

### Appendix A3.9: ML Regression

The input file for the regression program is similar to that described in Appendix G, except that following the columns of analytical data are columns of the areal percent of each rock type in each basin. The user must again specify the input file name, number of samples, format statement for the geochemical element being regressed, and whether or not the log-transformation is required.

The program produces two output files: OUTPUT.OUT, which contains the regression constant ( $b_0$ ), regression coefficients for each rock type ( $b_j$ ) and the standard deviation ( $s$ ) produced by the ML regression. The  $b_j$  values are the average amount of the element present in each rock type  $j$  underlying catchment basin  $i$ . These values can be used in Equation (5) in Section 4.4.1 to calculate the predicted value of an element ( $y_i$ ) based upon the geochemical composition of the sample and areal proportion of geologic units present in the drainage basin from which the sample was taken. These predicted values are calculated automatically and are found in the second output file, RESULT.OUT. It also contains a sequence number (related to the basin number), the original analytical value, and the calculated percentiles (discussed in Section 4.3.4) for each basin. Initially, the file is unordered, listing the basins with undetected values first. Once the "<" signs have been added if required, the file can be sorted by sequence number.

The percentile values represent the predicted range of values for that basin based upon the underlying lithology. The original analytical value for an element in each basin can be compared to this new "background scheme" and is recorded as falling into one of the following classes: <80th percentile, 80-90,

90-95th, 95-98th or >98th percentile. Data that are undetected are left in a special class so that the resulting geochemical maps clearly reflect the true character of the data.