GMS TUTORIALS

Geostatistics – 2D

Two-dimensional geostatistics (interpolation) can be performed in GMS using the *2D Scatter Point* module. The module is used to interpolate from sets of 2D scatter points to any of the other object types (meshes, grids, TINs). Several interpolation schemes are supported, including kriging. Geostatistics are useful for setting up input data for analysis codes or for site characterization.

The tools for manipulating scatter point sets and the interpolation schemes supported in GMS are described in this tutorial. The interpolation schemes presented in this tutorial will be easier to understand if you have read the *Interpolation* section of the *GMS Online Help*. This tutorial should be completed before attempting the *3D Geostatistics* tutorial.

1.1 Outline

This is what you will do:

1. Create a scatter point set from scratch.
2. Import scatter points from a file.
3. Create a bounding grid.
4. Interpolate the scatter points using various interpolation methods.
5. Use the data calculator to compare interpolations.

1.2 Required Modules/Interfaces

You will need the following components enabled to complete this tutorial:

- Grid

- Geostatistics

You can see if these components are enabled by selecting the *File | Register*. If you do not have these components enabled, you can complete the tutorial in *Demo Mode*. You can switch to *Demo Mode* by selecting the *File | Demo Mode* menu command.

2 Getting Started

Let's get started.

1. If necessary, launch GMS. If GMS is already running, select the *File | New* command to ensure that the program settings are restored to their default state.

3 Creating a Scatter Point Set

Interpolation in GMS is performed using scatter points. A set of 2D scatter points is defined by a set of xy coordinates. A group of scatter points is called a scatter point set. Each scatter point set may also have a data set associated with it. Data sets are scalar or vector values – one value at each scatter point. The scatter points represent the geometric objects while the data sets represent the data values at the points. The data sets may represent anything – elevation, heat measurements, concentration of a chemical etc. The data sets can be interpolated to a TIN, mesh, or grid.

2D scatter point sets can be created inside GMS using the *Create Scatter Points* tool.

1. In the *Project Explorer* right-click on the empty space and then, from the pop-up menu, select the *New | 2D Scatter Set* menu command.
2. Select the *Create Scatter Points* tool.
3. Click on the screen a few times in different places.

You are creating scatter points. When you create each point, GMS automatically assigns a data set value at the point. You can change it so that GMS prompts you for the data set values.

4. Select the *Scatter Points | Scatter Point Settings* menu command.
5. Turn on the *Confirm data set value* option and click *OK*.
6. Click on the screen.

Now you are prompted to enter a data set value every time you create a new point.

7. Click *OK* to exit the *Scatter Point Data Set Value* dialog.

You can only enter steady state data this way. To create a scatter point set with transient data, you would need to import it from a file. That's what we'll do next.

4 Importing a Scatter Point Set

Scatter point sets can be imported from an existing file using the *Import Wizard*. The *Import Wizard* allows you to import data into GMS from text files that are in columnar format. The file we will import was generated as an *Excel* spreadsheet and exported from *Excel* as tab delimited text. The file contains scatter points which represent locations where the concentration of a contaminant has been estimated using a soil gas survey. Our goal is to generate a map of the contaminant plume.

4.1 Open the File

1. Select the *New* button and select *No* when prompted to save changes.
2. Select the *Open* button.
3. At the bottom of the *Open* dialog, change the filter to **Text Files (*.txt,*.csv)**.
4. Locate and open the directory entitled **tutfiles\Geostatistics\geos2d**.
5. Select the file named **plumedat.txt**.
6. Click on the *Open* button.

The *Import Wizard* should appear. The first step of the *Import Wizard* allows you to specify how the data is delineated and where in the file the data begins. For this file, the first row contains column headings. The first column contains the point labels, the second column is the X values of the points, the third column is the Y values, and the fourth column is the data set values for the points.

