Paper 134-28

Custom Map Displays Created with SAS/GRAPH® Procedures and the Annotate Facility

Debra Miller, National Park Service, Denver, CO

ABSTRACT

The Annotate facility is a flexible system that you can use alone or with standard SAS/GRAPH procedures to create many types of custom graphical displays. This paper describes a specific application in which the Annotate facility is used with the GMAP procedure to generate custom maps.

Air pollution is frequently a regional phenomenon, as gases and particles can travel for hundreds or even thousands of miles to affect distant locations. Analysts interested in studying the causes of poor air quality in a particular location are interested in knowing the source regions of polluted air masses. The National Oceanic and Atmospheric Administration's Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) Model is used to create 'back trajectories' that estimate the paths traveled by arriving air masses. The model outputs ASCII files that are read into SAS and then merged with pollutant measurements to retrieve back trajectories for particularly hazy or clean days. PROC GMAP is used to create maps of the area of interest, and the trajectories are overlayed on the maps using the Annotate facility.

This paper is intended for users who have had some exposure to SAS/GRAPH.

INTRODUCTION

The National Park Service has an affirmative responsibility given by Congress to protect the lands it manages from the harmful effects of air pollution. This responsibility derives from the Organic Act that created the Park Service as well as the federal Clean Air Act.

To help fulfill this mission, the National Park Service monitors air quality in dozens of national parks. Samplers measure concentrations of small particles suspended in the atmosphere that cause haze that reduces visibility and impairs visitors' ability to observe the scenery. Other monitors measure concentrations of ozone, a gas that is harmful to people as well as vegetation.

There are many sources of man-made emissions that contribute to increased concentrations of these pollutants in our national parks. In many instances, these emissions travel many hundreds of miles before arriving at the parks. Analysts are interested in identifying potential source regions for pollutants that affect air quality at national parks.

The National Oceanic and Atmospheric Administration (NOAA) has developed a computer model that inputs meteorological data and generates 'back trajectories' for any desired location in the United States. Each back trajectory gives the path traveled by an air mass arriving at that location at a particular time and day. This model is called the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) Model.

The HYSPLIT model outputs text files that contain the back trajectories. The location of the air mass arriving at the chosen location is given in one-hour time steps, beginning with the requested time and moving backward in time. Back trajectories may be requested for up to 5 days in length, resulting in 120 individual air mass locations in latitude, longitude, and height.

In order to be useful, the back trajectory files need to be displayed graphically on a map. This is done using SAS software. First, the files are read into SAS to generate SAS data sets. The data sets are combined with air pollutant concentration data so that back trajectories associated with particularly clean or dirty days can be identified. These back trajectories are then overlayed on a map using SAS/GRAPH and the Annotate facility.

EXAMPLE BACK TRAJECTORY MAPS

Two examples of the output of this program are shown in Figure 1. The maps in the figure show back trajectories that arrived at Voyageurs National Park in northern Minnesota during the months of May through September of 2000 and 2001. The map in the upper portion of the figure shows the back trajectories that arrived on relatively 'clean' days, when the sulfate particle concentration measurements fell within the lowest 15% of the measurements collected in each year.

Trajectories are indicated on the map by colored lines, and the colors correspond to the month in which the air mass arrived. Individual dots along the lines show the 1-hour location time steps output by the HYSPLIT model. Trajectories have been truncated in length to 48 hours in the figure to keep the display uncluttered.

The map in the lower portion of the figure shows back trajectories that arrived on days when sulfate particle concentration measurements fell within the highest 15% of each year. It is interesting to note that the majority of air masses reaching Voyageurs on relatively clean days arrived from the north, while most of the air masses arriving on relatively dirty days approached from the south and southwest. This type of information can be helpful in identifying potential source regions for pollutants.

To generate these maps I used the Annotate facility along with the SAS/GRAPH procedure GMAP. The Annotate facility can be used with many SAS/GRAPH procedures to enhance the outputs and create custom displays, or it can be used to generate stand-alone custom graphics. Below is a brief introduction to the Annotate facility; there are many more features available than will be discussed here.

