Surfer offers many different grid based maps, such as 2D contour maps and 3D surface maps. To create a grid based map, you must have a grid file. To create a grid file, Surfer takes randomly spaced XYZ data and uses it to create a regularly spaced grid file, composed of grid nodes. Each grid node is located at a particular XY location and has a Z value associated with it. Although the algorithms are computed internally in Surfer, choosing the best gridding method for the data can be difficult. Surfer has several different options for gridding methods and each have its own set of gridding options. This newsletter article is going to focus on Data Metrics. More information about Kriging, Nearest Neighbor, Natural Neighbor, Local Polynomial, Radial Basis Function, and Triangulation with Linear Interpolation can be found in this newsletter article; the rest of the gridding methods will be covered in a future newsletter article.

The Data Metrics gridding method can be used to map statistical parameters of a data set. Although Data Metrics is a gridding method, it is different than most because it does not interpolate the data to obtain a Z value. Data Metrics is used to gain information about data points in the form a grid. It is recommended to use the same settings (i.e. grid line geometry, search, breakline, and fault parameters) as you would when gridding with another method.

There are five major groups of the Data Metrics gridding method: Z order Statistics, Z Moment Statistics, Other Z Statistics, Data Location Statistics, and Terrain Statistics. Each group is categorized based on the detail it is analyzing of the data. For each statistical parameter, the data is divided into search sections where the calculation will be performed. The value returned for the calculation will then be assigned to the grid node in that region. Data metrics allows you to define the size and specific nodes for each search parameter. This method is useful when determining specific information about the data that was calculated and created maps from those calculations.

**Z Order Statistics**

The Z Order Statistics provide specific statistical information about the data that is specified within the search radius. The Z grid node values will have the same units as the original data file. These data values can be significant to calculate if the goal is to demonstrate areas of statistical interest. The data that is found in the search radius will be sorted from least to greatest and then the following statistics can be calculated.
The image above shows a sorted set of data and the appropriate data metrics that are calculated based on the data.

- **Minimum**: The Minimum is calculated by returning the first result of the sorted data in the search parameters. An application of this gridding method would be to show the areas of lowest elevation in comparison to surrounding areas to represent the areas with the highest rate of erosion.

- **Median**: The median is the middle value in the data set. The median value can be useful if the data is skewed to one direction, thus leaving the average to be skewed as well. The median is a good measure of the middle value in a set of data, such as a grid of the middle points in the Rocky Mountains. Using the median will allow for the outliers not to be taken into account so the grid would be an accurate representation of the middle peaks.

- **Maximum**: The Maximum value is determined by returning the last value in the sorted data. This statistical representation can be useful when creating a canopy map of the forest that represents the greatest heights of the trees.
The Kakum National Park Tree Canopy grid, shown above as a 3D Surface, represents the tallest trees in the Kakum National Park, Ghana. The data used in the image is fictional and used to demonstrate an application of the Maximum statistical value.

- **Range**: The range of the data is taken by subtracting the minimum value from the maximum value. This value can be useful to determine the difference between the values that are located in each search section. An application of this method includes creating a grid file to represent the elevations used when creating a contour plot to verify the accuracy.

- **MidRange**: The midrange value is the half the value of the sum of the maximum and the minimum. The value can be more applicable than the median, since it takes into account the data range instead of the location of the middle number. The midrange value represents the true middle of the data range.
The model above displays a full data set, at the top, and then it split into the upper and lower half based on the median value, 8. The median is then taken of the upper and lower half to return the upper and lower quartile values, respectively.

- **Lower Quartile**: The data is divided into half, and the median of the lower half is returned. The value returned will represent the lower 25% of the data. The lower quartile can provide insight on which the range of the lower outliers exists.

- **Upper Quartile**: The data is divided into half, and the median of the upper half is returned. The value returned will represent the upper 25% of the data. The upper quartile can provide insight on which the range of the upper outliers exists.

- **Interquartile Range**: The interquartile range is the lower quartile subtracted from the upper quartile. The difference can be used to show the spatial variability. Since the data focus mostly on the center data, it does not take into account the distributions at the end of the data. The interquartile range can provide insight as to where most of the data lies since the outliers have been removed.

### Z Moment Statistics

The Z moment statistics use the data that is specified in the search radius to perform specific statistical calculations about the values in the node. The statistic values show the variability in the data set and how the values relate to one another. The grid node value assigned at the grid is in the same data units as the original Z values.

- **Mean**: The mean value is determined by summing up the data and then dividing by the number of points in the search section. The mean value is useful to determine the average value calculate in a specific area.
The contour map created above is displaying the mean population for the world based on the mean population of the major surrounding cities.

- **Standard Deviation**: This value is calculated by summing all the Z values in the search area to compute the variance, and then the square root is taken of the variance to determine the standard deviation. The standard deviation value identifies the variability of each data value from the mean. This is useful to summarize continuous data. Since this method returns values based on the mean, it is not recommended to use this method if the data set contains a high number of outliers.

- **Variance**: This value is the square standard deviation of the data defined in the search parameters. The variance value can depict the variability of the data. The closer the variance is to the zero, the closer the data is located near the mean. When the variance value is larger, then the data is scattered. Using this method can be useful in determining where the data is the most spread out to reveal inconsistencies in the data or identify areas of concern.

- **Coefficient of Variation**: The coefficient of variation is the ratio between the standard deviation and the mean. This calculation is used when interested in the variability of the data compared to the observation size.

**Other Z Statistics**

The statistical methods listed under the **Other Z Statistics** are used to analyze the data without requiring the data to be sorted. The values that are calculated for each node will have the same units as the Z value in the data set.