4.2 Import Wizard - Step 1

1. Change the dialog settings to match those shown in Figure 1 below.

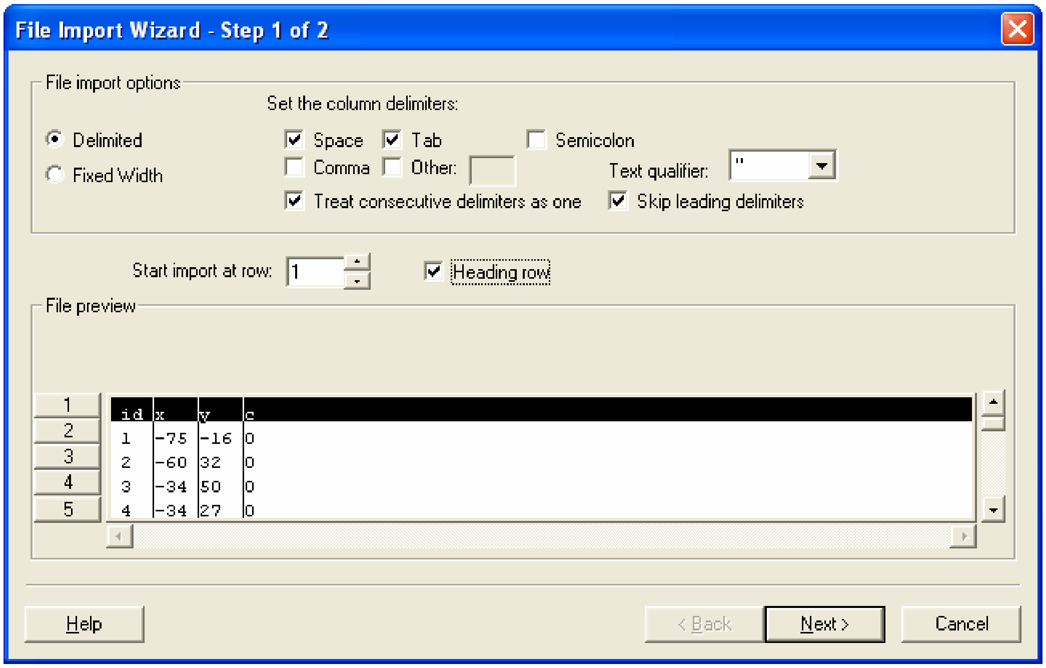

Figure 1 File Import Wizard, Step 1.

2. Select the *Next* button.

4.3 Import Wizard - Step 2

Step 2 of the *Import Wizard* allows you to specify what the data in the file means.

1. Select **2D Scatter Points** as the *GMS data type*.

The data associated with 2D scatter points is now displayed in the *Mapping options* section. Now we tell GMS what each column in the file means. We do this by selecting the correct data type in the *Type* row of the spreadsheet for each column in the file. The *Type* row is the first row in the spreadsheet. The options in the combo box changes depending on the GMS data type selected in the top of the dialog. Since we specified we were using a heading row in the first step, GMS looked at the headings and automatically found and mapped the X and Y columns by selecting the X and Y selection in the *Type* row of the spreadsheet.

2. Locate the *Type* row, the first row in the spreadsheet. The first column is ID's for each point. In the *Type* row for the first column select the **Label** option in the pull-down menu.

3. For the column with heading **c**, make sure the type is set to **Data set**.

The data set in this file represents concentrations of a contaminant. The *No data* option can be used to specify a key value in the file used to indicate a lack of information. For example, if no measurement was taken at one of the points, we might enter –999 or some other key value for the concentration and specify –999 as the *No data* value in the *Import Wizard*. GMS will then know to ignore these points when we do interpolation later. The key value should be a value that would not normally be encountered in the data set. In this case, we don't need to use this option.