USING THE ANNOTATE FACILITY

To use the Annotate facility you must create an Annotate data set. An Annotate data set contains variables with specific names that SAS recognizes; the values of the variables provide instructions for creating graphics. These Annotate variables define *what* annotation is to be generated, *where* the annotation is to be placed, and *how* the annotation is to be drawn. Other variables that may

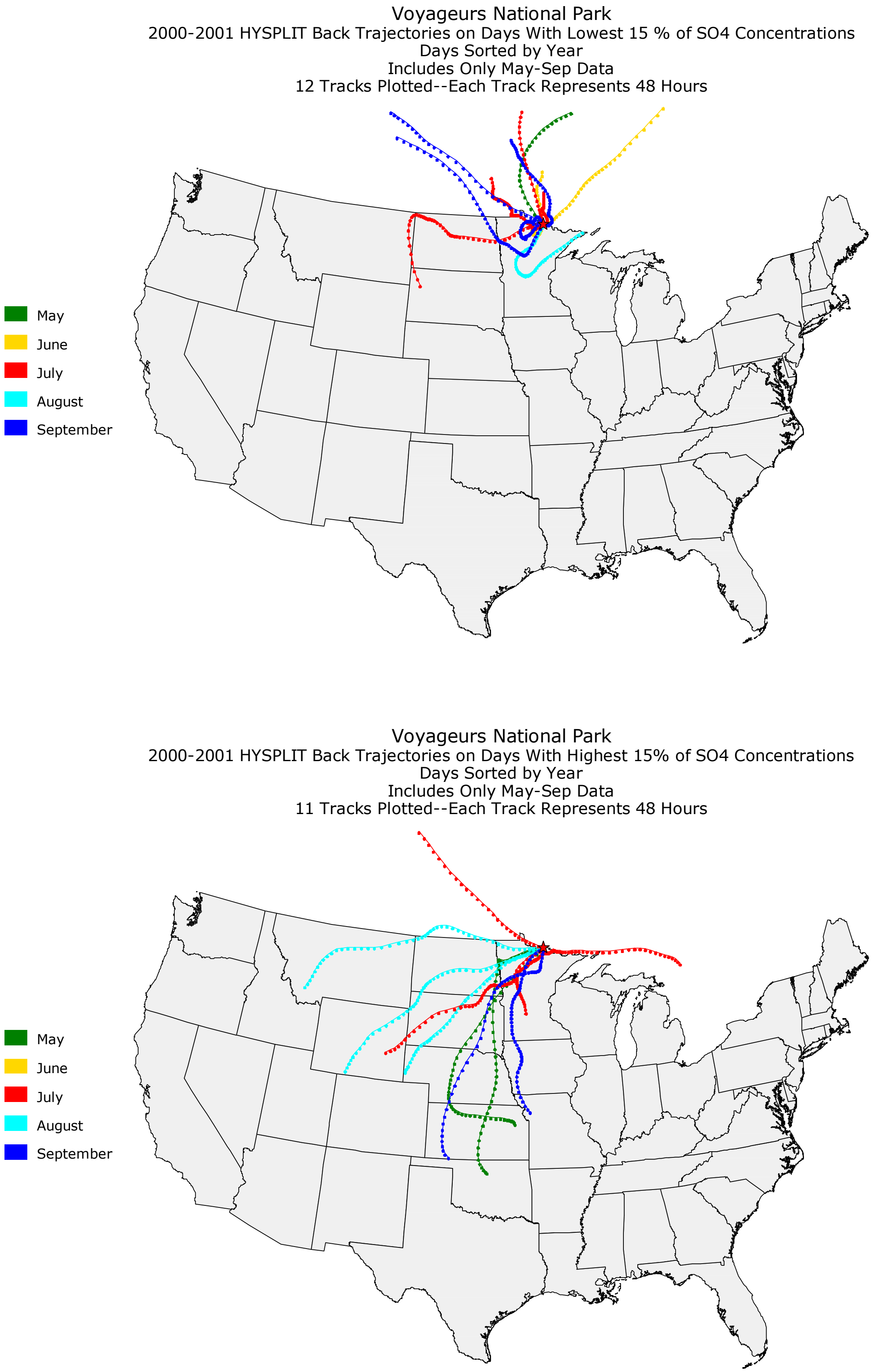

Figure 1. Sample Back Trajectory Maps

be present in an Annotate data set are ignored by the Annotate facility when generating graphics. Some examples of the Annotate variables are discussed below.

