SESUG Paper 191-2018

A Macro for Creating Time Trend Plots with Four Y Axes

Dennis J. Beal, Leidos, Oak Ridge, Tennessee

ABSTRACT

Time trend plots are used often in statistical analysis to see how variables change over time. SAS® can easily produce time trend plots with one or two Y axes, but displaying more than two Y axes can be challenging. If there are three or more Y variables with very different scales that are to be compared together across time on the same graph, placing these Y variables simultaneously on a two-dimensional graph can be done. This paper shows SAS code in a macro that uses annotation with `PROC GPLOT` to produce time trend graphs with four Y axes. This paper is for intermediate SAS users of Base SAS® and SAS/GRAPH®.

Key words: time trend plots, multiple Y axes, SAS/GRAPH, `PROC GPLOT`

INTRODUCTION

One statistical graph that is commonly used by data analysts to examine the relationship of variables with time is the time trend plot. This plot has the variable of interest on the Y axis and time on the X axis. Long-term trends, seasonal variability, short-term trends and changes in results over time can be examined using time trend plots. Two variables can also be plotted easily with time by having one variable on the left Y axis and the other variable on the right Y axis by using `PLOT2` within SAS/GRAPH. If the data analyst wants to plot more than two variables with different scales across time on the same graph, the annotation facility can be used.

An example using environmental data is shown by plotting the variables specific conductivity, pH, water elevation, and daily rainfall on separate Y axes and time on the X axis. Two of the variables are plotted on the left Y axis, while the other two are plotted on the right Y axis using `PROC GPLOT` and the annotation facility. By using different colors for each variable the data analyst can better examine how these variables interrelate across time.

This paper will show the SAS code in a macro that generates high quality time trend plots with four Y axes. The SAS code presented uses the SAS System for personal computers version 9.3 running on a Windows® 7 Professional platform.

EXAMPLE DATA SET

In order to produce the time trend plot with four Y axes, an example data set consists of the following variables:

- STATION has the groundwater well station name (e.g., Well A).
- DATETIME consists of the date and time the measurements were made in SAS `datetime19.` format from March 2, 2017 through May 8, 2017.
- WL_ELEV consists of the water elevation measurements in feet mean sea level (msl) within groundwater Well A which ranges from 835 to 839 feet msl.
- PH consists of the pH measurement (unitless) collected which ranges from 6.4 to 7.1 pH units.
- CONDUCTIVITY consists of the specific conductivity measurements in microsiemens per centimeter (µS/cm) which ranges from 670 to 705 µS/cm.
- RAIN consists of the daily total rainfall (inches) from hourly precipitation readings from a nearby meteorological station which ranges from 0 to 1.5 inches per day.

The goal is to plot each of these four variables with its own Y axis and scale with the common variable DATETIME on the X axis. Because these variables have very different scales, each variable must have its own Y axis. Two variables will share the left Y axis and the other two variables will share the right Y axis.

WL_ELEV is the primary variable of interest because we want to see how the other variables affect the water level elevation within the well across time.

The variables WL_ELEV, PH, and CONDUCTIVITY are shown as lines, each with its own color. However, RAIN is shown as vertical bars for each daily rainfall total at the bottom of the plot in green.

SAS CODE

The SAS code that generates the plot is described in this section. First, because the rainfall data is shown as bars at the bottom of the plot, we must subset the rain data and assign each value of RAIN zero in order to create a bar for each day. Each bar must have both a minimum and maximum value per day.

```
data rain0;
  set dataset;
    where station = 'RAIN';
  rain = 0; run;
```

Combining the data set `rain0` with the original data set provides a minimum and maximum value of RAIN for each day in order to create a bar.

```
data alldata;
  set dataset rain0; run;
```

