CONF-880385--2

DE89 012339

New Dimensions from Statistical Graphics for GIS Analysis and Interpretation

Raymond A. McCord and Richard J. Olson

Environmental Sciences Division
Oak Ridge National Laboratory[1]
Oak Ridge, Tennessee

ABSTRACT

Environmental research and assessment activities at Oak Ridge National Laboratory (ORNL) include the analysis of spatial and temporal patterns of ecosystem response at a landscape scale. Analysis through use of a geographic information system (GIS) involves an interaction between the user and thematic data sets frequently expressed as maps. A portion of GIS analysis has a mathematical or statistical aspect, especially for the analysis of temporal patterns. ARC/INFO is an excellent tool for manipulating GIS data and producing the appropriate map graphics. INFO also has some limited ability to produce statistical tabulation. At ORNL we have extended our capabilities by graphically interfacing ARC/INFO and SAS/GRAPH to provide a combined mapping and statistical graphics environment. With the data management, statistical, and graphics capabilities of SAS added to ARC/INFO, we have expanded the analytical and graphical dimensions of the GIS environment. Pie or bar charts, frequency curves, hydrographs, or scatter plots as produced by SAS can be added to maps from attribute data associated with ARC/INFO coverages. Numerous, small, simplified graphs can also become a source of complex map "symbols." These additions extend the dimensions of GIS graphics to include time, details of the thematic composition, distribution, and interrelationships.

Background

The Environmental Sciences Division (ESD) of Oak Ridge National Laboratory (ORNL) is one of the U. S. Department of Energy's (DOE's) leading facilities for environmental and ecological research. Over the past 30 years ESD has conducted numerous projects including environmental impact assessment, environmental and ecosystem modelling and simulation, ecosystem analysis for nutrient cycling and pollutant fate and effects, radiological and hazardous waste management, regional environmental monitoring, and data analysis and interpretation. The studies have included landscape scales ranging from a single watershed to global. The ecological and environmental projects have evaluated the effects and interrelationships of individual and multiple components.

[1]Operated by Martin Marietta Energy Systems, Inc. under contract number DE-AC05-84OR21400 with The U. S. Department of Energy.

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

In all of these categories of studies, ESD has made extensive use of computing technology for its data management, analysis, and presentation. The analytical tools frequently used with data in ESD studies include statistics, graphics, data integration, model construction and simulation, and geographic information systems (GISs). A typical research problem will require the use of most of these analytical techniques. The chronology of a project also frequently includes several interchanges between data management, statistical analyses, and mapping with a GIS. Because of these operational considerations, ESD scientists have emphasized the need to integrate the data, methods, and products from these analytical tools. Olson et al. (1988) discussed the needs and methods for integrating data from several sources for national GIS studies. This paper presents the results of an effort to integrate the graphics from the statistical software (SAS) and the GIS software (ARC/INFO) used by staff in ESD.[2]

The integration of SAS/GRAPH and ARC/INFO graphics was initially done to combine two complementary classes of graphical products. SAS and SAS/GRAPH are very comprehensive software systems that ESD has used for several years to conduct data management, statistical analysis, and data-intensive graphics (i.e., graphics which depend upon the processing and display of hunderds to thousands of data values). Because of SAS's ease of use, comprehensiveness, and power, these software systems are used by a majority of the ESD scientists performing data analyses. However, the mapping and GIS capabilities of SAS/GRAPH are limited. In contrast, ARC/INFO is a very comprehensive GIS and mapping software system with limited statistical analysis capability and essentially no data-intensive graphics capability for the display of statistical analyses. (The primitive graphics tools for drawing various graphs are found in ARCPLOT, but simple graphs can require hundreds of statements.) As result of these characteristics, we anticipated that these two systems would be very good complements to one another.

After the initial interface between these graphics systems was developed, we determined that the combination of map- and data-based graphs also adds new dimensions to data interpretation and analysis. While the combined graphic product could be simply adding a couple of graphs (e.g., frequency histograms and pie charts) to a map, this capability could also be used as a very sophisticated "symbol generator." If the graphs are simplified (i.e., contain only very simple or no axis labelling; bold, but unpatterned, symbols; and uniform size, scale, and axes ranges), then they can be made relatively small

---

[2]SAS and SAS/GRAPH are registered trademarks of SAS Institute, Inc., Cary, North Carolina. ARC/INFO is a registered trademark of Environmental Systems Research Institute, Inc., Redlands, California. The use of trademarks in this paper does not constitute an endorsement of these products. Full identification of the trademarks can be found in the references.