SAS determines *what* graphics are to be created from the values of the Annotate variable FUNCTION. Some possible values are:

LABEL: creates a text annotation
DRAW: draws a line
BAR: draws a bar
MOVE: updates the current graphics drawing position without creating any graphics
SYMBOL: places a symbol such as a star, square or circle

There are also other values that can be assigned to the variable FUNCTION that make it possible to create a wide variety of graphical elements.

SAS determines *where* you want graphics to appear from the values of several other variables. The location is determined from the Annotate variables X and Y (and Z if using the G3D procedure), which specify the coordinates to be used for the annotation. For functions such as ‘bar’ and ‘draw’, the X and Y coordinates specify the ending location of the graphics element. The starting location for these elements is determined from two internal SAS variables, XLAST and YLAST. These variables retain the ending coordinates of the last graphics element drawn. They cannot be assigned explicitly, but are automatically updated by non-text functions such as ‘bar’ and ‘draw’. You can also update them without generating any graphics by using the FUNCTION value ‘move’.

Two analogous internal variables, XLSTT and YLSTT, retain the coordinates of the last text element created, and are automatically updated by the FUNCTION variable values ‘label’ and ‘symbol’.

In order to place annotations, you also need to specify the type of coordinate system to be used. Coordinate systems can be absolute or relative. An absolute coordinate system has a fixed origin, and the numerical X and Y variables specify the precise location of the annotation relative to the origin; in a relative system the coordinates are located relative to another graphics element.

The coordinate system chosen specifies the total available area in which you can place annotations. The total area is defined as one of the following: the data area, which is the area in which data may be displayed (such as the area defined by coordinate axes); the graphics output area, which is all available space; or the procedure output area, which is the available graphics area reduced to allow space for titles and footnotes. These three area definitions are shown in Figures 2-4 below in the context of the GPLOT procedure, with each illustrated area highlighted in gray:

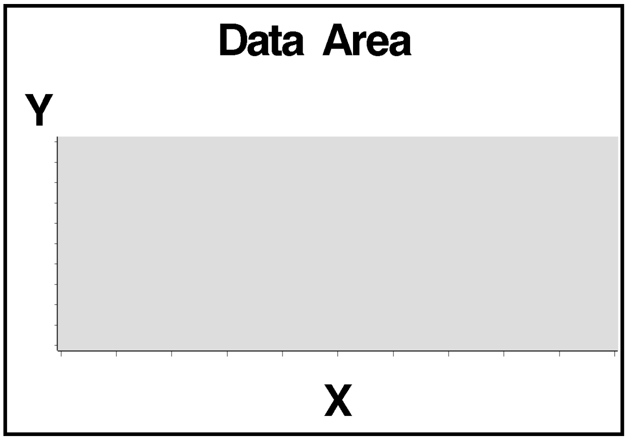

Figure 2.

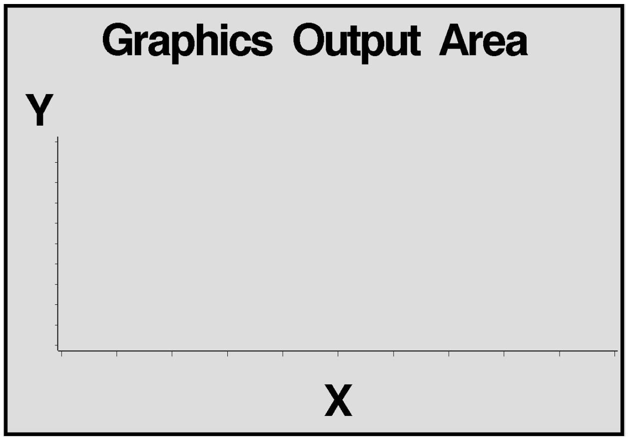

Figure 3.

