Determining a Useful Interpolation Method for Surficial Sediments in the Gulf of Maine
Ian Cochran

ABSTRACT

This study was conducted to determine if an interpolation of surficial sediments in the Gulf of Maine could be made which could be used to predict mean grain size within the study area. Data from the U.S. Geological Society along with Inverse Distance Weighting and Kriging processes was used. It was determined that a Kriging interpolation that does not take anisotropy into consideration provides the best predictive surface.

BACKGROUND

The study area for this project is illustrated in figure 1. The region was chosen because of the proximity to Colby College. Data for this region was available from the U.S. Geologic Survey Open-File Report 03-001. The region also has some interesting geologic history that causes the Gulf of Maine to be a good place to try and create a predictive surface.

Fig. 1: The area of study for this project is the Gulf of Maine, George’s Bank and New England shelf regions shown above.
DATA

The hathaway71 shape file dataset consisted of 1252 data points collected between 1955 and 1970 in the Gulf of Maine, Georges Bank and portions of the New England shelf. This dataset was chosen from the numerous data sets in the U.S. Geological Survey Open-File Report 03-001 because the information about the sediments collected at each of the sample sites was adequately extensive to eliminate the need for preprocessing for this project. The dataset contained latitude, longitude, depth if available, gathering method if available, year of sampling, lithology, percent sand, percent clay, percent silt, percent gravel, mean phi size, median phi size, skewness, kurtosis, and percent distribution for phi sizes from –11 to 21. Other data sets were more limited in their geologic distribution or would have required preprocessing to turn a lithologic description into a useful ratio of component types using appropriate tables. The mean phi size was chosen to be the category on which the interpolations were done because it is a fairly useful, simple characterization of the sample.

Within the area of study, the data was not evenly distributed. There was a higher concentration of points inshore, especially in the western Gulf of Maine north of Cape Cod. There was a grid-like distribution in the rest of the study area that resulted in reasonable data distribution and enough spread across the area to allow for minimal error in the interpolations.

Figure 2: A map showing the distribution of data points from the hathaway71 dataset.
DATA PROCESSING

Minimal preprocessing of the data was necessary. The data was available as shapefiles and so there was no conversion or alteration needed. The only preprocessing that was necessary was the conversion of all of the various data sets to a common geographic coordinate system. This was done so that the data could be projected on the fly in ArcMap. Changing all of the data to a projected coordinate system prior to loading into ArcMap was not successful for unknown reasons. The conversion of the coordinate systems to one geographic coordinate system was done by using either the “Define Projection Wizard” in ArcToolbox, or by altering the shapefile properties in the Properties window through ArcCatalog.

Due to the large number of data points contained in the data set, it was necessary to change the “Maximum # of inputs” to 1500 in the AdvancedArcMapSettings program. Changing the maximum number of included points allows for the full data set to be used in the processing that will ensue. Without changing this parameter, some of the data points would be left out and the final product would suffer.

The first analysis of the data was done by changing the symbology of the hathaway71 data so that it was displayed by categories (Figure 3).

![Figure 3: A map showing the hathaway71 data categorized for mean phi size.](image)
The category used was the mean phi size and each phi size was given an individual color. This altering of the symbology allowed for visual identification of any trends that may exist as well as getting a feel for the distribution of the data. The initial representation of the data showed a general trend of coarser sediments forming a band that runs roughly from southwest to northeast. This representation of the data provides clues to the general trends, but is not useful in the final project as it is not a predictive tool.

The first attempt at creating a surface that had predictive value was an Inverse Distance Weighted (IDW) interpolation. An IDW interpolation was chosen because the process is as simple an interpolation as possible. The IDW interpolation simply uses the distance from data points and the value of these data points to determine a probable value at an unknown location. For the IDW done for this project, the default parameters were used. A neighborhood search pattern with 15 desired neighbors and a minimum of 10 was utilized. A circular search pattern was used because anisotropy in the data was not a concern. A power of two was used. The power is related to how quickly the influence of a point decreases with distance. The weight is the inverse of the distance raised to the power so in this case, as the distance increases, the weight decreases by power of 2. A power of two is the default setting for the IDW.