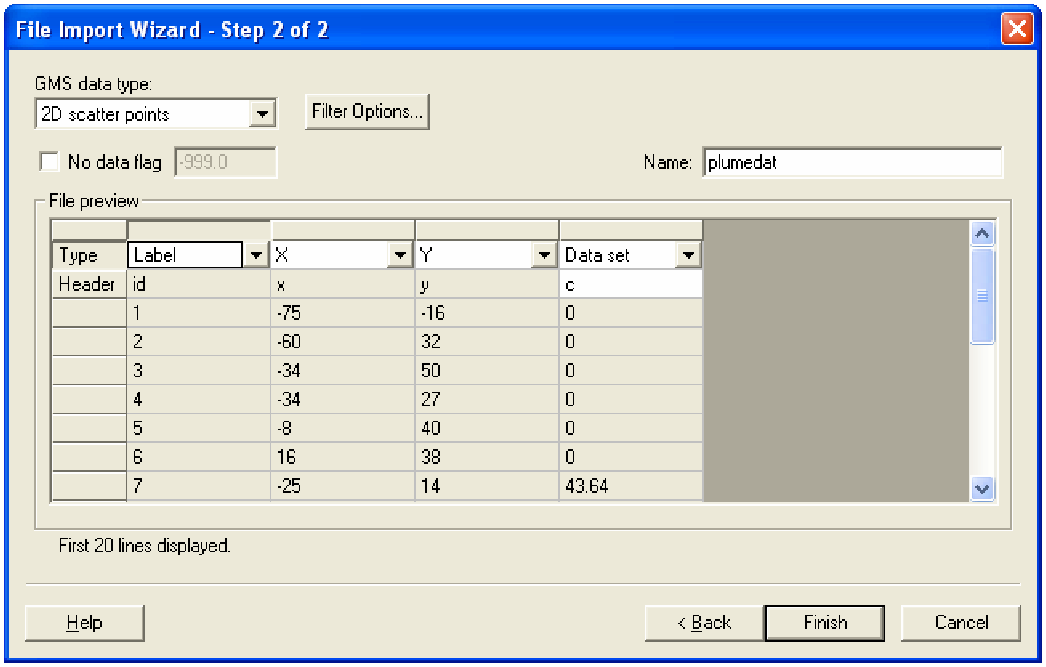

Figure 2 File Import Wizard, Step 2.

4. Make sure the dialog looks like Figure 2 and click the *Finish* button.

A set of points should appear on the screen.

5 Changing the Display Options

You can change the appearance of the scatter points using the *Display Options* dialog:

1. Select the *Display Options* button .
2. Select the *2D Scatter Data* item  from the list on the left.
3. For the *Scatter point symbols*, change the *Color* option to **Data**.
4. Select the button to the right of the *Scatter point symbols* toggle.
5. Choose one of the triangle shaped symbols.
6. Select the *OK* button to exit the *Symbol Attributes* dialog.
7. Select the *OK* button to exit the *Display Options* dialog.

Each of the points should now be displayed with a colored triangle. The color of the symbol represents the relative concentration of the contaminant at the point. When displaying colored symbols, it is useful to also display a color legend.

8. Select the *Data | Color Ramp Options* menu command.
9. Turn on the *Legend* option.
10. Select the *OK* button.

Notice that the concentrations vary from 0.0 to about 100.0.

6 Creating a Bounding Grid

The goal of this tutorial is to generate a series of contour plots illustrating the plume. To do this we will first create a grid that bounds the scatter point set and then we will interpolate the concentrations from the scatter points to the grid nodes. The grid will then be contoured.

1. In the *Project Explorer*, right-click on the **plumedat** scatter point set and select the *Bounding 2D Grid* menu command.

Notice that the x and y dimensions of the grid are already defined. The default values shown in the dialog cause the grid to extend beyond the scatter points by 10% on each side.

2. In the *X-Dimension* section, enter **60** for the *Number of cells*.
3. In the *Y-Dimension* section, enter **40** for the *Number of cells*.
4. Select the *OK* button.
5. Select the *Frame* button.

A grid should appear on the screen that just encompasses the scatter point set.