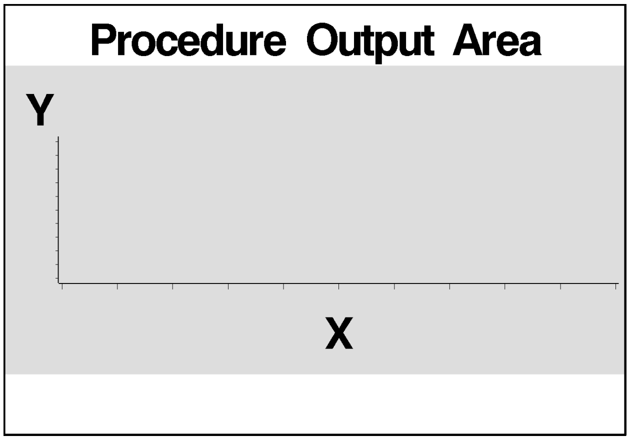

Figure 4.

The three areas are also shown below in Figures 5-7 using the GMAP procedure:

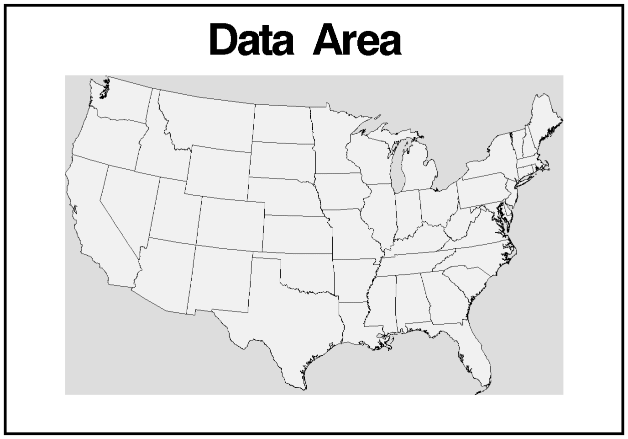

Figure 5.

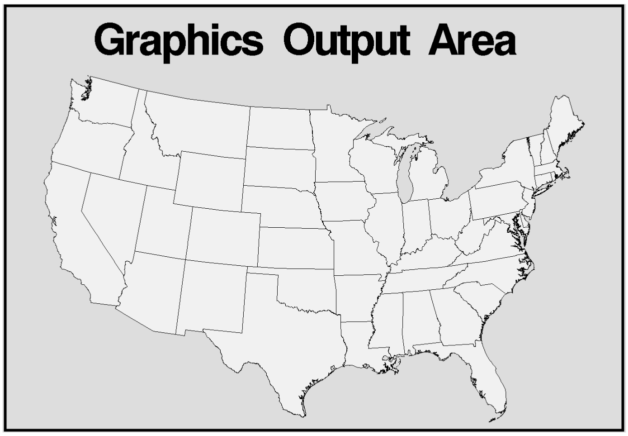

Figure 6.

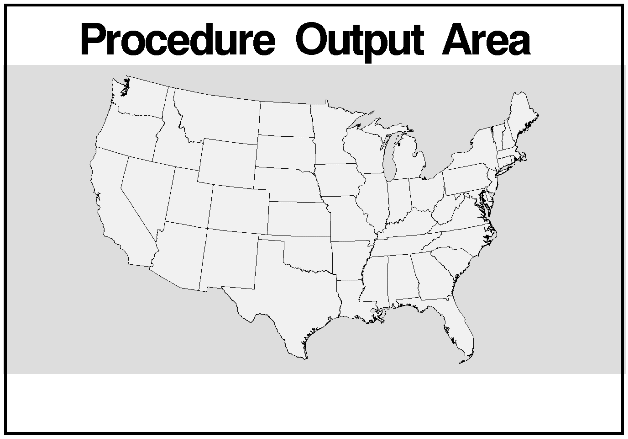

Figure 7.

The coordinate system used also determines the type of units. Units may be data values (if using a data coordinate system), percentages of the total available area (if using a procedure or graphics output coordinate system), or cells (also used with a procedure or graphics output coordinate system). The number of cells available depends on the particular graphics device being used (such as a specific terminal or printer), and the number of rows and columns made available by the graphics options and device parameters in effect.

The coordinate systems to be used are determined by the values of the variables XSYS and YSYS (as well as ZSYS for PROC G3D), and can be different for each axis. The coordinate system values can also be changed in each observation of the Annotate data set if desired.

