DISPLAYING FOREST RESEARCH DATA WITH SAS/GRAPH® SOFTWARE, PROC GPLOT, PROC GMAP, and PROC G3D

David C. Chojnacky, Intermountain Research Station
Caroline K. Wraith, Intermountain Research Station
Mark E. Rubey, Intermountain Research Station

ABSTRACT: Graphic data display is an integral part of research at the Intermountain Research Station's Forest Survey Project of the Forest Service, U.S. Department of Agriculture. Graphics are used to explore data relationships and to produce customized graphs and maps of research results. Applications of SAS/GRAPH procedures GPLOT, GMAP, and G3D have been used to study tree growth relationships, map forest resources, and display response surface models. This paper discusses GPLOT, GMAP, and G3D as well as several graphics enhancements, the ANNOTATE feature, NOTE statement, AXIS statement, and LEGEND statement. Also mentioned are experiences running SAS/GRAPH software on a Data General MV8000 minicomputer and plotting results on Hewlett-Packard HP7475 and HP7550 pen plotters. Examples are given of graphs with oversized labels suitable for camera-ready slide production.

INTRODUCTION

Graphic data display is an integral part of research at the Intermountain Station's Forest Survey Project, Forest Service, U.S. Department of Agriculture. Graphics are used to examine data relationships and to produce customized graphs of research results. Because we rely on SAS® System for computing most statistical analyses and data reductions, SAS/GRAPH is a convenient choice for data display. SAS data files from prior analyses are available for SAS/GRAPH without additional input or reformat steps. This paper highlights some of our experiences using PROC's GPLOT, G3D, and GMAP. Focus will be SAS/GRAPH used for visual data analyses and for making customized camera-ready graphs for producing slides. Familiarity with the SAS System is assumed.

PROC GPLOT

Exploratory data plotting with simple two-dimension graphs is a first step in most of our modeling analyses. To enhance use of GPLOT, we often group data into classes to display three or more categories on a single GPLOT. By using data classes for the "zvariable" in the PLOT statement, different colors and shapes can be associated with additional data dimensions. For example, four data categories in a juniper volume relationship are shown in figure 1.

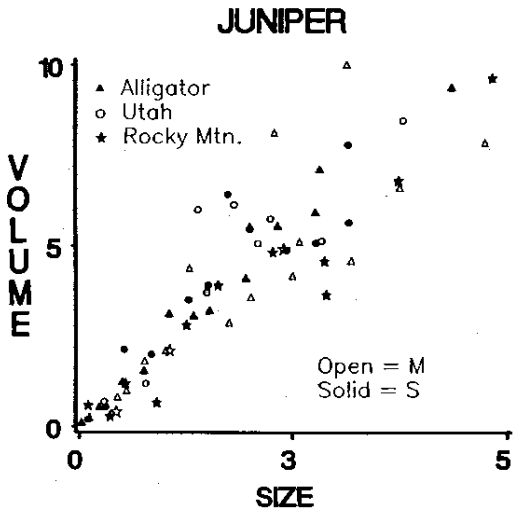

Figure 1--Four data dimensions on a two-dimension graph. In addition to tree volume and size, symbol shape indicates species and symbol fill indicates multiple (M) or single (S) stem characteristics.

Data display and axis labeling (with an AXIS statement) as in figure 1 is relatively fast and simple. However, customizing the legend with NOTE statements in upper left and lower right corners of the figure is by far the most time-consuming task. This is because repeated "trial and error" programming is required for exact placement of legends. Also, font changes from symbol to text within a single NOTE creates spacing and alignment problems. The process is further complicated by differences between our graphics terminal device driver (Data General D460) and our pen plotter device drivers (Hewlett-Packard HP7550 and HP7475).

Because electronic terminals display graphs much faster than do pen plotters, it makes sense to do initial plotting on a terminal. But due to differences between SAS/GRAPH device drivers, a NOTE statement appearing perfectly positioned on our Data General (DG) terminal screen is shifted when the same code is displayed by a Hewlett-Packard (HP) pen plotter. We have also found similar positioning problems among device drivers for other options in the ANNOTATE feature.