($1\ cm^2$). If the graphs are made small enough, then numerous (tens to hundreds) graphs can be plotted on a map as if they were "symbols."

These "symbols" can then readily add large amounts of information and new dimensions to the data displayed. A map can display the two or three spatial dimensions of the thematic data being analyzed. The symbols from points, lines, and polygon shading can display additional categorical changes and spatial overlap of the attributes of the data in a GIS system. In addition to these features, the graphs on the map could display the following types of new details about differences between subcomponents of the thematic information:

- thematic values versus time,
- frequency distribution for sample populations,
- multivariate relationships,
- sample compositions,
- enumerated data plots or lists, and
- statistical summary tables.

Approach

The interface between SAS/GRAPH and ARC/INFO was developed in the following hardware and software environment:

SAS/GRAPH	Hardware:	IBM PC/AT, EGA monitor;
	Software:	PC-DOS 3.2, SAS, SAS/GRAPH 6.03.
ARC/INFO	Hardware:	MicroVAX 3500, Imagen 2308 laser printer, Matrix PCR film recorder;
	Software:	VAX/VMS 4.7a, ARC/INFO 4.01, VAX FORTRAN 4.8.

While the above specifications describe the systems used at ESD, the interface can be implemented between any hardware system(s) that can execute the current versions of SAS, SAS/GRAPH, and ARC/INFO and transfer character data.

The first step in developing the interface was to modify the specifications of the metagraphics driver provided by SAS with PROC GDEVICE (SAS 1988) to simplify the hardware capabilities assumed by the driver. The metagraphics driver can be altered to control the assumed device characteristics such as page size, resolution, number and name of colors, hardware character and symbol generation, polygon fill, hardware arc/curve generation, and hardware line types. To make the interface as universal as possible, hardware capabilities for character, symbol, line type, and curve generation were not allowed in the metagraphics device driver. The resulting output file from SAS/GRAPH then contained only instructions for pen selection, movement, and drawing. The metagraphics output file was defined to be in ASCII so that it could be transported between systems (PC to VAX) without data and record conversion problems.

After the metagraphics file from SAS is transferred to the ARC/INFO host (a VAX in this case), it is processed by a FORTRAN program that produces an ARC/INFO plot file. The program that converts the metagraphics file to ARC/INFO plot file format includes the following steps:

- unneeded SAS/GRAPH plot header information is bypassed,
- SAS page units are converted to inches,
- X and Y minimum and maximum are calculated, and
- SAS pen control commands are converted to ARC/INFO plot file syntax.

Incorporation of the graphs in ARC/INFO plot file format into a final map can be accomplished with the PLOT command in ARCPLOT. The graphs can be incorporated into a final map with or without the map composition capability of ARCPLOT. Map composition commands can be used to reposition and rescale the graphs as needed.

When generating graphic "symbols" in SAS to be displayed by ARC/INFO, each "symbol" is transferred as a separate plot file. Often the plots are generated by using a SAS BY statement with the site or plot ID value. Unique plot names which incorporate the ID can then be used to associate the symbol with map coordinates to control plotting within the ARCPLOT program. The following schematic AML program illustrates cross referencing of plot filename for graphs with the coordinates by the ID from an attribute file and the calculation of the BOX coordinates for the PLOT command in ARCPLOT (ESRI 1987):

```
ARCPLOT
      ... and other statements setting up map and drawing other map elements...
&S OUNIT [OPEN TEST.DAT STAT -R]
      ... TEST.DAT is a file containing ID and coordinates from PAT for plotting
      graphs... ... scan data file to determine number of record...
&DO INDEX = 1 &REPEAT %INDEX% + 1 &UNTIL %INDEX% = %COUNT%
      &S VALUE [READ %OUNIT% STAT]
            ... Read TEST.DAT to generate symbol values to be used for filename
            building and box coordinate calculation ...
      &IF [EXISTS U1:[LTS.HUSTON]THARP%ID%.PLT -FILE] &THEN
            ... Test for existence of plot file from symbolically derived name
      PLOT U1:[LTS.HUSTON]THARP%ID%.PLT BOX %XMIN% %YMIN% %XMAX% %YMAX%
            ... example of PLOT command with BOX option using generated filename
            and calculated BOX coordinates.
&END
```

If the coordinates were also stored with the attribute data in SAS, then a SAS program could be used to generate an AML program that also uses the PLOT command of ARCPLOT to position and scale the "symbols." The BOX option on this PLOT command is a very useful way to control the position and size of the "symbols."

