Improving National Uranium Resource Evaluation (NURE) Data Using Resampling and Statistical Adjustment:
A Methodology

Robert Barber-Delach – Oak Ridge Associated Universities, USGS National Center, Reston, VA 20192

Introduction

The National Uranium Resource Evaluation (NURE) program was initiated by the Atomic Energy Commission (AEC) in 1973 with a primary goal of identifying uranium resources in the United States. After a couple agency changes during the 1970's, the program was continued under the Dept. of Energy in 1977. The Hydrogeochemical and Stream Sediment Reconnaissance (HSSR) program was one of nine components of NURE. Planned systematic sampling of the entire United States was conducted by four national laboratories, but only just over half of the 1°x2° quadrangles were completed by the end of the program in 1983-84. Each DOE laboratory developed its own sample collection, analytical, and data management methodologies, which has led to biases in the HSSR/NURE data (Smith 1999), not only between quadrangles sampled by different labs, but sometimes also between quadrangles sampled by the same laboratory.

NURE samples are archived by the USGS in Denver, CO and are available for re-analysis. The USGS archive reportedly contains about 380,000 original sediment samples from all four laboratories, about 250,000 replicates, splits, size fractions or other samples (Smith 1999).

Concentrations in the NURE stream sediment data show quadrangle-based biases in virtually all elements for which a significant number of samples exceed detection limits (23 elements). The scatter plot for Magnesium concentration (figure 1) show how strongly quadrangle-biased the NURE data can be, and further demonstrate how both quadrangles vary significantly from our re-analysis of the data. This consistency of bias within a quadrangle allows derivation of regression equations that can adjust the original NURE data to correct for previous analytical biases and produce a better accord with new analytical data. Although this technique has inherent statistical uncertainty, that uncertainty is quantifiable, and the resulting element maps are significantly more accurate than those made from the original NURE data. Although this technique does require the cost and time of re-analyzing a significant number of NURE samples, the result produces a much improved data set without the cost of re-analyzing the entire sample set.

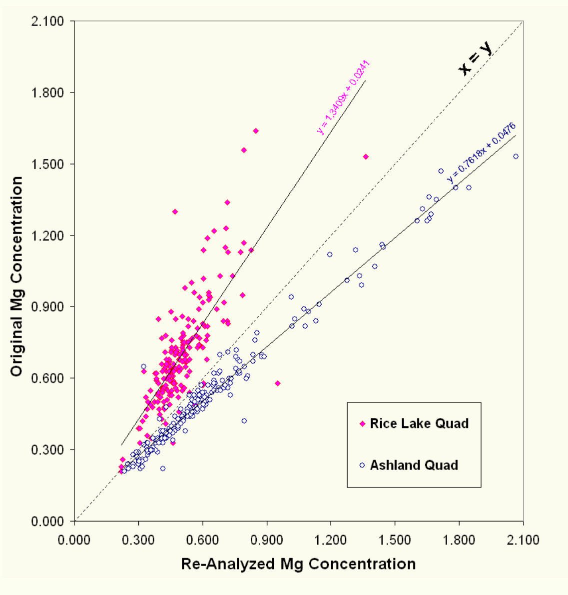

Figure 1. Scatter plot of magnesium concentration demonstrating the quadrangle bias of the original NURE stream sediment data, with equations of fitted regression lines.

Methodology

Analyzing splits of the archived NURE samples with modern lab analysis techniques allows for statistical comparisons to the old NURE data. We re-analyzed a randomized sample of approximately 200 samples per 1°x2° quadrangle. Our analyses were done using a routine 40-element inductively coupled plasma-atomic emission spectroscopy (ICP-AES) technique supplemented by a more precise single-element technique for arsenic. Analytical precision and accuracy were monitored by replicate analyses of standard samples.

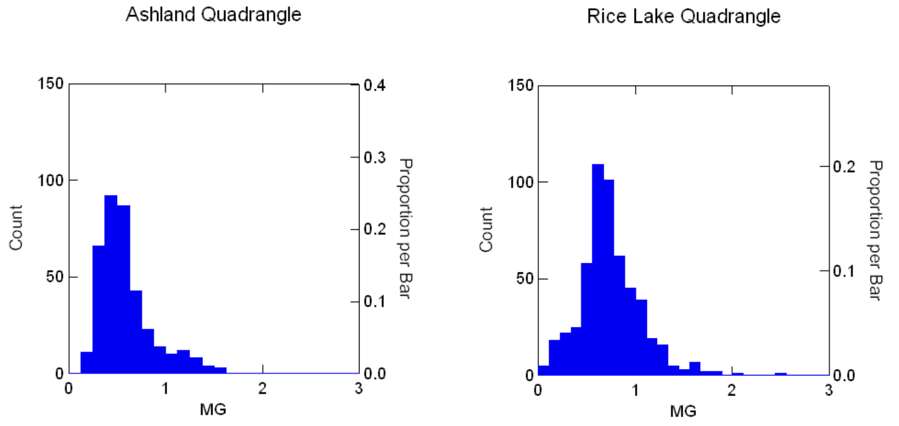

Figure 2. Histograms of the distribution of Magnesium concentration from original NURE stream sediment data, by quadrangle.