Given that there are multiple groundwater wells within the data set, a SAS macro is defined to efficiently plot each well. The macro variable PATHPLOT assigns the subdirectory where SAS will store the resulting plot.

```
%let pathplot = C:\SAS\My SAS Files\;

%macro plotem(STA, ELEV_MIN, ELEV_MAX, ELEV_BYVAR, COND_MIN, COND_MAX,
COND_BYVAR, PH_MIN, PH_MAX, PH_BYVAR, RAIN_MIN, RAIN_MAX, RAIN_BYVAR, PCT);
```

The following macro variables are specified based on the data for Well A:

STA	= Well A	= the station name
ELEV_MIN	= 835	= the minimum value of the left Y axis for water level elevation
ELEV_MAX	= 840	= the maximum value of the left Y axis for water level elevation
ELEV_BYVAR	= 1	= the value that the major tick marks for elevation are incremented by
COND_MIN	= 670	= the minimum value of the left Y axis for conductivity
COND _MAX	= 705	= the maximum value of the left Y axis for conductivity
COND_BYVAR	= 5	= the value that the major tick marks for conductivity are incremented by
PH_MIN	= 6.4	= the minimum value of the right Y axis for pH
PH _MAX	= 7.1	= the maximum value of the right Y axis for pH
PH _BYVAR	= 0.1	= the value that the major tick marks for pH are incremented by
RAIN_MIN	= 0	= the minimum value of the right Y axis for daily rainfall
RAIN_MAX	= 8	= the maximum value of the right Y axis for daily rainfall for the entire right Y axis
RAIN_BYVAR	= 0.5	= the value that the major tick marks for daily rainfall are incremented by
PCT	= 0.50	= 50 percent of vertical graph area reserved for conductivity and pH

The actual scale on the right Y axis is for daily rainfall. Therefore, the scales for conductivity and pH must be rescaled into daily rainfall units (inches). The variable `YAXIS_MIN` is the starting value of the scales for conductivity and pH in daily rainfall units. The variable `PH_RESCALED` converts pH units into daily rainfall inches for the right Y axis. The variable `COND_RESCALED` converts conductivity units into daily rainfall inches for the right Y axis (although this will appear on the left Y axis). The variable `PH BYVAR RESCALED` converts the incremental “by” variable for pH into daily rainfall inches for the right Y axis. Similarly, the variable `COND_BYVAR_RESCALED` converts the incremental “by” variable for conductivity into daily rainfall inches for the right Y axis. The variables `YAXIS_MIN`, `PH_BYVAR_RESCALED` and `COND_BYVAR_RESCALED` are then stored into macro variables to be used later in `PROC GPLOT`.

```
data plotem;
  set alldata;
    where station in ("&STA" 'RAIN');
  YAXIS_MIN = &PCT * &RAIN_MAX;
  ph_rescaled = (ph - &PH_MIN) / (&PH_MAX - &PH_MIN) * (&RAIN_MAX -
   YAXIS_MIN) + YAXIS_MIN;
  ph_byvar_rescaled = &PH_BYVAR / (&PH_MAX   - &PH_MIN  ) * (&RAIN_MAX -
   YAXIS_MIN);
  cond_rescaled = (conductivity - &COND_MIN) / (&COND_MAX - &COND_MIN) *
   (&RAIN_MAX - YAXIS_MIN) + YAXIS_MIN;
  cond_byvar_rescaled = &COND_BYVAR / (&COND_MAX - &COND_MIN) * (&RAIN_MAX
   - YAXIS_MIN);
  call symput('YAXIS_MIN', yaxis_min);
  call symput('PH_BYVAR_RESCALED', ph_byvar_rescaled);
  call symput('COND_BYVAR_RESCALED', cond_byvar_rescaled);
  run;

proc sort data=plotem; by datetime ; run;
```

The name of the plot is specified and a `goptions` statement is executed.

```
filename plot "&pathplot.&STA elev ph cond rain with 4 Y axes.cgm";

goptions reset=all display gsfmode=replace noprompt gsfname=plot
   device=EMF targetdevice=EMF rotate=portrait htext=1.0
   ftext="Times New Roman" horigin=0 in vorigin=0 in vsize=6.5 in
   hsize=6.8 in cback=white ;
```

