Investigation of data mining and data clustering techniques for automatically generating proxy geology maps from airborne geophysical data.

This project, will investigate automated methods for generating proxy geology maps from airborne geophysical data using recent advancements in data mining and data clustering techniques. The techniques will be tested using data-sets from Newfoundland and Labrador. To first develop experience and expertise in current interpretation techniques, traditional processing and inversion methods will be applied to the Deer Lake, Indian Head, and Gros Morne blocks of the airborne magnetic data-set acquired in 2008-2009 for the Energy Branch and Nalcor.

The proposed research will focus on the following objectives:

1. Processed maps and localized 2D & 3D Earth models of features of interest for the Deer Lake, Indian Head, and Gros Morne blocks of the airborne magnetic data-set collected for the Energy Branch and Nalcor Energy in 2008-09, and for the data-set to be acquired off-shore western Newfoundland in the summer of 2012.

2, Geological interpretations of these data-sets done in collaboration with Energy & Mines Branch personnel.

3. Access to the new methodologies for automatically generating proxy geology maps from airborne geophysical data that are investigated and developed during the course of the Ph.D. research project.

4. Results of testing the new methodologies on data-sets from Newfoundland and Labrador.

Data mining and data clustering techniques such as principal component analysis, K-means clustering, fuzzy c-means clustering, and self-organizing maps aim to segment a multicomponent data-set into groups of data points that have similar properties. An example of a multi-component data-set is the combination of magnetic, gravity, and radiometric data acquired over a geographical area by one or more airborne surveys. The particular magnetic, gravity and radiometric values observed at an observation location depend strongly on the dominant rocktype at and around the measurement location.

Certain rock-types will give high gravity and magnetic readings, other rock-types will give high magnetic but low gravity readings, others low magnetic and high gravity readings, and so on. Data clustering techniques attempt to recognize automatically these different groupings of readings in the input geophysical data-sets, that is, which measurement locations have high magnetic and high gravity values, which locations have high magnetic but low gravity values, etc. The groupings, or clusters, that are identified in the geophysical data can then be associated with particular rock-types. The geographical distribution of measurement locations belonging to each cluster then amounts to a proxy or pseudo geology map.

The goal of this project, following the lead provided by Fraser and Dickson (2007) and Carneiro et al. (2011), is to investigate the use of modern data mining and clustering techniques for the automatic generation of proxy geological maps, or the refinement of existing geological maps, from airborne geophysical datasets.

Geophysical data-sets from Newfoundland and Labrador, and from elsewhere in Canada if necessary, will be used to test the data mining and clustering techniques. Data-sets from areas for which the geology is well known will be used to examine just how well the proxy geology maps produced by these techniques resemble the maps derived from traditional geological mapping. Data-sets from regions for which the geology is poorly known, because of cover, lack of outcrop or accessibility, will then be considered.

Before investigating the new data mining and data clustering techniques mentioned above, traditional processing, inversion and interpretation techniques will be applied to the Deer Lake, Indian Head, and Gros Morne blocks of the airborne magnetic data-set acquired for the Energy Branch of NL Department of Natural Resources and Nalcor Energy from October 2008 to May 2009, and to the data-set to be acquired off-shore western Newfoundland, summer 2012. This will provide a solid, thorough background in the methods currently used to interpret geophysical potential field data, and thus will allow a critical comparison of the proposed new methods with current practice. Also, this will result in a deeper, more substantial analysis and interpretation of this data-set than has so far been possible.