Finally, there are Annotate variables that tell SAS *how* an Annotation is to be drawn by defining its attributes. Variables in this group include COLOR, SIZE, STYLE, POSITION, and LINE. Some of these attribute variables, such as COLOR and STYLE, produce results that vary with the value of the FUNCTION variable. For example, the value of the STYLE variable might define the type of font used if the FUNCTION variable equals 'label' or 'symbol', or the type of fill pattern to be used if the FUNCTION variable equals 'bar'.

CREATING THE ANNOTATE DATA SET

The sample of code below illustrates the generation of the Annotate data set BT_ANNO. It begins with the data set BTPLOT. This data set contains all of the back trajectories associated with the clean or dirty days; each observation in the data set contains the location of an air mass for an individual 1-hour time step of a particular back trajectory.

The first step in creating the Annotate data set from the back trajectories is to generate an observation in the output data set that will result in a dot being placed at that location. The latitude and longitude at each 1-hour step are converted to radians, since this is required by the GMAP procedure for plotting. **(1)** The variables X and Y are used by the Annotate facility to locate the annotations. Additional variables that are carried over from the BTPLOT data set such as LAT, LONG, and MONTH are ignored by the facility.

Units for the X and Y axes are chosen to be data values. **(2)** Annotate variable FUNCTION is set to 'label', variable STYLE is set to 'special', and TEXT is set to 'J'. **(3)** This causes a text label to be placed at the location of the time step. Using a value of 'J' with the Special font generates a dot at the location. The color of the label is assigned according to the month of the observation **(4)**, and then the observation is output. Because the text function 'label' is used, the graphics internal variables XLAST, YLAST are not updated in this step.

The next step is to generate a line from each time step to the next. This connects the dots in the back trajectory. The data set variable FIRST is set in an earlier data step to 1 at the first time step of each back trajectory, and 0 for all subsequent time steps. If the observation is the first one in the back trajectory **(5)**, then the Annotate variable FUNCTION is set to 'move'. This updates the internal variables XLAST and YLAST to the location of the first step in the back trajectory without generating any graphics.

If the variable FIRST is set to 0 **(6)**, indicating that the observation is not the start of the back trajectory, the Annotate variable FUNCTION is set to the value 'draw'. This results in a line being drawn from the XLAST, YLAST location to the location of the current observation. The 'draw' function also automatically updates the XLAST, YLAST variables to the current observation for use by the next graphics command. This observation is then also output to the Annotate data set.

```
data bt_anno;
  length text $20 color function style
$8;
  set btplot;
/* Convert latitude and longitude to */
/* radians.                          */
  x=(atan(1)/45)*long; (1)
  y=(atan(1)/45)*lat;
/* Coordinate systems are data */
/* values.                     */
  xsys = '2';  ysys = '2'; (2)
/* Annotate is applied after the */
/* procedure output.             */
  when = 'A';
  state=%eval(&state);
/* Generate text label at airmass */
/* location (a dot).              */
  function='label'; (3)
  style='special';
  text='J';
  position='5';
  size=0.8;
/* Assign colors by month. */
  select (month); (4)
    when (1) color="&color1";
    when (2) color="&color2";
/*  (...same for months 3 to 11)  */
    when (12) color="&color12";
    otherwise color="&default";
end;
output;
/* These statements below put the    */
/* lines between the dots.           */
/* The move statement is necessary   */
/* to first update the XLAST, YLAST  */
/* coordinates at the beginning of   */
/* each back trajectory; after that  */
/* a draw frunction is used to       */
/* create the lines.                 */
if first=1 then do; (5)
     function='move';
     output;
     end;
if first=0 then do; (6)
     function='draw';
     output;
     end;
run;
```

CREATING THE MAP

The map is created using the GMAP procedure, which is part of SAS/GRAPH. PROC GMAP requires two types of data sets as inputs: a map data set and a response data set. The map data set contains the boundaries of the map. Boundary information must be contained in the variables X and Y. These values are Cartesian coordinates. The response data set contains the variable of interest that you wish to plot, such as state population. Both data sets must have a common ID variable that identifies the map areas.

Map locations are specified as latitude and longitude and then projected using the GPROJECT procedure. This converts the spherical coordinates to Cartesian coordinates that can be plotted on a flat map. The latitude and longitude must be in radians, and have the variable names 'X' and 'Y'. Three types of projections are available; in this case I used the Albers equal-area projection. The output data set contains the variables X and Y as Cartesian coordinates.