The data set `anno1` creates the Y axis for pH as well as minor tick marks.

```
data anno1;  ** for pH axis;
  length function $ 8;
  retain xsys '2' ysys '2' color 'orange' when 'a';
  size=1.05;
  function='move';  x='06may2017:00:00:00'dt; y=&YAXIS_MIN; output;
  function='draw';  x='06may2017:00:00:00'dt; y=&RAIN_MAX;  output;
  do i = &YAXIS_MIN to &RAIN_MAX by &PH_BYVAR_RESCALED;
    function='move';  x='06may2017:00:00:00'dt; y=i ; output;
    function='draw';  x='07may2017:00:00:00'dt; y=i ; output;
    function='label'; x='09may2017:00:00:00'dt; y=i+0.05;
   text=trim(left(put(&PH_MIN + (&PH_MAX - &PH_MIN)/(&RAIN_MAX -
   &YAXIS_MIN)*(i - &YAXIS_MIN), 4.1))); output;
  end;
  do i = &YAXIS_MIN + &PH_BYVAR_RESCALED / 2 to &RAIN_MAX -
   &PH_BYVAR_RESCALED / 2 by &PH_BYVAR_RESCALED;    ** minor tick marks;
    function='move';  x='06may2017:00:00:00'dt; y=i ; output;
    function='draw';  x='06may2017:12:00:00'dt; y=i ; output;
  end;
run;
```

Similarly, the data set `anno2` creates the Y axis for conductivity as well as minor tick marks.

```
data anno2;  ** for conductivity axis;
  length function $ 8;
  retain xsys '2' ysys '2' color 'red' when 'a';
  size=1.05;
  function='move';  x='02mar2017:00:00:00'dt; y=&YAXIS_MIN; output;
  function='draw';  x='02mar2017:00:00:00'dt; y=&RAIN_MAX; output;
```

```
    do i = &YAXIS_MIN to &RAIN_MAX by &COND_BYVAR_RESCALED;
      function='move';  x='02mar2017:00:00:00'dt; y=i ; output;
      function='draw';  x='03mar2017:00:00:00'dt; y=i ; output;
      function='label'; x='05mar2017:00:00:00'dt; y=i+0.05;
     text=trim(left(put(&COND_MIN + (&COND_MAX - &COND_MIN)/(&RAIN_MAX -
     &YAXIS_MIN)*(i - &YAXIS_MIN), best5.))); output;
    end;
    do i = &YAXIS_MIN + &COND_BYVAR_RESCALED / 2 to &RAIN_MAX -
     &COND_BYVAR_RESCALED / 2 by &COND_BYVAR_RESCALED;  ** minor tick marks;
      function='move';  x='02mar2017:00:00:00'dt; y=i ; output;
      function='draw';  x='02mar2017:12:00:00'dt; y=i ; output;
    end;
  run;
```

The data set `anno3` inserts the station name label at the top of the plot.

```
  data anno3;
    length function color style $ 8 text $15;
    function='LABEL';
    size=1.3;
    xsys='1';
    ysys='1';
    hsys='4';
    y = 98;
    style=' ';
    x = 43;
    text="&STA";
    color='black';
    when='a';
    position='6';
   output;
   run;
```

The data set anno combines data sets `anno1` and `anno2`.

```
  data anno;
    length text $9;
    set anno1 anno2;
  run;
```