7 Selecting an Interpolation Scheme

The next step is to select an interpolation scheme. Several interpolation schemes are supported in GMS because there is no one interpolation scheme that is superior in all situations. Typically, the best approach is to try several schemes and then determine which scheme is giving the most reasonable results. GMS has been structured in such a way that several different schemes can be tested quickly and easily.

8 Linear Interpolation

First we will try simple linear interpolation.

1. Select the *Interpolation* | *Interpolation Options* menu command.
2. Select the *Linear* option.
3. Select the *OK* button.

To interpolate to the grid:

4. Select the *Interpolation* | *Interpolate → 2D Grid* menu command.
5. Click OK.

9 Viewing the Results

A set of contours should now be displayed. The concentrations have been interpolated to the grid.

1. Select the *Oblique View* button.

Notice that the grid has been deformed to match the contours. Color shading the entire surface can provide an even more effective display of the grid.

2. Select the *Display Options* button.
3. Select the *2D Grid Data* item from the list on the left.
4. Make sure the *Contours* option is checked and select the *Options* button to the right of the *Contours* option.
5. For the *Contour method*, select **Color Fill**.
6. Turn **on** the *Smooth* option in the *Fill options* section.
7. Select the *OK* button twice to exit both dialogs.

Notice that the outer part of the grid still has a concentration value equal to zero. When linear interpolation is performed, the scatter points are triangulated to form a temporary TIN. A plane equation is computed for each triangle in the TIN and the coefficients of the plane equation are used to interpolate to points inside the triangle. Therefore, linear interpolation cannot be performed for grid nodes outside the convex hull of the TIN (the boundary of the TIN). As a result, these nodes are assigned a value of zero. However, for this application, a value of zero is appropriate since the concentrations of the scatter points on the perimeter of the scatter point set are zero.

10 Viewing the Elliptical Control Function

The scatter points were generated from an elliptical control function. It's interesting to see how the different interpolation schemes compare to the original control function. To view the original control function:

1. Right-click on the *2D Grid Data* folder in the *Project Explorer* and select the *Import Data Set…* option.
2. Select the file **tutfiles\Geostatistics\geos2d\ellipse2g.dat** and click *Open*.
3. If necessary, select the **control_function** data set in the *Project Explorer* to make it active.

11 Simple IDW Interpolation

The next scheme we will try is a simple form of inverse distance weighted (IDW) interpolation.

1. Click on the *2D Scatter Data* folder in the *Project Explorer*.
2. Select the *Interpolation* | *Interpolation Options* menu command.
3. Select the *Inverse distance weighted* option.
4. Select the *Options* button to the right of the *Inverse distance weighted* option.
5. In the *Nodal function* section at the top of the dialog, select the *Constant* (Shepard's method) option.
6. In the section entitled *Computation of interpolation weights*, select the *Use all points* option.
7. Select the *OK* button to exit the *2D IDW Interpolation Options* dialog.
8. Select the *OK* button to exit the *2D Interpolation Options* dialog.

To interpolate to the grid:

9. Select the *Interpolation* | *Interpolate* → *2D Grid* menu command.
10. Select the *OK* button.

The *IDW* scheme is a simple moving weighted averages scheme. To interpolate a value at a point, a weighted average of the nearby scatter points is used. The weights are an inverse function of distance. The closer a scatter point is to the interpolation point, the greater the weight given to the scatter point.

12 IDW Interpolation With Gradient Planes

One of the problems with simple IDW interpolation is that the interpolated data set always tends toward the mean of the data set in the voids between scatter points. As a result, local minima or maxima in the voids in the scatter point set are not properly inferred. To overcome this problem, a "nodal function" can be computed at each scatter point. A nodal function is a plane or quadratic function that is forced to pass through the scatter point and approximate the nearby scatter points in a least squares sense. When the interpolation is performed, rather than computing an average of the data set values at the scatter point locations, an average is computed of the nodal functions of the nearby scatter point evaluated at the interpolation point. This approach allows local trends to be inferred and often results in a more accurate interpolation.