In theory, using VSIZE, HSIZE, VPOS, and HPOS GOPTIONS and an appropriate coordinate system should solve the device driver difference problem. We have discovered a few solutions to specific problems by "trial and error", but a general solution for the device driver difference problem is lacking.

The use of trade or firm names in this paper is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

Most useful for our DG/HP device driver situation is use of GOPTIONS VPOS=80 without specifying anything for HPOS, VSIZE and HSIZE.

In addition to plotting data, we use many of the handy options in the SYMBOL statement to overlay regression equations, plot confidence intervals, and smooth data (fig. 2). These are useful for quick testing of hypotheses when constructing mathematical models with our data.

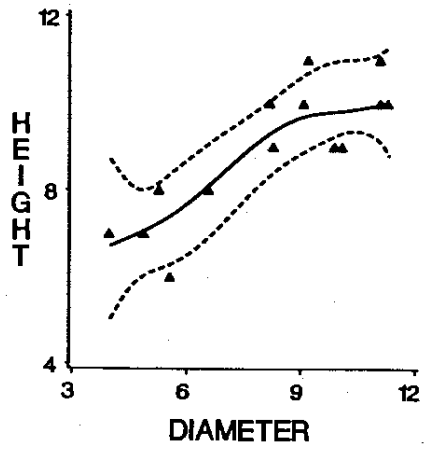

Figure 2--Data smoothed (solid line) by a spline option and 95 percent confidence intervals from cubic regression (dashed lines) overlaid.

PROC G3D

Because many of our analyses result in complex mathematical models, PROC G3D is used to graph finished work (fig. 3). For a model with more than three variables, we generate test data in BY groups to produce multiple plots corresponding to fixed values of the fourth and higher dimension variables.

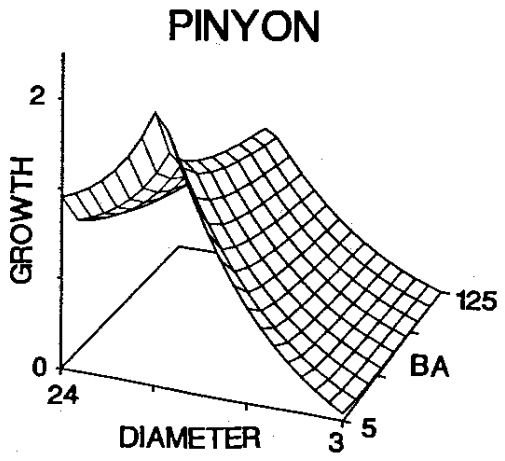

Figure 3--G3D graph of a pinyon pine growth model.

Overlaying more than one response surface on a single graph is not an option with G3D. However, we approximated this effect by graphing points of several response surfaces on a single G3D plot with a SCATTER statement. This is done by adding a small value to the "yvariable" values for each response surface. G3D then plots the "yvariables" almost on top of each other, approximating an overlay of the response surfaces (fig. 4).

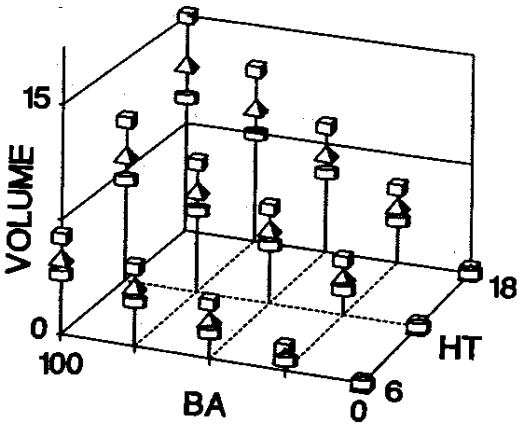

Figure 4--Approximation of three response surfaces overlaid. Cylinders, pyramids, and cubes correspond to volume at three successive time periods.

Customizing options for G3D are lacking. There is no AXIS statement for G3D. All axis labels must be put on with the NOTE statements or with the ANNOTATE feature. Because of the complexities of three axes and device driver differences, we make all labels with PROC GSLIDE and then "cut and paste" to position these labels on G3D graphs.

