

## APPENDIX

## 1

# The G4GRID Procedure

---

<i>Introduction</i>	93
<i>Data Considerations</i>	94
<i>Terminology</i>	94
<i>Using the Graphical Interface</i>	94
<i>Procedure Syntax</i>	95
<i>The PROC G4GRID Statement</i>	95
<i>The GRID Statement</i>	97
<i>The BY Statement</i>	98
<i>Examples</i>	98
<i>Output Volume Grid Ranging from Minimum to Maximum</i>	98
<i>Output Volume Grid Ranging from 1 to 10</i>	99
<i>References</i>	99

---

## Introduction

The G4GRID procedure creates a SAS data set to be used with the SAS/SPECTRAVIEW three-dimensional visualization software. The procedure enhances SAS/SPECTRAVIEW software's ability to create two- and three-dimensional graphic images from data.

PROC G4GRID allows you to interpolate three-dimensional data contained in a SAS data set. While PROC G4GRID is not required to load a data set into SAS/SPECTRAVIEW, the procedure does provide advantages:

- It enables the loading of data sets that could not otherwise be loaded due to memory constraints. SAS/SPECTRAVIEW creates a three-dimensional volume grid with values for each unique x,y,z coordinate contained in the data set. For sparse data sets, this can result in data points, many of which may be missing. By using PROC G4GRID prior to loading the data set, you can replace filler points with interpolated values or resize the data set as required.
- It creates larger or smaller resolutions of a data set to enhance or facilitate your view of the data. Using PROC G4GRID, you can create a data set of almost any desired resolution for use in SAS/SPECTRAVIEW.
- It creates more meaningful volume data from randomly sampled data points.

PROC G4GRID works by interpolating randomly distributed points in three dimensions and by creating a volume grid with data points at each grid location. The procedure uses a Gaussian interpolation function that varies as the inverse of the square of the distance from surrounding data points. While PROC G4GRID is similar in syntax and function to the G3GRID procedure, G4GRID uses a different interpolation method and does not render the same values as G3GRID.

You can run the procedure using either of these methods:

- PROC G4GRID syntax
- graphical interface.

---

## Data Considerations

Each data set processed by the G4GRID procedure must contain at least four numeric variables:

- three axis variables
- one or more response variables.

If you specify more than one response variable, the procedure performs a separate analysis and produces interpolated values for each response variable. Unlike the G3GRID procedure, the G4GRID procedure does not require that the input data coordinate values be unique. It uses all coordinate values that fall within the grid volume.

---

## Terminology

The following terms are specific to the G4GRID procedure:

coordinates	the values representing the location of a data point along the X, Y, Z axes.
interpolate	to estimate values between two or more known values.
response	a value representing a measured response associated with each data point in the volume.
variables	the SAS data set variables whose values are plotted on the X, Y, and Z axes.

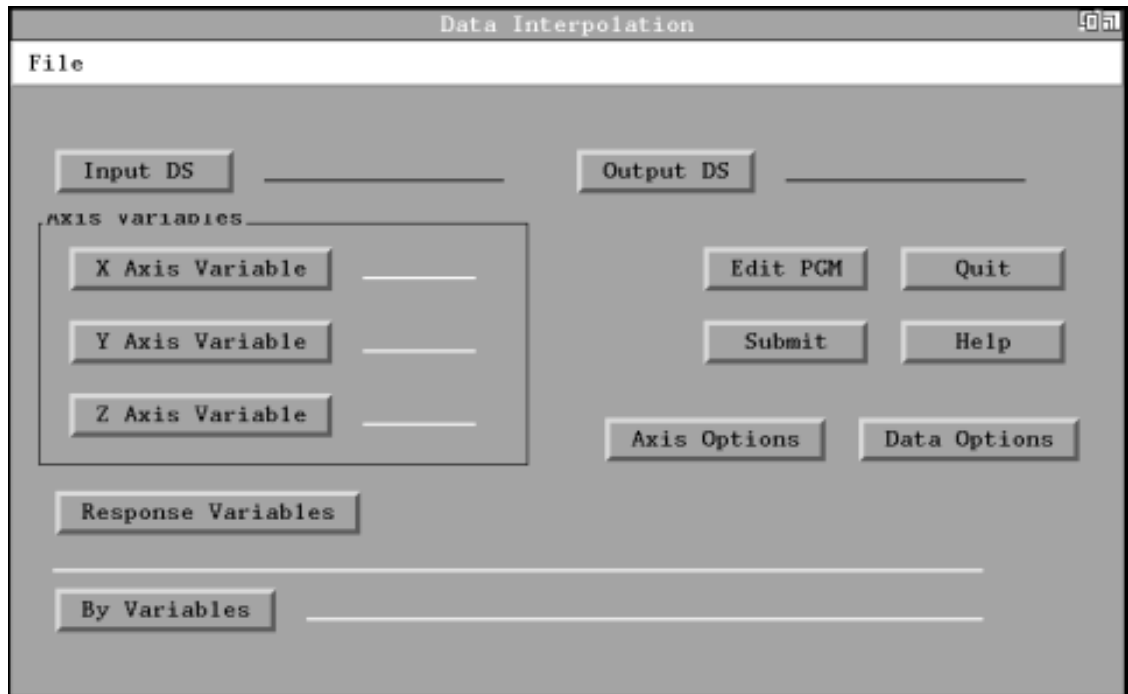
---

## Using the Graphical Interface

A FRAME entry application gives you a graphical interface for the G4GRID procedure so that you can process a G4GRID request without knowing the PROC G4GRID syntax. To use the G4GRID interface, issue the command:

```
af c=sashelp.g4grid.g4grid.frame
```

When the Data Interpolation window displays, see its online help for how to use it.




