Grid Enabled Data Mining
High Performance Computing
Martin Keniron – NUI Galway
GeoHAB Conference 2004

What does Grid have to offer in Data mining on the National seabed survey?

- Security
- Processing Power
  Both Distributed and Parallel Computing
- Distributed Data Storage?

The Computational Grid

Grid computing (or the use of a computational grid) is applying the resources of many computers in a network to a single problem at the same time.
More information acquired about the surface of Venus than 70% of the Earth’s surface.

Dramatic Improvements in Sonar data acquisition.

10 times Ireland’s land area - Area 525,000 sq km.

Currently 4.2 Tbytes - expected to grow to 8 Tbytes.

Data Mining

Field data:
- Complete raw data

Raw Data:
- Tidal Corrections
- Water velocity Corrections
- Boat motion

Processed:
- Despiked
- Delete “dodge” Data
- Bined
- Destriped

The data-mining problem

Datasets are geological interpreted by visual inspection.

This results in a very complete analysis but slow.

As more datasets are remotely sensed – vast datasets

Automated algorithms – Computational cost

Linear Features Why?

Linear features - objects of geological interest.

Fractures (i.e. faults, joints and veins) and Fold crests and troughs all appear as sublinear features.

Layered sedimentary rocks bed terminations also appear as linear features.

Identification of linear features is vital for interpreting the geological history of the area.

The presence of fractures and fold crests are evidence of geological deformation and the orientations help identify the principal tectonic forces at work during deformation.
Edge Detection Algorithms
- not looking for edges

Neural Networks
- Neural Networks are an approach to computing that involves developing mathematical structures with the ability to learn.
- Neural networks have the ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complicated to be noticed by the human expert.
Another approach was to find a way to represent a linear feature mathematically so an automated algorithm could be able to pick such a feature from the $x,y,z$ bathymetry.

- A linear feature - a rapid change in slope in the bathymetry.
- Tessellated in small pieces in order to find such rapidly changing slopes.
- The slope and the gradient of each slope worked out for every region.
- Calculate the variance of the gradients in a region indicating if there is a rapidly changing slope present.

**Mathematically based Algorithm**

**Implementation**

- Triangle the in order to work out the slope and gradient of each tessellated region.
- Delaunay triangulation is one of the most popular for unstructured mesh generation.
- Delaunay triangulation – in circle test

**The Algorithm**

**Implementation**

The Convex Hull of a set of points $S$ is the smallest convex region that contains all points in $S$. The boundary of the region is called the convex hull of $S$, and is usually specified by a set of points and their neighbour relation.

There are several different paradigms for creating a convex hull:

- Plane Sweep
- Naive Algorithm
- Divide and Conquer
- Redirection from a Convex Hull
The Incremental Algorithm in 2-D.

When the data is tesselated in 2-D.

Now we can apply our mathematical algorithm to work out the each face of the 2-D Delaunay triangulation.

In order to check for a rapidly changing slope every triangle face that shares a vertex with a selected face will be include in its variance check.
Thank You

Questions !!