Examples of Applications

Dimensions of Time

One example of the analysis incorporating the dimension of time includes a map (Figure 1) showing the location and water level history of monitoring wells in the White Oak Creek Watershed, which is a region on the DOE Oak Ridge Reservation being studied for waste cleanup procedures (Voorhees et al. 1988). The line graphs of water level (Y) versus time (X) all have a uniform size and scale to show the details for the hydrographs. The dimensions of the information displayed here could be increased if more than one dependent variable was displayed on the graphs.

Dimensions of Composition

The map shown in Figure 2 displays differences in the size distribution for trees in different forest types on the Walker Branch Watershed (Dale et al. in prep.). The graphs shown here are simple vertical bar charts, but other graphs reflecting sample composition or distribution could also be used. This figure is also an example of one that used an AML to position and scale all of the graphs.

Dimensions of Interrelationships

Statistics provide the capability to determine significant differences in the relationship between two variables for several populations with analysis of covariance. Figure 3 shows a map of the eastern United States with scatter plots containing regression lines showing the changing relationship between surface water pH and elevation for different regions studied by EPA's National Surface Water Survey (Linthurst et al. 1986). Additional dimensions describing relationships between measurements could also be displayed with three-dimensional (3-D) surface displays generated in SAS/GRAPH.

Dimensions of Complex Symbols

Bar charts have been used as complex "symbols" in Figure 2 on a map of the Walker Branch Watershed, showing changes in the stem density for tree size classes from forest sample plots on the watershed. Other examples of graphs that can be converted to complex symbols include stacked vertical, horizontal, 3-D bar charts, and pie charts.

Other Data Dimensions

Examples of other dimensions of data include the incorporation of statistical summaries, analytical results, and brief data tables on the maps. SAS/GRAPH has a GPRINT (SAS 1987) procedure that will convert any portion of text output into