---

## Procedure Syntax

```

PROC G4GRID<DATA=SAS-dataset>
  <OUT=SAS-dataset>
  <GAMMA=gamma-factor>
  <D=amplitude-factor>
  <RADIUS=search-radius-value>;
GRID x*y*z=variable <...variable-n>< / <grid-options>>;
BY <options> variables;

```

“The PROC G4GRID Statement” on page 95 initiates the procedure.

“The GRID Statement” on page 97 specifies the variables for interpolation.

“The BY Statement” on page 98 specifies the variable or variables by which data is grouped for processing.

---

## The PROC G4GRID Statement

```

PROC G4GRID <DATA=SAS-dataset>
  <OUT=SAS-dataset>
  <GAMMA=gamma-factor>
  <D=amplitude-factor>
  <RADIUS=search-radius-value>;

```

The PROC G4GRID statement initiates the G4GRID procedure. If necessary, it can specify the input data set, the output data set, or both. This statement is required.

## Options

### **DATA=** *SAS-dataset*

specifies the input data set. By default, the procedure uses the most recently created data set, or you can use the DATA= option to specify a data set. If no data set has been created in the current SAS session and the DATA= option is not supplied, an error occurs and the procedure stops.

The input data set must contain at least four numeric variables, and it must contain the numeric variables specified in the GRID statement used with the PROC G4GRID statement.

### **OUT=** *SAS-dataset*

specifies the output data set. The data set will contain any BY variables specified, the values for the volume variables at the grid coordinates, and interpolated values for each response variable specified. By default, the output data set uses the *variable-n* naming convention as the data set name.

### **GAMMA=** *gamma-factor*

is the first of three parameters that control the interpolation of data points. The GAMMA= value controls the "locality" of the data interpolation.

The routine makes two passes through the data. In each pass, a Gaussian weight is applied to each observation based on its distance from the grid point in question. On the second pass, the interpolation function weighting is narrowed so that points nearer the grid point location contribute more than points farther away. The numeric value of *gamma-factor* controls the narrowness of this second Gaussian pass.

The valid range for the GAMMA= value lies between zero (non-inclusive) and one (inclusive). A smaller value means closer points contribute more heavily during the second pass than when a larger value is specified. A normal range for gamma is 0.3 to 0.5, with 0.3 being the default.

### **D=** *amplitude-factor*

is the second of three parameters that control the interpolation of data points. The D= value controls how the routine tries to resolve wavelengths embedded in the data.

The procedure tries to resolve some fraction of a target wavelength. The D= option indicates what fraction the procedure should attempt to resolve. Shorter wavelengths are resolved with greater amplitudes, while longer wavelengths are resolved with lesser amplitudes. The target wavelength that the procedure attempts to resolve is twice the average spacing of the input coordinates (that is, the smallest justifiable wavelength).

For example, assume that a data set has an average spacing of 3.5 units. The target wavelength for the data set is 7.0 units. The procedure attempts to resolve the D= fraction of the amplitude of this wave. If D=0.6 is specified, then the routine attempts to resolve 60 percent of amplitude of wavelengths in the data of 7.0 units. Longer waves have less than 60 percent of their amplitude resolved, while shorter waves have more than 60 percent resolved. The larger the value, the greater the detail resolved. However, too high a value results in spurious waves of smaller wavelengths being introduced into the data. That is, the larger the value specified, the more detailed the result, but also the more noise is introduced. The default value is 0.8.

### **RADIUS=** *search-radius-value*

is the third of three parameters that control the interpolation of data points. The RADIUS= value sets the limit of the search radius during interpolation.

Points outside the search radius do not contribute to the interpolated value of the grid point in question. Data points that lie farther from a grid point contribute less to the interpolated value of that grid point. Some points are far enough away that

their contribution can simply be ignored. Specifying a search radius that ignores them can significantly speed the data interpolation.

The valid range lies between zero (non-inclusive) and one (inclusive). For example, a search radius of 0.5 uses only data points that lie within half the grid volume distance from the grid location in question. To use all data points in the grid volume for each grid point, specify **missing**. The default value is 0.5.

*Note:* If an output grid contains missing values, increasing the search radius may eliminate them.  $\Delta$

---

## The GRID Statement

**GRID** *x\*y\*z=variable* < ...*variable-n*> </ <*grid-options*>>;

The GRID statement specifies the variables for interpolation. These include the volume variables to be used as coordinates for the output grid, as well as one or more response variables to be interpolated. You can specify multiple response variables for the three volume variables that you specify. Options for the GRID statement control the observations produced in the output data set. You can specify only one GRID statement.

