers and Geosciences: Special Issue on Exploratory Cartographic Visualization. 4 (1); 371-385, web material at www.elsevier.nl/locate/cgvis.

Cruz-Neira, C.: 1993, SIGGRAPH '93 Course Notes 23, 18 pages.

Cruz-Neira, C.. 1995, PhD thesis, University of Illinois at Chicago.

Klemm, D. J. and Lazorchak, J. M. editors.: 1995, "Technical Report EPA/620/R-94/004", U.S. Environ. Protection Agency, Office of Res. and Dev., Environ. Monitoring Systems Laboratory, Cinncinnati, Ohio.

Mitasova, H., Mitas, L., Brown, W. M., Gerdes, D. P., Kosinovsky, I., and Baker, T.: 1995, *Intl. J. of Geographical Info. Systems*, **9** (4):433–446.

Pimentel, K. and Teixeira, K.: 1995, "Virtual Reality through the new Looking Glass (Second Edition)", McGraw-Hill, New York, NY

Raper, J.: 1989, "Three Dimensional Applications in Geographic Information Systems", Taylor Francis, London, UK.

Roy, T., Cruz-Neira, C., and DeFanti, T. A.: 1995, "Presence: Teleoperators and Virtual Environments", **4** (2):121–129.

Sutherland, I. E.: 1965, Proc. IFIP 65, 2, pages 506--508, 582–583.

Swayne, D. F., Cook, D., and Buja, A.: 1991, "ASA Proceedings of the Section on Statistical Graphics", pages 1–8, Alexandria, VA. American Statistical Association.

Symanzik, J., Cook, D., Kohlmeyer, B. D., Lechner, U., and Cruz-Neira, C.: 1997, IASC Proceedings, Forthcoming.

Van Teylingen, R., Ribarsky, W., and Van Der Mast, C.: 1997, *IEEE* "Transactions on Visualization and Computer Graphics", **3** (1):65–74.