PROC GMAP

All of the data collected by our Forest Survey Project are organized by counties and States and are geographically referenced. By using PROC GMAP and the ANNOTATE feature we have greatly enhanced the mapping of Forest Survey data (SAS 1985, p. 316-317, for an example). Our data are referenced with Universal Transverse Mercator (UTM) coordinate system, which is not compatible with GMAP. But we have developed a conversion macro (Chojnacky and Tymcio 1987) to convert UTM coordinates to latitude and longitude, the required coordinates for GMAP.

In some instances we prefer to map our data without State or county boundaries shown. This is done by creating a new boundary data set to use in place of the usual SAS/GRAPH county or State boundary data. For example, to map information collected on the Hopi Indian Reservation in Arizona, we created a four observation boundary data set. Each

observation had a unique county code (any integer) and appropriate X,Y coordinates corresponding to latitude and longitude of the northwest, northeast, southeast, and southwest corners of the Hopi Indian Reservation. The four observation boundary data set provided GMAP X,Y coordinates to properly scale the map for the Hopi data, but unique county codes for each boundary observation prevented any boundary lines from being drawn. The resulting Hopi map, including the four boundary points, as shown in figure 5, illustrates the technique.

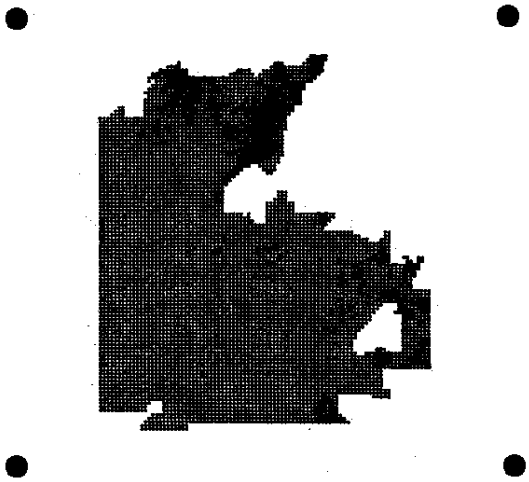

Figure 5--Map of the Hopi Indian Reservation including the four boundary points necessary for producing the map.

One major inconvenience in our experience with the Data General version of GMAP is using ANNOTATE to overlay resource data onto maps. GMAP registers county boundary data precisely to latitude and longitude coordinates, but with some type styles the ANNOTATE data set is registered differently. For example, ANNOTATE data using SPECIAL font type are registered to the center of the graphics cell closest to a given latitude, longitude coordinate. As a result, data symbols for a given county are sometimes plotted just outside the appropriate county boundary (fig. 6).

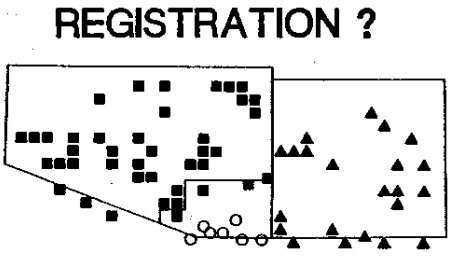

Figure 6--Registration problem illustrated: each southern Arizona county should have similar symbols within its respective boundary.

CONCLUSION

SAS/GRAPH PROC'S GPLOT, G3D, and GMAP are important for our Forest Survey research effort. These SAS products are easy to use and provide many options for creative data analysis. However, when adding the "bells and whistles" for camera-ready copy, considerable effort is required especially if ANNOTATE is used.

Suggestions for improvements to meet our needs include:

1. Adding a mouse to the ANNOTATE feature to position labels interactively at a terminal.

2. Adding a GOPTION to automatically rectify positioning differences when plotting the same SAS code with different device drivers.

3. More enhancements for PROC G3D, such as an AXIS statement and a response surface overlay option.

REFERENCES

Chojnacky, D.C.; Tymcio, R.P. 1987. Conversion of UTM coordinates to geographic coordinates for SAS/GRAPH software PROC GMAP display. In: Proceedings of the 12th annual SAS Users Group International Conference; 1987 February 8-11, Dallas, TX. Cary, NC: SAS Institute, Inc.: 489-494.

SAS Institute, Inc. 1985. SAS/GRAPH user's guide, version 5 edition. Cary, NC: SAS Institute, Inc. 596 p.

Programming details for SAS/GRAPH examples described are available from the authors:

Intermountain Research Station
507 25th Street
Ogden, UT 84401
(801) 625-5402