We first evaluated the distributions of both the original and reanalyzed data for normality by quadrangle. This step is important to ensure that assumptions of normality for the statistical tests used during the adjustment of the data are not violated. Nearly all the trace elemental data for both quadrangles were strongly skewed, and were subsequently transformed using a $\log_e$ transformation. The major elements recorded in percent concentration (Al, Ca, Fe, K, Mg, and Na) were fairly normally distributed, and did not seem to benefit from a transformation (see figure 2 for an example). The major elements in parts per million (ppm), Phosphorus and Titanium, were log$_e$ transformed. In some circumstances, an arcsine transformation may be appropriate for percent data (Zar 1996) such as concentrations of major elements. Non-detect values in the datasets were replaced with 50% of the non-detect value in each instance to more accurately represent those values for each element.

Upon the completion of transformations, I computed simple linear regressions from the data, with the dependant variable set to the reanalyzed data and the independent variable to the original, for each element. If the regression for a given element identified outliers in the data I removed them and re-ran the regression until I got a regression result without outliers. The dropping of outliers before performing a simple linear regression is important because they can violate the assumptions of normality and homogeneity of variance in the data (Zar 1996).

I adjusted the original data using both the coefficient (slope) and the constant (Y intercept) from the regression for each element as follows:

$$\textbf{adjusted} = (\textbf{original} * \textbf{B}) + \textbf{A}$$

where: A is the y-intercept constant, and B is the slope coefficient. The next step was to untransform the adjusted data, and then evaluate the success of the adjustment by displaying the adjusted data vs. the re-analyzed data, as in Figure 3. In the example in Figure 3, the between quadrangle bias was significantly reduced by the adjustment of the data.

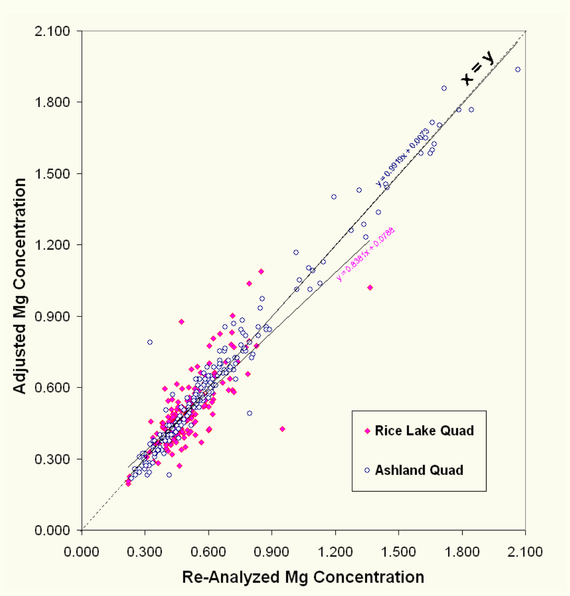

Figure 3. Scatter plot of magnesium concentration after adjustment of the original data using the regression method, with equations of fitted regression lines.

Table 1. List of elements reanalyzed, type of transformation employed, and units.

Element	Transformation	Concentration Units
Aluminum (Al)	none	percent
Arsenic (As)	n/a	ppb
Barium (Ba)	Log$_e$	ppm
Calcium (Ca)	none	percent
Cobalt (Co)	Log$_e$	ppm
Cromium (Cr)	Log$_e$	ppm
Copper (Cu)	Log$_e$	ppm
Iron (Fe)	none	percent
Potassium (K)	none	percent
Lithium (Li)	Log$_e$	ppm
Magnesium (Mg)	none	percent
Manganese (Mn)	Log$_e$	ppm
Sodium (Na)	none	percent
Niobium (Nb)	Log$_e$	ppm
Nickel (Ni)	Log$_e$	ppm
Phosphorus (P)	Log$_e$	ppm
Scandium (Sc)	Log$_e$	ppm
Strontium (Sr)	Log$_e$	ppm
Thorium (Th)	Log$_e$	ppm
Titanium (Ti)	Log$_e$	ppm
Vanadium (V)	Log$_e$	ppm
Yttrium (V)	Log$_e$	ppm
Zinc (Zn)	Log$_e$	Ppm

Three trace elements in our dataset did not significantly benefit from the adjustment. The adjusted result for Niobium (Nb) and Nickel (Ni) were not significantly better than the original data before adjustment. For Niobium (Nb), this seemed to be due to its non-linear distribution, and for Nickel (Ni) due to its apparent lack of quadrangle bias prior to adjustment. Arsenic (As) shows a strong quadrangle bias, yet was not improved by the adjustment; therefore, we chose to adjust it using differences in the means of the original and re-analyzed data. This adjustment did not change the statistical distribution of the As data, but corrected the means of the original data to those of the re-analyzed values.

References

Smith, S.M. 1999. Reformatted Data from the National Uranium Resource Evaluation (NURE) Hydrogeochemical and Stream Sediment Reconnaissance (HSSR) Program. U.S. Geological Survey Open File Report 97-492. URL: http://greenwood.cr.usgs.gov/pub/open-file-reports/ofr-97-0492/index.html. Version 1.10 updated July 29, 1999.

Zar, J.H. 1996. Biostatistical Analysis. Prentice-Hall, Inc. NJ.