- **Sum**: This value will add up all the Z values that are included in the search region. This calculation can be useful in determining the total values for a given region. For example, this can represent regions with high contamination values as shown in the plot below.
Gold Concentration

The map above was created using a shaded relief map overlaid with a contour plot. The grid file was generated using the Sum statistical count to represent the areas with the highest concentration of gold.

- **M.A.D:** The median absolute deviation value is calculated by determining the median in the search range and then deviations are taken for each value. Once the deviation values are determined, then the median is taken from that set and assigned as the Z value for that grid node. This method is more resilient to outliers and ideal for data sets that do not contain a median or a variance.

- **R.M.S.:** The root mean square value is calculated by taking the square root of the average of all the values squared. This means, using the data set \(x_1, x_2, x_3, x_4, x_5\):
  1. The data values are squared and summed:
     \[
     x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2
     \]
  2. The average is taken of the value from Step 1:
     \[
     \frac{(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2)}{5}
     \]
  3. The square root of the value in Step 2 is calculated to obtain the RMS value:
     \[
     RMS = \sqrt{\frac{(x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2)}{5}}
     \]

This method is very useful in the electrical industry when determining the effective power value.

**Data Location Statistics**
The statistics for Data Locations are concerned with the location of the data points, unlike the methods mentioned above. The location of data points is often useful when determining the density or the distance from each other. The values calculated in the statistics are in the same units as the original data set. Since the values are not concerned about the Z value at each location, the data metrics are calculated based on the XY data points.

For example, a use of the Count Data Metrics can be represented in the map below. The grid represents the number of health care facilities per area in Kenya. This map can then be used to represent a link between the number of health care facilities and the rate of illness.
The map above was created using the Count statistics to represent the number of health care facilities in the entire region of Kenya, Africa. The color scale, on the right, shows the colors that correspond to the specific Z values.

- **Count**: The count value is determined by simply calculating the number of points inside the specified search area. The count of the data points in a particular area can be useful in generating a probability map to demonstrate the likeliness of an even occurring in a given location. This is often useful when determining areas of high risk or repetitive events.

- **Approximate Density**: Data density within the search. This can be helpful if you want to determine the number of points that are within a certain area, also known as a density map.

- **Distance to Nearest**: The distance returned here is the distance between the grid node and the nearest data point. This can be useful in determining which areas are more clustered in the data set.

- **Distance to Farthest**: The distance calculated is the distance between the grid node and the furthest data point in the search range. This can be useful in determining how spatial the data is and how far the data points being calculated are.

- **Median Distance**: Similar to median calculation, this value returns the median distance based on the distance values for each point in the search parameter. Using the method can return a more accurate representation of the data because it does not give much weight to outliers in the data.

- **Average Distance**: This is the average distance value from the grid node to the data point. This can be useful in determining the average distance between the data and the grid node.

- **Offset Distance**: The first thing calculated is the center location of the data points. Then the distance between that centroid and the grid node are calculated and returned as the Z value for this offset distance. This method will provide insight if the data is located near the grid node or further away.
**Terrain Statistics**

Similar to the Grid | Calculus commands, the Terrain Statistics will provide specific information on the slope and aspect of the grid. Using Data Metrics will allow the data to be sampled in sections instead of as an entire grid, as it is done with the Grid | Calculus command. This will allow for a more accurate representation of the slope and aspect since the search parameters allow a single section to be focused on at a time. For example, both of the Terrain Slope and Terrain Aspect values can be used in determining the strike and dip values for the data set.

- **Terrain Slope**: The value returned for the slope measures the degree of inclination relative to the horizontal plane. This method is useful to calculate the slope in each search parameter and represent areas of a constant slope in a contour map. The value returned is in degrees. A possible error is this gridding method is obtaining a horizontal plane. The result of this is because the search parameters are not defined correctly for the data set. It is recommended to reduce the search radii to search a more specific area.

- **Terrain Aspect**: The value calculated for each search parameter is the dip of the area, or the direction which the download slope is facing. The terrain aspect value is returned in degrees and is meaningless if the slope is zero because the area is flat.

![Terrain Aspect of Aspen, Colorado](image)

*The contour map above displays a grid file created using the Terrain Aspect measurement to represent Aspen, Colorado. The contour plot was exported as a KML file and displayed in Google Earth.*

**Advanced Options**
To specify the advanced options of the Data Metrics gridding method, in the Grid Data dialog, click Advanced Options.

Click the Advanced Options button to open the Data Metrics Advanced Options dialog.

The General page, of the Data Metrics Advanced Options displays the statistical options to use when gridding the data.

The General tab contains the different statistical options to create the desired grid file.

The Search page allows you to specify the search options of which points are considered when interpolating grid nodes.
The **Search** tab contains setting to modify the search parameters of the selected gridding method.

The **Search Ellipse** specifies the size of the local neighborhood in which to look for data, this can be impacted by the search rules set. The values of radius 1 and radius 2 are the distance in positive data units. The radii values are the lengths that are searched in the direction indicated by the angle. This means if trying to search an area of 5, then the search radii values should be half, $5/2 = 2.5$. This value identifies to search in a circular area around the grid node with a diameter of 5m.

The **Search Angle** represents the orientation between the positive x axis and the radius 1 of the ellipse axis. If there are no points found in the search ellipse then that grid node will be assigned the blanking value. By default, the search ellipse is circular giving equal weight to both directions surrounding the grid node. The default length is half the diagonal distance of the data.