The IDW output feature, seen in figure 4, shows the same trends as were observed in the raw data. The southwest-northeast trend is visible, as well as regions of finer sediments on either side of the coarser sediments.

Although the IDW surface appears to be a reasonable representation of the surficial sediments, there are some problems with the surface. The change in mean phi size for sediments across a large area would be expected to be fairly
constant. There are not many processes in an ocean setting that would create
dramatic, localized changes in the grain size of the sediment. Infact, natural
processes such as tides and currents would be expected to create a much more
gradual change in grain size. Yet, there are many localized dramatic changes in
grain size in the IDW output. These can be seen in a transect of the output seen
in figure 5.

The rapid variance in the interpolated IDW surface is due to the method by
which the interpolation is conducted. The relationship between distance and
variance is fixed so as you get farther from the points, the influence decreases at
a constant rate, albeit exponentially. There is no accounting for a special way in
which distance and variance may relate for a particular data set. A more
sophisticated interpolation method is needed to take this into account.

The Ordinary Kriging method for interpolation uses specifics gathered
from the data to determine how to relate variance and distance when
interpolating points.

The first Ordinary Kriging interpolation in the project was done using the
default settings. The default settings were used to determine if the Kriging
process was better at modeling the surficial sediments than the IDW process. A
prediction map was generated since predicting the surficial sediment grain size
was the goal of the project.
Using the Geostatistical Wizard in the Geostatistical Analyst extension, in the Ordinary a spherical model was used with major and minor ranges of 7.8965. A neighborhood search was used with preferably 5 neighbors and a minimum of 2. There was no accounting for anisotropy in this interpolation. The search range was circular so no weighting was given to any particular direction. The result of this Ordinary Kriging can be seen in figure 6.

A comparison of the IDW and Ordinary Kriging methods was made to determine what changes had been made by switching the interpolation methods. The Raste Calculator of the Spatial Analyst extension was used to subtract the IDW raster from the Ordinary Kriging raster.

To improve the interpolation of the data, the anisotropy in the data needs to be accounted for. The simplest way of doing this is to incorporate the anisotropy into the Ordinary Kriging process. The anisotropy Ordinary Kriging prediction interpolation was similar to the first Ordinary Kriging interpolation except that the anisotropy feature was selected in the Geostatistical Analyst extension’s Geostatistical Wizard.
The anisotropy feature allows for the search area for neighboring points to be asymmetrically shaped. This results in a search area which is preferentially skewed towards one direction. In the case of the anisotropic Ordinary Kriging, the search ellipse had a major semiaxis of 5.6589 and a minor semiaxis of 1.2135. These were the program defined outputs for the anisotropy parameters. User defined parameters were not used because of the complexity of the Kriging process and the scope of the project. Figure 8 shows the output of the anisotropy Ordinary Kriging process.

Figure 8: A map showing the output of the Ordinary Kriging interpolation when anisotropy is considered.
ANALYSIS/INTERPRETATION

All three of the interpolation methods resulted in surfaces that were understandable and reasonably showed the surficial sediment size in the Gulf of Maine region. The goal of the project was to create a surface that was useful in predicting mean grain size at any location within the study area. Each of the interpolations has its benefits and drawbacks.

The IDW interpolation has the benefit of being a quick process, but it lacks the accuracy of the more sophisticated Kriging methods. The fact that the weighting method is based solely on the power defined by the user. In this case, the power was 2 so as the distance increased, the weighting changed by inverse of the square of the distance. This is a fixed ratio so there is no accounting for any variance in the data. Some data may have a relationship between variance and distance that is not an inverse square, it may be more or less complicated and may have some outer limit where the effect of the point becomes minimal. The lack of sophistication in the data is manifested in the fact that the IDW surface has many pock marks or circular, dramatic changes in phi size. This variance is due to the fact that the IDW interpolation only looks at distance from a
data points to determine values. If there is a point that is dramatically different from those around it, it will create a ring around it due to its influence. There is no way for the other points to counteract this influence.

The IDW process also yields a surface that is not as exact as desired. Between points, the IDW interpolation results in dips. This is because of the decreasing influence of the points. These dips in the surface are detrimental when trying to predict mean phi size. A smooth, general trend is much more likely than a series of dips between known data points. For these reasons, the IDW interpolation is not useful as a predictive tool.