`PROC GPLOT` is executed to produce the plot.

```
  proc gplot data=plotem annotate=anno3;
    plot wl_elev*datetime / vaxis=axis1 haxis=axis2 vm=1 hm=4 grid ;
    plot2 rain*datetime ph_rescaled*datetime cond_rescaled*datetime /
     overlay vaxis=axis3 nolegend vm=1 annotate=anno ;
    format wl_elev 9.0 rain 4.1 ;
    axis1 width=2 value=(h=1.05 c=blue) order=(&ELEV_MIN to &ELEV_MAX by
&ELEV_BYVAR) label=(f="Times New Roman" a=90 r=0 h=1.05 j=l c=blue
"Groundwater Elevation (ft msl)" j=r c=red "Specific Conductivity (µS/cm)");
    axis2 width=2 offset=(0.05 in) order=('02mar2017:00:00:00'dt to
'11may2017:00:00:00'dt by 432000) label=(f="Times New Roman" j=c h=1.05 'Date
(2017)') value=(h=1.0
       t=1 c=black j=c '3/2'
       t=2 c=black j=c '3/7'
       t=3 c=black j=c '3/12'
       t=4 c=black j=c '3/17'
       t=5 c=black j=c '3/22'
       t=6 c=black j=c '3/27'
       t=7 c=black j=c '4/1'
```

```
      t=8 c=black j=c '4/6'
      t=9 c=black j=c '4/11'
      t=10 c=black j=c '4/16'
      t=11 c=black j=c '4/21'
      t=12 c=black j=c '4/26'
      t=13 c=black j=c '5/01'
      t=14 c=black j=c '5/06'
      t=15 c=black j=c '5/11');
        axis3 width=2 value=(h=1.05 c=green) order=(&RAIN_MIN to &RAIN_MAX
   by &RAIN_BYVAR) label=(f="Times New Roman" a=90 r=0 h=1.05 j=l c=green
   "Daily Rainfall (in.)" j=r c=orange 'pH');
   label wl_elev='00'x datetime='00'x rain='00'x;
   symbol1 f='marker' h=0.8 v=none i=join    l=1 w=1 c=blue   cv=blue ;
   symbol2 f='marker' h=0.6 v=none i=box00tf l=2 w=1 c=green bwidth=1
   co=green  ;
   symbol3 f='marker' h=0.8 v=none i=join  l=1 w=1 c=orange cv=orange ;
   symbol4 f='marker' h=0.8 v=none i=join  l=1 w=1 c=red    cv=red  ;
   run; quit;
%mend plotem;
```

The macro is executed with the following statement:

```
%plotem(Well A, 835, 840, 1, 670, 705, 5, 6.4, 7.1, 0.1, 0, 8, 0.5, 0.5)
```

THE GRAPH

After the macro is executed, the graph is edited in the SAS 9.3 ODS Graphics Editor. The unneeded portion of the blue axis is deleted at the top of the left Y axis, and axis labels are moved next to the correct axes. The unneeded portion of the green axis is deleted beyond 1.5 on the right Y axis. The pH axis is moved out of the graph area to the top of the right Y axis. The file is saved and closed. Then to save these changes the file is printed by going to File > Print. Click “Print to File” and OK. The finished graph is then brought into Microsoft Word. **Figure 1** shows the groundwater elevation is directly affected by rainfall. Conductivity also has an effect; while pH appears to have little effect on groundwater elevation.

CONCLUSION

When creating time trend plots, sometimes two Y axes are not enough. There are times when the data analyst wants to see the relationship between three or four Y variables over time. If these Y variables have very different scales, each must have its own axis. This paper shows SAS code that can produce a time trend plot with four Y axes using `PROC GPLOT` and the annotation features. By applying different colors for the four Y variables, the graph becomes easier to interpret.