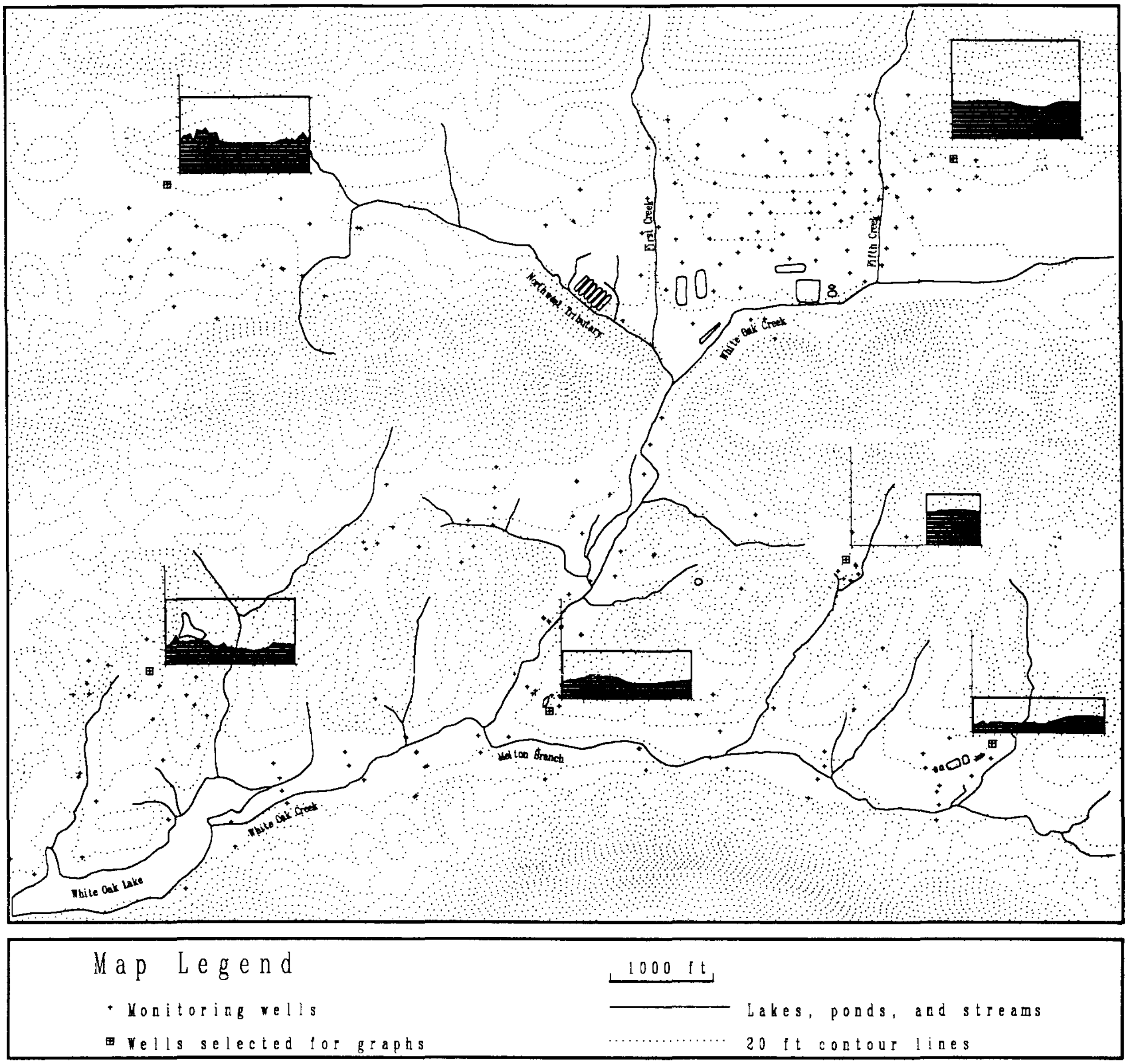

Figure 1. Example of plotting hydrographs for a selected subset of monitoring wells in the White Oak Creek Watershed near ORNL. The hydrographs display data from 10/86 - 10/89. The shaded area represent the water level and the empty area above represents the elevation to the ground surface. All of the hydrographs have the same vertical scale to illustrate the differences between the ground surface elevation and depth to the monitored water level for the wells.

