VISUALIZING HIGH-DIMENSIONAL, MULTI-PARAMETER GEOLOGICAL DATA

Matthew O. Ward[1], Frank Horowitz[2]
[1]Department of Computer Science, Worcester Polytechnic Institute, Worcester, MA USA
[2]CSIRO Division of Exploration and Mining, PERTH, WA.

Summary - This paper presents an overview of several methods for graphically displaying multivariate data and how these methods can be applied to geological data. A software package, named *XmdvTool* has been developed which allows users to examine their data sets in a variety of ways.

1 INTRODUCTION

Traditional 2-D and 3-D visualization techniques are limited in their effectiveness at conveying details about data sets which are high-dimensional or multivariate in nature, yet this is exactly the type of data being confronted by geologists and geophysicists today. The basic problem is that the number of data attributes to be displayed exceeds the number of graphical attributes (e.g. location, color, translucency) typically used in visualization.

In this presentation, *XmdvTool*, a software package for exploring high-dimensional data, will be described. The system allows users to display and interact with data sets using four classes of projection techniques. A major strength of the system is the ease at which the user can switch from one projection technique to another, which assists in the detection of interesting data features. Another important capability is the ability to define an *N-dimensional brush*, which allows users to highlight data points which fall within an arbitrary subsection of the N-dimensional space. Several examples of using the system to explore geological data sets will be presented, including the analysis of drill hole data, comparison of results from kriging routines, and the simultaneous display of multi-channel remotely sensed data.

2 MULTIVARIATE DATA DISPLAY

A wide variety of techniques have been developed over the years for mapping high-dimensional data onto a two-dimensional screen. Some of these include;

scatterplots: 2 of the N dimensions are used to control the position of the data point on the display. For N dimensional data, there are $N * (N-1)/2$ distinct scatterplots, which are usually arranged in a grid on the screen.

parallel coordinates: N axes are vertically aligned, and each data point maps to a polyline through the N axes.

glyphs: Each data point controls the drawing of a shape on the screen, and each dimension specifies one of the attributes of the shape.

hierarchical techniques: each dimension is broken into a discrete finite set of values, and 2 dimensions are selected to initially divide the screen into fixed size subscreens based on their cardinality. These subscreens are then divided by two more dimensions, with this process repeating until all dimensions are assigned. Each data point then maps to a particular section of the screen depending on the order of assigning dimensions.

Each technique has its strengths and weaknesses in terms of the number of dimensions which may be effectively displayed, the number of data points which can be accomodated, and the ease in which features of interest within the data may be detected and classified. The effectiveness of each technique may be enhanced by providing users with interactive controls over viewing parameters, color maps, and lighting characteristics.

Exploring the data using the graphical presentation can be facilitated by using an *N-D Brush*. A user specifies a brush size (a certain percent of the range for each dimension) and a brush location (a point for each dimension), and all data which falls within this subspace can be highlighted. This capability is very important, as any projection technique loses information on spatial relationships. Thus, without brushing it is very difficult to judge the true distance between data points.

3 GEOLOGICAL CASE STUDY FOR XMDVTOOL

As an example, we show below scatterplots from a combined SPOT, magnetics, and radiometrics dataset from the Grant's Patch region of Western Australia. The task is to locate in space the two spike anomalies evident in the Magnetics/Thorium scatterplot (found in the first panel diagonally up from the bottom left). One highlights the spikes by selecting the brush size and midpoint such that only the spikes are covered in that panel, and then adjusting the brush such that all of the other variables are fully covered. The results may be displayed in a spatially ordered map view by showing the glyphs. Assuming that the correlations come from a common rock type, this technique effectively identifies and roughly maps the two rock units that correspond to the two spikes. The procedure and resulting map will be demonstrated at the conference.

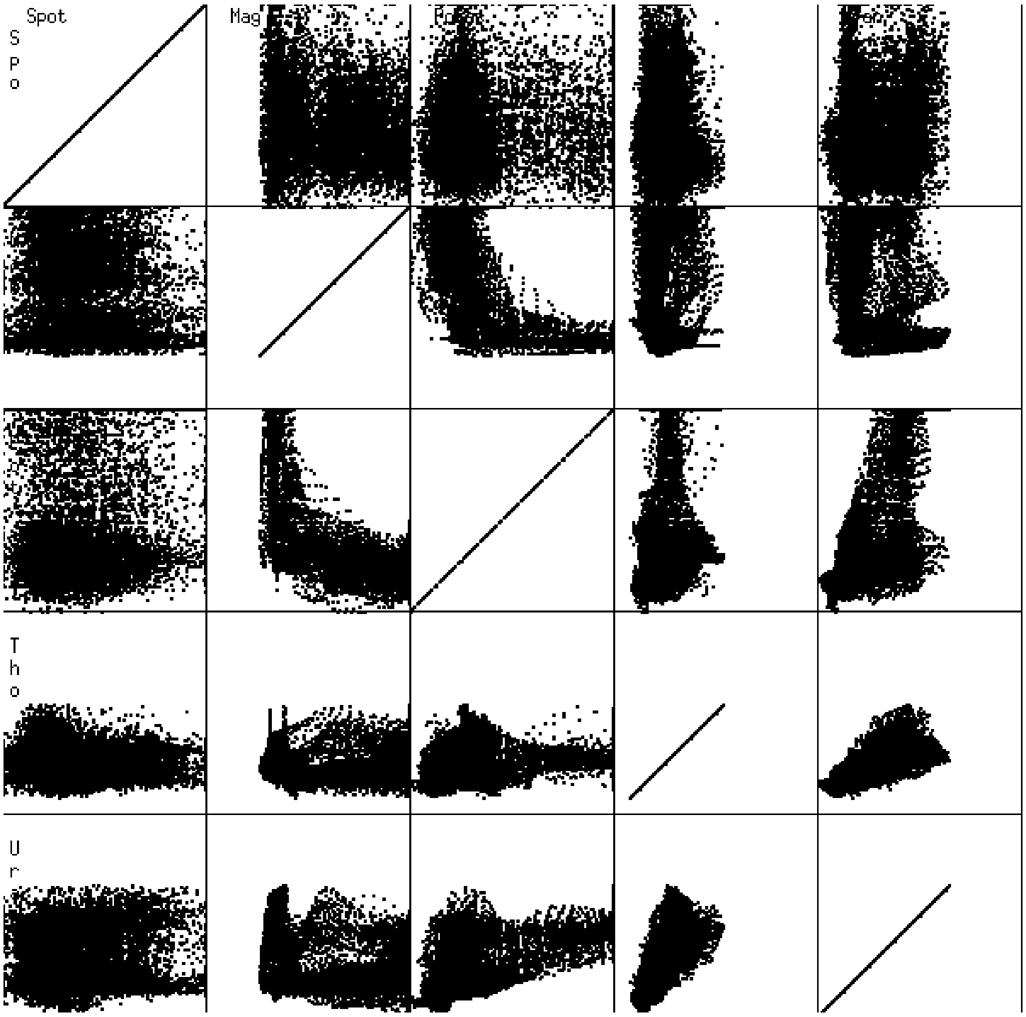

Figure 1: Scatterplots for the Grant's Patch region.