### Options

Options affect all output produced by the GRID statement. If you use any of these options, separate them from the grid request with a slash (/). If you do not use any options, omit the slash.

**XAXIS=***list-of-coordinates*

**YAXIS=***list-of-coordinates*

**ZAXIS=***list-of-coordinates*

specifies a list of numeric values to be assigned to the respective axis. The list does not have to be in order; it will be arranged in ascending numeric order. The list should not contain duplicates. The list overrides the number of values defined by the NXAXIS=, NYAXIS=, or NZAXIS= options. The list of coordinates may be specified in the following ways:

```
n n ...n
n,n, ... ,n
n TO n <BY increment>
<n ... n> n,n ... ,n TO n <BY increment> <n, ... ,n>
```

**NXAXIS=***n*

**NYAXIS=***n*

**NZAXIS=***n*

specifies the number of coordinates along the respective axis. The default is 11. The minimum and maximum values along each axis are used as its starting and ending points. Each axis is divided into *n*-1 equally spaced sections. These options are ignored if the corresponding XAXIS=, YAXIS=, or ZAXIS= options are specified.

**XEXTRA=***n*

**YEXTRA=***n*

**ZEXTRA=***n*

specifies the number of temporary grid cells to be added to the interpolation along the respective axis. Normally, only points within the grid volume are included in the interpolation. However, this can result in some edge effect errors, since no data

outside this volume are used. To avoid such conditions, you can set the option to a positive non-zero value.

The distance of the new points from the end, which is the size of each "border," is the width of the first and last coordinate pair, respectively. All points that fall within the new coordinate boundaries are used in the interpolation. The size of the output grid is not affected.

#### **UNIFORM**

instructs the G4GRID procedure to assume that all coordinates along each axis are equally spaced. This allows the procedure to take some short cuts to slightly speed processing. The UNIFORM option is not necessary unless XAXIS=, YAXIS=, or ZAXIS= options are specified.

If XAXIS=, YAXIS=, or ZAXIS= is specified, PROC G4GRID must assume that the spacing between output grid points is not equal. As a result, special processing must occur for each correction pass.

If none of the above options are specified or the UNIFORM option is specified, the procedure assumes all points in the output grid are equally spaced and omits the special processing. If you specify one of the above options and the output grid points are known to be equally spaced, specifying the UNIFORM option may slightly speed processing. If, however, the output grid points are unevenly spaced to a significant extent, and the UNIFORM option is specified, erroneous results can occur.

UNIFORM is the default unless XAXIS=, YAXIS=, or ZAXIS= is specified.

---

## The BY Statement

**BY** *< options > variables;*

The BY statement specifies the variable or variables by which data is grouped for processing. A separate grid is produced for each value of the BY variable. The BY statement is optional.

### Options

#### **DESCENDING**

indicates that the input data set is sorted in descending order. This is the default.

#### **NOTSORTED**

indicates that observations with the same BY variable values are to be grouped as they are encountered, without regard for whether the values are in numerical order.

---

## Examples

The following examples illustrate various G4GRID invocations.

---

## Output Volume Grid Ranging from Minimum to Maximum

Assume that data set MYLIB.SAMPLE contains the variables MY\_X, MY\_Y, and MY\_Z. At each x,y,z coordinate location specified by these variables, there is a value for

the response variable RANDAT. Furthermore, there are multiple groups of these values specified by the BY variable BYIT.

To create a data set with an output volume grid that ranges from the minimum to the maximum along each coordinate for each BY group, you could use the following syntax:

```
proc g4grid data=mylib.sample;
  grid my_x*my_y*my_z=randat / nxaxis=5 nyaxis=5 nzaxis=5;
  by byit;
run;
```

---

## Output Volume Grid Ranging from 1 to 10

Using the same data set, assume there is also a variable RANDAT2 for which you would like output volume grids. This time, you would like the grid to run from 1 to 10 with values at each even unit distance. You could use the following syntax:

```
proc g4grid data=mylib.sample;
  grid my_x*my_y*my_z=randat randat2 /
  xaxis=1 to 10 yaxis=1 to 10 zaxis=1 to 10;
  by byit;
run;
```

---

## References

Barnes, Stanley L., "Mesoscale Objective Map Analysis Using Weighted Time-Series Observations," United States Oceanic and Atmospheric Administration, Environmental Research Labs, Norman, OK, *Technical Memorandum (NOAA TMM ERL NSSL-62)*, March 1973.





The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/SPECTRAVIEW® Software User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

**SAS/SPECTRAVIEW® Software User's Guide, Version 8**

Copyright © 1999 by SAS Institute Inc., Cary, NC, USA.

ISBN 1-58025-462-4

All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, by any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute, Inc.

**U.S. Government Restricted Rights Notice.** Use, duplication, or disclosure of the software by the government is subject to restrictions as set forth in FAR 52.227-19 Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

The Institute is a private company devoted to the support and further development of its software and related services.