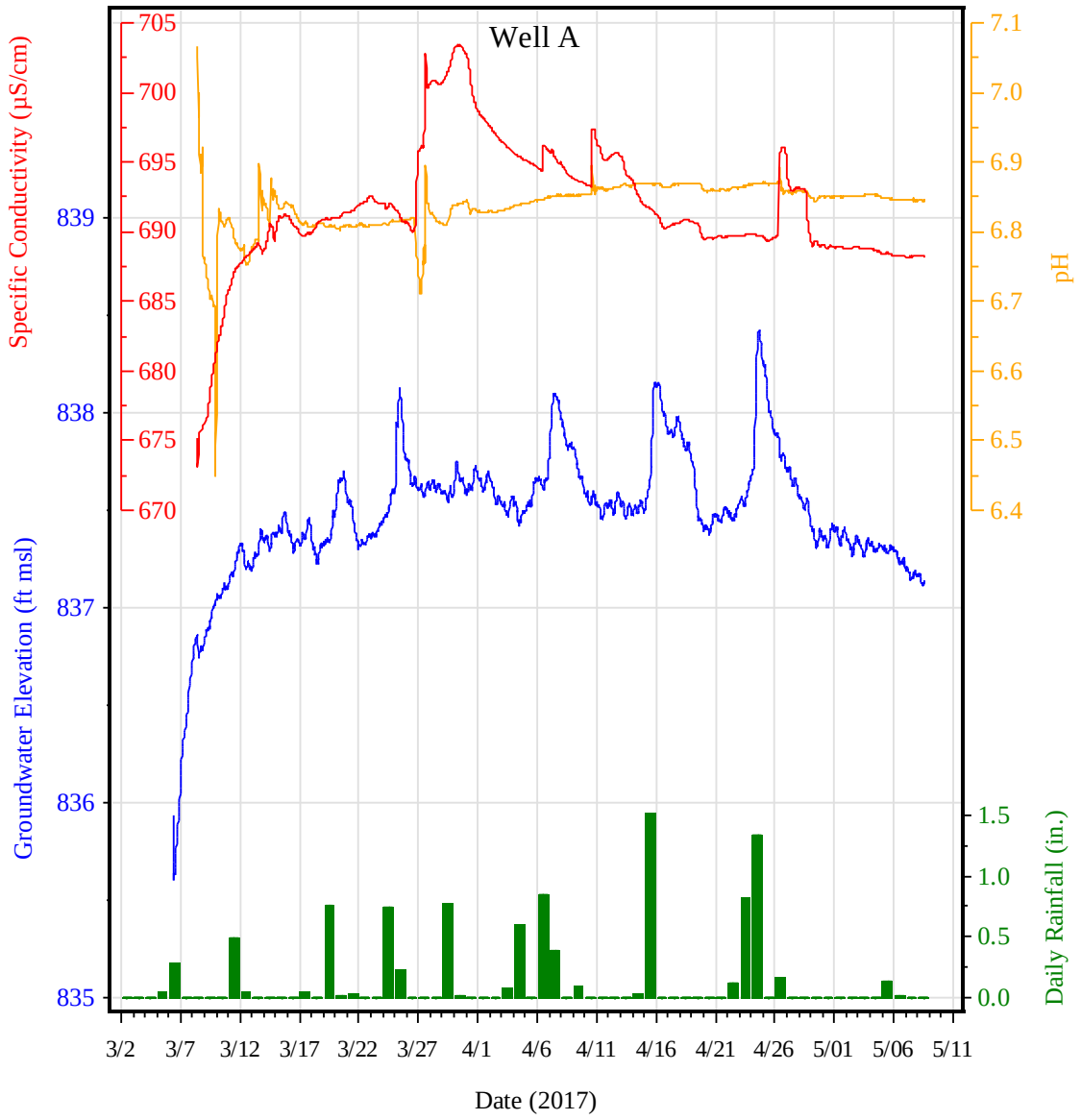

Figure 1. Time trend plot of groundwater elevation, conductivity, and pH with daily rainfall for Well A

CONTACT INFORMATION

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

Dennis J. Beal, Ph.D.
Senior Statistician/Risk Scientist
Leidos
301 Laboratory Road
P.O. Box 2502
Oak Ridge, Tennessee 37830

phone: 865-481-8736
e-mail: beald@leidos.com