The next scheme we will try is IDW interpolation with planar nodal functions.

1. Select the *Interpolation* | *Interpolation Options* menu command.
2. Select the *Options* button to the right of the *Inverse distance weighted* option.
3. In the *Nodal function* section at the top of the dialog, select the *Gradient plane* option.
4. Select the *OK* button to exit the *2D IDW Interpolation Options* dialog.
5. Select the *OK* button to exit the *2D Interpolation Options* dialog.

To interpolate to the grid:

6. Select the *Interpolation* | *Interpolate →2D Grid* menu command.
7. Select the *OK* button.

13 Using the Horizontal Anisotropy Option

The interpolation can be “stretched” in a horizontal direction by using the horizontal anisotropy option in conjunction with a specified azimuth. This option might be used if, for example, you know a contaminant plume is elongated in a particular direction due to groundwater flow.

To change the horizontal anisotropy:

1. Select the *Plan View* button .
2. Select the *Interpolation* | *Interpolation Options* menu command.
3. Change the *Horizontal anisotropy* value to **2.0**.
4. Change the *Azimuth* value to **90.0**.
5. Select the *OK* button.

6. Select the *Interpolation* | *Interpolate →2D Grid* menu command.

7. Enter **c_idw_grad_h2** for the new data set name.

8. Select the *OK* button.

As can be seen, the data has been stretched in the horizontal direction. You might want to experiment with different values for the horizontal anisotropy.

14 IDW Interpolation With Quadratic Nodal Functions

The nodal functions used in IDW interpolation can also be quadratic functions which are constrained to pass through the scatter point and approximate the neighboring scatter points in a least squares fashion. The averaging or blending of the quadratic functions during the interpolation stage often results in a very smooth surface.

1. Select the *Oblique View* button .

2. Select the *Interpolation* | *Interpolation Options* menu command.

3. Select the *Options* button to the right of the *Inverse distance weighted* option.

4. In the *Nodal function* section at the top of the dialog, select the *Quadratic* option.

5. Select the *OK* button to exit the *2D IDW Interpolation Options* dialog.

6. Select the *OK* button to exit the *2D Interpolation Options* dialog.

To interpolate to the grid:

7. Select the *Interpolation* | *Interpolate →2D Grid* menu command.

8. Select the *OK* button.

15 Truncation

Notice that the minimum value listed in the color legend is a negative number. Of course, this is impossible since there is no such thing as a negative concentration. By inferring trends, the nodal functions can sometimes project the plume values beyond zero and into the negative range. This type of error can be easily fixed using truncation.

1. Select the *Interpolation* | *Interpolation Options* menu command.

2. Turn on the *Truncate values* option.

3. Select the *Truncate to specified range* option.

4. Enter **0.0** for the *Min* value and enter **150.0** for the *Max* value.

We don't want the concentrations to go below zero but we will allow the interpolation scheme to infer a maximum concentration greater than the maximum measured value.

5. Select the *OK* button.

6. Select the *Interpolation* | *Interpolate →2D Grid* menu command.

7. Enter **c_idw_quad_trunc** for the name of the new data set.

8. Select the *OK* button.

Notice that the concentrations are now mostly zero around the perimeter of the map.

16 Kriging

The last interpolation scheme we will test is kriging. Kriging is based on the assumption that points that are near each other have a certain degree of spatial correlation, but points that are widely separated are statistically independent. Kriging is a set of linear regression routines that minimize estimation variance from a predefined covariance model.

1. Select the *Interpolation* | *Interpolation Options* menu command.

2. Select the *Kriging* option.

3. Select the *Options* button to the right of the *Kriging* option.

16.1 Creating the Experimental Variogram

There are a large number of options to be specified in the *Kriging Options* dialog. Fortunately, the defaults shown are adequate in most cases. However, a variogram must always be defined.

