

Extrapolation technique used for geology mapping

The extrapolation technique was developed by the Enhanced Resource Assessment Team at NRE's Centre for Land Protection Research. The method is based on a technique established by Bui, Moran and Simon (1998) involving data extrapolation. For this study extrapolation techniques were used in generation and refinement of NRE-Minerals and Petroleum geology data into a seamless geo-spatial data layer.

The extrapolation process uses training data (environmental data attributes sampled according to geological mapping) to create rulesets that that can then be applied to areas of unknown or lower implied accuracy.

The extrapolation process involved:

- 1. Compilation of existing data.** The compilation of the pre-existing geology mapping in an ArcInfo coverage format. Given the number of mapping surveys (some with overlap), these surveys were assembled into a consistent data format.
- 2. Conversion to grid format.** The geology coverage is converted into a grid format using ArcGrid with cells 100 x 100 m in size. This grid cell resolution was chosen as it was considered significant for mapping at 1:100 000 scale.
- 3. Data sampling.** The area of data sampling (Figure 1) was chosen on the basis that it was similar to the region of extrapolation. While sampling from a large portion of the pre-existing geology grid may have provided useful information on all lithological types, the sample datasets were found to be extremely large and often resulted in software failures. Strategically, training data was sampled close to the area of extrapolation, as it was assumed that the local geology would represent the best match to the area of extrapolation.

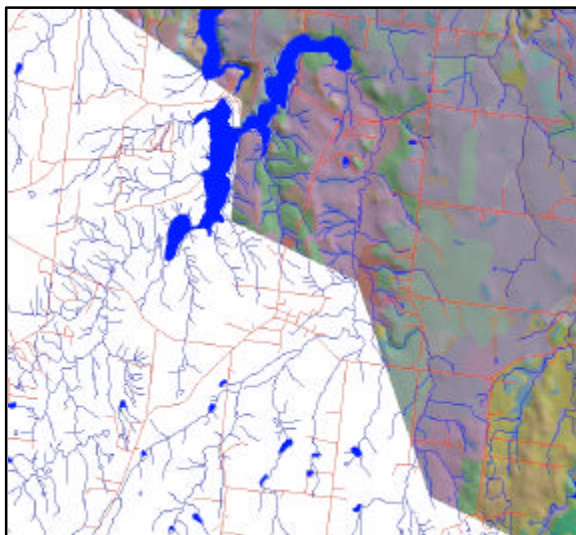


Figure 1 Existing survey data

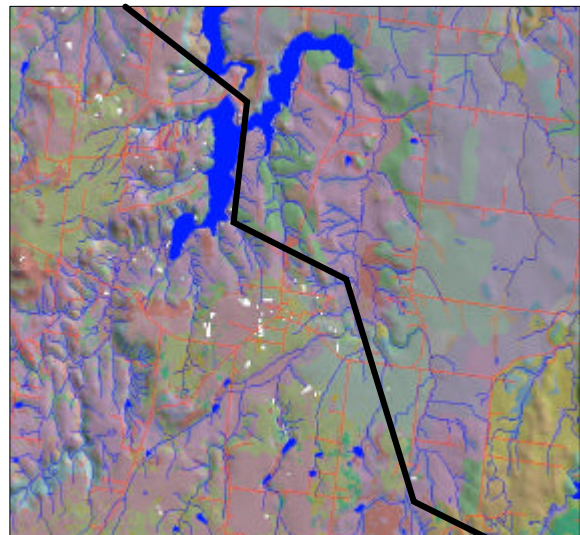


Figure 2 Extrapolated data from existing

The sampling procedure as developed by CLPR uses a sliding scale method where data is sampled according to an area-weighted random sampling approach (Burrough

1991). The data (training data per geology unit cell) was sampled using this approach through the application of a sampling AML (arc macro language) run in ArcInfo.

- If geological unit area is less than 10% of the sampled area, 100% of that unit type is sampled.
- If geological unit area is between 10-30% of the sampled area, sample according to their relative area using a simple linear regression.
- If geological unit area is greater than 30% of the sampled area, 30% of that unit type is sampled.

This sampling procedure ensures that a representative number of samples have been collected per geological unit.

4. Conversion of sampled data. The sampled data were then converted into a format that can be processed in a regression tree mining package (See5) where rules for the sampled data are formulated. The dataset was produced as an ascii file listing x and y co-ordinates of sampled cells, and the respective cell values of all layers specified. Some reformatting of this data was required to allow processing of data in See5. Also required to run the regression modelling in See5 is a names file that lists the geological units sampled as numbers.

5. Regression analysis. The data was processed in See5 where rulesets were developed. The output file was imported into Microsoft Access where the rules were viewed and manipulated. Sorting these rules allowed a grid output to be produced with an output of the highest attainable accuracy according to the implied accuracy of the output rules. This was exported in a format to run in ArcGrid via an AML.

6. Extrapolation applied. This extrapolation was then applied to the area of interest producing an output grid (geology) for that region (Figure 2).