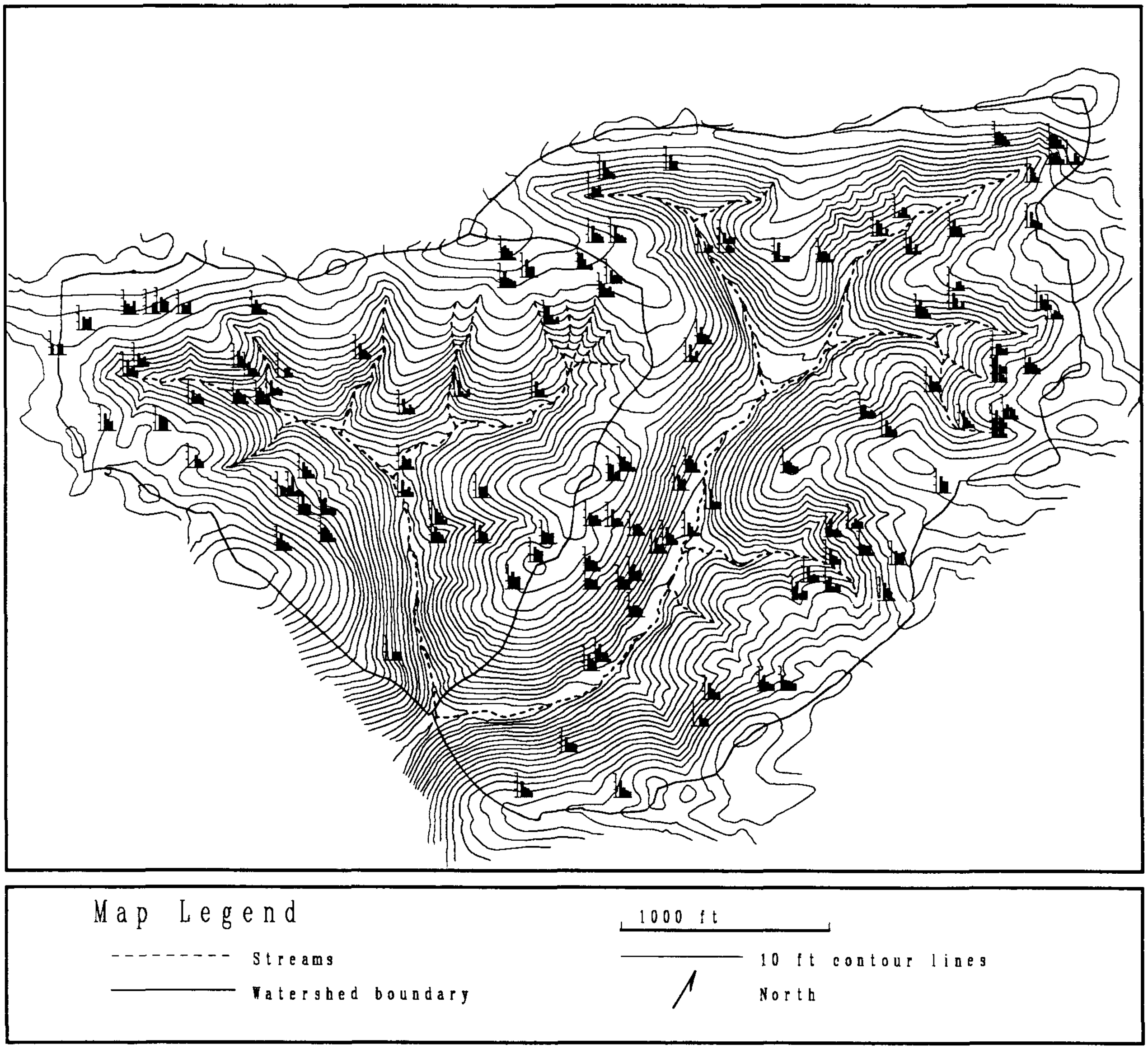

Figure 2. Example of plotting vertical bar charts of the log of stem density by tree size classes for forest study plots on the Walker Branch Watershed near ORNL. The graphs are a combined example of displaying distribution and using graphs as "symbols".

