BWMining

QUEST Project: Geoscience BC Using Geochemistry and Neural Networks to map Geology under Glacial Cover

Supplementary remarks on preprocessing the geochemical data

Colin T Barnett* Peter M Williams†

April 2010

1 The leveling process

The aim of releveling is to provide a unified picture of the geochemistry on a regional scale. The available data, however, comprise several different surveys, carried out at different times, collected from different sample media, and analysed using different laboratory techniques. It is necessary, therefore, to relevel the various surveys to minimise mismatches between populations, so as to present a composite picture suitable for regional exploration.

The approach adopted was an empirical one designed to match statistics across the various surveys. Ideally it would be possible to compare assays from the various surveys directly at the same stations. In reality, however, none of the stations coincide and, indeed, few of the survey areas even overlap. This creates a problem, because simple matching of statistics, across the whole of neighboring survey areas, would imply that these statistics are constant over distances of the order of 100km. Although this might sometimes turn out to be approximately true, it should not be assumed.

The chosen solution was to compare the statistics only of small immediately abutting areas, comprising border strips from neighboring survey areas. Although these areas are still generally disjoint, their separation is only of the order of a few kilometers, rather than hundreds of kilometers for the survey areas as a whole, so that much less spatial variation is to be expected. These areas, however, are necessarily quite small, and may not include enough sample stations to give reliable results; furthermore, the actual number of stations will depend on which neighboring surveys are being considered. Consequently, in order to obtain sufficient consistent sample sizes for neighboring surveys, each of the geochemistry

*BW Mining, Boulder, Colorado, USA
†BW Mining, Brighton, Sussex, UK

elements was interpolated onto a regular grid. As a result, border strips of the same width on either side of the boundary between two survey areas will always contain the same number of grid cells, and hence the same number of data points if data are now taken from the grids. The gridding method, applied to the log assays, is described below.

The general procedure for leveling two neighboring surveys, for a given element, is now as follows. On either side of their common border, we take a 10km wide strip and assemble the log assays, for the given element, at grid cells within these two border strips. This provides two sets of samples, of approximately the same size, from each of the two neighboring surveys. If the two surveys were already consistently measuring the same quantity, we should expect any sample statistics of these sets of values, such as their means and standard deviations, to be the same. This suggests a way of leveling the two surveys by transforming one or other, or both, to ensure that their means and standard deviations agree. Such a transformation involves a linear shift, in one direction or the other, combined with a positive multiplicative rescaling: in other words, a simple linear transformation of log assays, applied subsequently throughout the surveys in question.

Although we used such a linear transformation in log space for releveling surveys, we adopted a different way of determining its coefficients. The reason is that the mean, and especially the standard deviation, can be sensitive to outliers. An equally valid alternative would be to ensure that the two sample collections have the same upper and lower quartiles. For example, if there were 100 assays in each of the two border strips, that would imply an adjustment so that the 25th largest and the 75th largest in each of the two collections have the same magnitude. That can be achieved by a linear transformation, and in fact a unique one. Such an approach has the merit of being independent of the magnitude of outliers, since the 75th largest, for example, does not depend on how much larger or how much smaller other values are, only on their positions in the ordering.

It should be noted however that, even if the outer quartiles agree, the medians of the two samples may still differ. This is because there are only two free parameters of a linear transformation, allowing only two percentiles to be fitted exactly. There is no reason, however, for giving priority to the outer quartiles over the median. We might therefore choose to use the best compromise fit to all three, even if the fit is not exact to any of them. Taking this a step further, it becomes apparent that there is nothing special about quartiles. We could equally use the best fit to the four quintiles or to any other collection of percentiles. The solution adopted in this study was to find the best fit to the nine deciles, in other words the values that divide the sorted data into ten equal parts. The reason for choosing as many as nine percentiles is that the results are then more sensitive to a wider range of assays values. If a larger number were used, the percentiles would risk becoming noisy as a result of small sample sizes.[1]

It is important to note that, using this procedure, the exact distributions of outliers, in the top 10% or bottom 10%, have no effect when calculating the releveling parameters.

