Geophysical Research Abstracts
Vol. 14, EGU2012-540, 2012
EGU General Assembly 2012

Lithological Uncertainty Expressed by Normalized Compression Distance

J. Jatnieks, T. Saks, A. Delina, and K. Popovs
Faculty of Geography and Earth Sciences, University of Latvia, Riga, Latvia

Lithological composition and structure of the Quaternary deposits is highly complex and heterogeneous in nature, especially as described in borehole log data. This work aims to develop a universal solution for quantifying uncertainty based on mutual information shared between the borehole logs. This approach presents tangible information directly useful in generalization of the geometry and lithology of the Quaternary sediments for use in regional groundwater flow models as a qualitative estimate of lithological uncertainty involving thousands of borehole logs would be humanly impossible due to the amount of raw data involved. Our aim is to improve parametrization of recharge in the Quaternary strata. This research however holds appeal for other areas of reservoir modelling, as demonstrated in the 2011 paper by Wellmann & Regenauer-Lieb.

For our experiments we used extracts of the Quaternary strata from general-purpose geological borehole log database maintained by the Latvian Environment, Geology and Meteorology Centre, spanning the territory of Latvia. Lithological codes were generalised into 2 aggregation levels consisting of 5 and 20 rock types respectively.

Our calculation of borehole log similarity relies on the concept of information distance proposed by Bennet et al. in 1998. This was developed into a practical data mining application by Cilibrasi in the 2007 dissertation. The resulting implementation called CompLearn utilities provide a calculation of the Normalized Compression Distance (NCD) metric. It relies on the universal data compression algorithms for estimating mutual information content in the data. This approach has proven to be universally successful for parameter free data mining in disciplines from molecular biology to network intrusion monitoring.

To improve this approach for use in geology it is beneficial to apply several transformations as pre-processing steps to the borehole log data. Efficiency of text stream compressors, such as prediction by partial matching (PPM), used for computing the NCD metric, is highly dependant on context. We assign unique symbols for aggregate lithology types and serialize the borehole logs into text strings, where the string length represents a normalized borehole depth. This encoding ensures that both lithology types as well as depth and sequence of strata is comparable in a form most native to the universal data compression software that calculates the pairwise NCD dissimilarity matrix.

The NCD results can be used for generalization of the Quaternary structure using spatial clustering followed by a Voronoi tessellation using boreholes as generator points. After dissolving cluster membership identifiers of the borehole Voronoi polygons in GIS environment, regions representing similar lithological structure can be visualized. The exact number of regions and their homogeneity depends on parameters of the clustering solution.

This study is supported by the European Social Fund project No. 2009/0212/1DP/1.1.1.2.0/09/APIA/VIAA/060

Keywords: geological uncertainty, lithological uncertainty, generalization, information distance, normalized compression distance, data compression

References:

Bennett C. H., Gacs P., Li M., Vitanyi P., Zurek W. 1998, Information Distance, IEEE Transactions on Information Theory, 44(4), 1407-1423.

Cilibrasi R., 2007., Statistical Inference Through Data Compression, ILLC Dissertation Series DS-2007-01, Institute for Logic, Language and Computation, Universiteit van Amsterdam.

Wellmann J.F., Regenauer-Lieb K., 2011, Uncertainties have a meaning: Information entropy as a quality measure for 3-D geological models, Tectonophysics, Elsevier (in press).