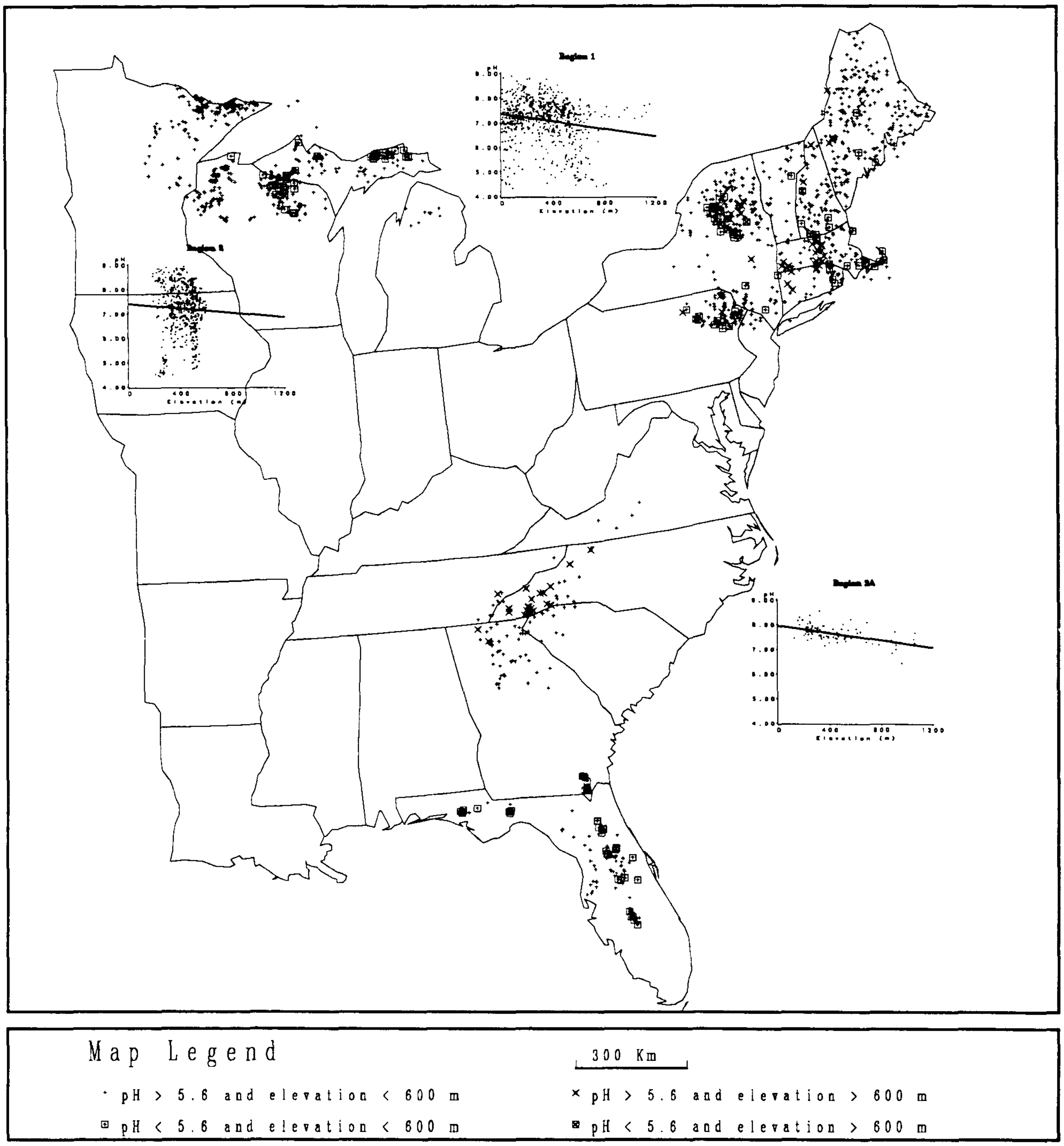

Figure 3. Example of scatter plots showing details of the relationship between pH and elevation by region from National Lake Survey samples. The map also shows the spatial relationship between lakes with low pH and high elevation.

a graphics-based font so that it can be plotted. The conversion of portions of SAS statistical or listing outputs to graphics is a very simple procedure with PROC GPRINT. Examples of maps with this additional type of dimension have not been included here because they are best suited for poster-size maps.

Summary and Conclusions

The graphical interface between SAS/GRAPH and ARC/INFO has resulted in not only the addition of data-intensive graphs in chart form to maps, but also resulted in an increase in the dimensions available for display and analysis. Numerous, small, simplified graphs can become a complex sort of "symbol" for the interpretation of changes in the details of the data. These graphs can show (1) changes in measurements over time, (2) sample composition, (3) statistical distribution, and (4) multivariate relationships. All of these details represent information beyond the spatial and categorical information displayed on the typical map produced by a GIS system. ARC/INFO and SAS/GRAPH were found to be excellent complements to each other in the production of analytical graphics.

The interface is relatively easy to use for the scientist who has an operational knowledge of SAS, SAS/GRAPH, and ARC/INFO. The major challenge in using the interface is to avoid overcrowding the final graphic product by adequately simplifying the map and the graphs to be included. Specifying the automatic page position in page units also has limitations because it presumes detailed knowledge about the location and scale of the final map product.

References

Dale, V. H., L. K. Mann, R. J. Olson, D. W. Johnson, and K. C. Dearstone. in prep. The Long-Term Influence of Past Land Use on the Walker Branch Forest.

ESRI. 1987. ARCPLOT Users Guide, Map Display and Query. Environmental Systems Research Institute, Redlands, CA.

Linthurst, R. A., D. H. Landers, J. M. Eilers, D. F. Brakke, W. S. Overton, E. P. Meier, and R. E. Crowe. 1986. Characteristics of Lakes in the Eastern United States. Volume 1: Population Descriptions and Physico-Chemical Relationships. EPA/600/4-86/007a, U.S. Environmental Protection Agency, Washington, D.C. 136 pp.

Olson, R. J., R. A. McCord, K. C. Dearstone, and S. P. Timmins. 1988. Regional Environmental Studies Using National Databases and GIS. Presentation at Eighth Annual ESRI Users Conference, March 21-25, 1988, Palm Springs, CA.

SAS. 1987. SAS/GRAPH Guide for Personal Computers, Version 6 Edition. SAS Institute, Inc., Cary, NC. 534 pp.

SAS. 1988. SAS/GRAPH Hardware Interfaces for Personal Computers, Version 6 Edition. SAS Institute, Inc., Cary, NC. 353 pp.

Voorhees, L. D., L. A. Hook, M. J. Gentry, M. A. Faulkner, K. A. Newman, and P. T. Owen. 1988. Data Base Management Activities for the Remedial Action Program at ORNL: Calendar Year 1987. ORNL/TM-10694. Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN. 109 pp.

Acknowledgements

The authors would like to acknowledge the contribution of maps and graphics from Kris Dearstone. Aaron Rosen, Dave Shriner, and Steve Hildebrand are also acknowledged for making helpful comments while reviewing the manuscript. Publication No. 3315, Environmental Sciences Division, ORNL.