[1]It has recently come to our attention that Daneshfar & Cameron [2] adopted a similar approach to the leveling problem: fitting a straight line to matching percentile points of the distributions in neighboring border strips taken from adjacent sheets.

However, once these releveling parameters have been determined by the positions of the decile points, the same releveling procedure is applied to all samples, including any outliers.

1.1 Streams and lakes

The previous section has described the general way in which two surveys, or composites of surveys, were releveled. In practice it was found best to form separate blends of the stream sediments and of the lake sediments on their own, and then finally to blend the two composites as wholes.

The order of leveling the dozen or so surveys that cover the QUEST area was essentially based on trial and error. The only difficulty comes towards the end of the process when there are just a few blocks remaining which have to be slotted in to match their neighbours on three or four sides. After considerable experimentation on a few key elements like Au, Cu and Mo, an order was established which was then routinely applied to all 42 elements.

After releveling, the various component grids were stitched together using a cosine taper. The overall level was taken from the most recent stream sediment survey, collected in 2007 from quad sheet 93O.

2 The gridding process

Gridded data are the result of converting scattered individual data points into a regular grid of interpolated values. This conversion process is called gridding. The resulting grid is often easier to analyze and display than the original scattered points.

The methods most widely used for gridding geochemical data are minimum curvature and kriging. Both methods have their merits. On balance, kriging seems better suited for localized rather than regional studies. Indeed the original problem studied in [3] was to estimate the properties in the gaps between rock samples already drilled on a regular grid. Kriging is optimal if its underlying statistical assumptions are met and the covariance function of the process can be reliably estimated. In practice regional geochemical measurements only approximately form a Gaussian process and, even then, it is not easy to obtain a good estimate of the covariance function from sparse and irregular data. In common with other investigators [4, 5] therefore, we have used an adjustable tension, continuous curvature, surface gridding algorithm for this regional study. Variation of tension is equivalent to changing the kriging covariance function, but in a more directly manageable way [6].

3 The imaging process

The imaging process used to present the geochemical data in our report was a superposition of two visualization techniques known as shaded relief and colour coding. The shading was derived by treating the geochemistry as a pseudo-topographic surface and calculating the slopes and illumination of this surface by sunlight shining from the north. The colouring is directly related to the amplitude of the geochemical samples, with purple assigned to the

lower elevations, through blue, green and yellow to red at the higher elevations. The result is an image which is highly sensitive to small variations in the data. This allowed us to carefully monitor the leveling process for satisfactory blends between adjoining surveys. It also revealed strong correlations with the geology, which in turn lead naturally to the idea of using neural networks to map the bedrock concealed beneath a thin veneer of glacial cover.

References

[1] Barnett, C. T. and Williams, P. M., Using geochemistry and neural networks to map geology under glacial cover, *Geoscience BC Report 2009-3*, 2009.

[2] Daneshfar, B. and Cameron, E., Leveling geochemical data between map sheets, *Journal of Geochemical Exploration*, **63**, 189–201, 1998.

[3] Krige, D. G., A statistical approach to some basic mine valuation problems on the Witwatersrand, *J. of the Chem., Metal. and Mining Soc. of South Africa* **52**, 119–139, 1951.

[4] Kelley, K. D., Eppinger, R. G. and Walton-Day, K., Selected geochemical studies related to mineral deposits by the U.S. Geological Survey (USGS), *Explore*, **116**, 2002.

[5] Folger, H., Yager, D., Ludington, S. and Stillings, L., Landscape geochemistry for the Humboldt River Basin, *Geological Society of America*, Seattle Annual Meeting, 2003.

[6] Smith, W. H. F. and Wessel, P., Gridding with continuous curvature splines in tension, *Geophysics*, **55**, 293–305, 1990.