With the GMAP procedure you can create block, prism, surface, and choropleth maps. This program uses PROC GMAP to create a choropleth map. This is a map where the levels of the response variable determine the colors and patterns of the map areas. If you specify the same data set as both the map data set and the response data set, PROC GMAP creates a map containing only the boundaries of the map areas.

The code below shows how the map is created. The data set MAPS.STATES, which is supplied by the SAS Institute and contains the state boundaries, is stored to the data set JUSTMAP. This data set contains unprojected latitude and longitude locations in radians stored in the variables X and Y. The data set can be clipped if desired by specifying maximum and minimum latitude and longitude parameters using the GPROJECT procedure and specifying PROJECT=NONE. The output data set PROJMAP thus also contains unprojected values.

```
/* First clip the map set using  */
/* project=none.                 */

proc gproject data=justmap out=projmap
   longmax=&longmax latmax=&latmax
   longmin=&longmin latmin=&latmin
   project=none;
   id state;
run;
```

In order to use the Annotate data set with the map data set, both must be projected together using PROJECT= ALBERS. It is important to use the GPROJECT option ASIS, which allows duplicate observations in the input data set, any time you are projecting map and Annotate data together.

```
/* Combine the annotate data sets */
/* with the unprojected map data  */
/* set to be projected together.  */

data all;
   set bt_anno projmap;
run;

/* Project the maps. Use option ASIS */
/* to allow duplicate observations   */
/* in the data set whenever Annotate */
/* observations are included in a    */
/* projection.                       */

proc gproject data=all out=allp asis
project=albers ;
   id state;
run;
```

After being projected together, the map data set and annotate data set are again separated into different data sets, MAP and ANNO.

```
/* Separate the projected data back  */
/* into two distinct data sets.      */
```

```
/* ANNTATE=1 designates the observa- */
/* tion as an annotate observation   */

data anno map;
   set allp;
   if anntate=1 then output anno;
    else output map;
run;
```

Finally, the map can be generated using the GMAP procedure. The map data set is specified to be both the map data set and the response data set so that only the state outlines are plotted. Using the CHORO statement option ANNO=ANNO tells the procedure to use the data set ANNO to place annotations on the base map. This overlays the back trajectories on the map.

```
/* Produce the map with annotation. */

proc gmap data=map map=map ;
   id state;
   choro state / anno=anno
   coutline=black nolegend;
   pattern1 r=50 v=solid c=grayf0;
run;  quit;
```

CONCLUSION

Although it may seem a bit daunting at first glance, the Annotate facility is not difficult to use. It is really just a matter of building a data set of graphics commands. These commands tell SAS what, where, and how to draw the desired annotations. You can use Annotate with many standard SAS/GRAPH procedures such as PROC GCHART, PROC GPLOT, and PROC GMAP to enhance the procedure output, thus creating more informative displays.

This paper has provided only a brief introduction to the topic; there are many more tools in the Annotate facility than could be covered here. Hopefully you will find many new ways to use the facility to create valuable custom graphics displays.

REFERENCES

Draxler, Roland. “HYSPLIT 4 Online User’s Guide.” September 2002. <http://www.arl.noaa.gov/data/models/hysplit4/win95/user_guide.pdf> (December 9, 2002).

Gebhart, Kristi and Malm, William (1994), “Source Attribution and Statistical Summary of Data Measured at Grand Canyon National Park During 1989-1991”, in *Aerosols and Atmospheric Optics: Radiative Balance and Visual Air Quality*, Volume B, 1098-1124. Air and Waste Management Association and American Geophysical Union.

SAS Institute, Inc. (1999), *SAS/GRAPH Software: Reference, Version 8,* Cary, NC: SAS Institute Inc.

ACKNOWLEDGMENTS

I am indebted to Kristi Gebhart of the National Park Service for the use of her programs to generate and combine multiple HYSPLIT back trajectory output files. I also wish to thank David Joseph of the National Park Service for his invaluable comments on this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Debbie Miller
National Park Service Air Resources Division
P.O. Box 25287
Denver, CO 80225-0287
Work Phone: (303) 987-6947
Fax: (303) 969-2822
Email: debbie_c_miller@nps.gov