The Kriging interpolation method is a much more sophisticated method. The semivariance of the data is compared with distance to determine how variance changes with distance. It is the resulting equation that is used to predict values at points in between data points. The fact that the weighting method is derived from the data causes the interpolation to be more representative of the data. As a result, the interpolation is more accurate to what is expected from natural processes.

The Ordinary Kriging output shown in figure 6 is a much smoother than the IDW output. There are fewer of the dips and overall, the data is much more gentle in its transition between grain sizes. The IDW surface had a much less smooth transitions between grain sizes and contained many abrupt changes in grain size that would not be expected in the natural world. Environmental processes would create a gradual gradient even if the initial boundary was abrupt. Waves, currents and other actions move sediment around and would mix boundary zones to create a gradient between the sediment sizes.

The Ordinary Kriging output has odd looking areas in the southeast area of the study. These are most likely a by-product of the Kriging process. The Kriging process chooses the data points closest to the point where the interpolation is occurring. In the area where this trending is occurring, there are no data points so all of that area uses the same data points, and the distance between interpolated points is small compared to their distance from the data points, the interpolated points all receive the similar values. The geometry of the data points creates these streaks of sediments. The fact that there are no local data points makes this region of the interpolation suspect and it should not be used for prediction.

The Ordinary Kriging output is a reasonable predictive tool but it does not take into consideration the anisotropy in the data. The second Kriging method accounts for the anisotropy and incorporates it into the processing. The output from the anisotropy Kriging looked similar to the normal Kriging. The anisotropy Kriging seemed to have much more north-south trending which is most likely artifact due to the search ellipse. The search ellipse defined by the major and minor semiaxes for the anisotropy Kriging was very long in the north-south direction and very thin in the east-west direction. This would bias the interpolation to a north-south trend. These trends can be seen in figure 8. There are probably not numerous local north-south trends in the sediment sizes. These trends are artifact due to the Kriging process.
The best way to determine which of the Kriging methods provides the most reliable surface is to use the prediction errors generated by the software during the interpolation process. ArcMap gives mean, root-mean-square, average standard error, mean standard error and root-mean-square standardized errors. According to ESRI's “Using ArcGIS Geostatistical Analyst” document “…a model that provides accurate predictions, the mean error should be close to 0, the root-mean-square error and average standard error should be as small as possible (this is useful when comparing models), and the root-mean square standardized error should be close to 1.” As seen in figure 9, the anisotropy Kriging method has the mean error closer to 0, the smaller average standard error and the root-mean-square standardized error closest to 1. The non-anisotropic Kriging method has the smaller mean standardized error and root-mean-square error. The splitting of the errors means that it is difficult to determine which is the better predictive Kriging method.

Given that the errors do not help to determine the better predictor of mean sediment size, the fact that the non-anisotropy Kriging has fewer obvious artifacts due to the Kriging process, it is more likely to be a reliable predictor of surficial sediments in the study area. It is reliable in the manner that the major
east-west trend of coarser sediments corresponds with Long Island and Cape Code, both of which are known to be end moraines from glaciers. It is logical that the end moraines would continue off shore. The Kriging output has shown that there is a definite off shore trend that might be this end moraine. The finer sediments inshore from the end moraine might be sediments eroded from fine inland formations such as the Waterville Formation. These sediments are washed into the Gulf of Maine and are accumulated in higher concentrations due to the end moraine preventing flow out to sea. These are hypotheses but they correspond to known data and are not unreasonably far fetched.

CONCLUSIONS

The project was sucessful in some ways and unsuccessful in others. It was sucessful in that a surface was created that reasonably models known geological features. It can be used as a predictive tool, but the accuracy cannot be known until samples are taken an compared to the predicted values. A model was produced, but the usefulness of the model is not known.

One problem is that there is not a good way of determing which of the two Kriging surfaces is the more accurate. The prediction errors are split fairly evenly between the two. Further study into the errors as well as refinement of the Kriging process are needed to fully determine which of the two Kriging processes is the more reliable and accurate.

In future studies, I would do more research into anisotropy and the Kriging process in order to minimize prediction errors and create a more reliable map. Kriging is not a simple process and it is necessary to take time to learn how the process works if you want to maximize its potential.