1. Select the *Edit Variograms* button to bring up the *Variogram Editor*.

2. Select the *New* button in the section entitled *Experimental variogram.*

3. Select the *OK* button to accept the defaults.

A curve should appear in the upper window of the *Variogram Editor*. This curve is called an experimental variogram. The experimental variogram is found by calculating the variance in data set values of each scatter point in the set with respect to each of the other points and plotting the variances versus distance between the points. As can be seen in the plot of the experimental variogram, the shape of the variogram indicates that at small separation distances, the variance is small. In other words, points that are close together have similar data values. With many data sets, after a certain level of separation, the variance in the data values becomes somewhat random and the variogram oscillates about a value corresponding to the average variance. However, with concentration data, many of the points have zero values and this tends to pull the experimental variogram back down.

16.2 Creating the Model Variogram

Once the experimental variogram is computed, the next step is to define a model variogram. A model variogram is a simple mathematical function that models the trend in the experimental variogram. The model variogram is used in the kriging computations.

1. In the section entitled *Nested structure*, select the *New* button.
2. For the *Model function* select the **Gaussian** option.
3. Enter a value of **25.0** for the *Nugget*, **1965.0** for the *Contribution*, **63.0** for the *Range*.

At this point there should be a reasonable fit between the model and the first part of the experimental variogram. The second part is difficult to fit in this case because of the zero values described above.

4. Select the *OK* button to exit the *Variogram Editor*.
5. Select the *OK* button to exit the *Kriging Options* dialog.
6. Select the *OK* button to exit the *2D Interpolation Options* dialog.

16.3 Interpolating to the Grid

To interpolate to the grid:

1. Select the *Interpolation* | *Interpolate → 2D Grid* menu command.
2. Select the *OK* button.

Notice that this interpolation scheme results in an upward curvature towards the outside edges of the grid. If you wanted to correct this, you could add more scatter points with a concentration of 0.0 in these outlying areas.

17 Switching Data Sets

Now that we have interpolated to the grid using several different interpolation schemes, we may wish to review the results by replotting some of the interpolated data sets. We can switch back to one of the previous data sets using the *Project Explorer*.

1. Select the *Plan View* button .

2. In the *Project Explorer*, select the *c_linear* data set under the *2D Grid Data* folder.

3. Use the up and down keys to switch between the datasets and watch the contours change.

18 Using the Data Calculator

Occasionally, it is useful to use the *Data Calculator* to compare two data sets generated by interpolation. As an example, we will use the *Data Calculator* to compute the difference between the kriging and natural neighbor data sets.

1. Select the *Data* | *Data Calculator* menu command.

The currently available data sets are listed in the top of the dialog. Each data set is assigned a letter. Data sets are referenced in the mathematical expression using the letters. The "c_krig" data set should be labeled "j" and the "idw_quad" data set should be labeled "g".

The next step is to enter an expression to compute the absolute value of the difference between the krig and nn data sets.

2. In the *Expression* field, enter **abs(g-j)**.

3. In the *Result* field, enter **Difference**.

4. Select the *Compute* button.

Now that we have computed the difference between two data sets it is helpful to view some basic statistics related to the new data set.

5. Select the *Data Set Info* button.

The resulting dialog displays basic statistics related to the active data set such as minimum, maximum, and mean data values.

6. Select the *OK* button to exit the *Data Set Info* dialog.

7. Select the *Done* button to exit the *Data Calculator* dialog.

8. Select the *Frame* button.

The contour plot now displayed represents the data set we just computed. Any new data set computed using the *Data Calculator* is automatically designated the active data set.

19 Conclusion

This concludes the tutorial. Here are some of the key concepts in this tutorial:

- You can interpolate from a sparse set of points to a different set of points (usually more dense and defining a surface) by using 2D scatter points.
- 2D scatter points can be created by hand or imported from a file.
- There are several interpolation algorithms available in GMS.
- Linear algorithms do not interpolate beyond the convex hull of the scatter points.
- Anisotropy and truncation can be used to help control the interpolation process.