Appendix 4.3. Additional Notes on Kleiner-Hartigan Tree Classification

A-4.3.1. Introduction to major and trace-element classification methods

This appendix details the application to Nahanni spring waters of a method of grouping multivariant chemical data known as Kleiner-Hartigan Trees (Kleiner and Hartigan, 1981).

The chemical analyses (Appendix 4.2) include the seven major ions and 20 trace elements. Three of the trace elements (Au, Sb and Pb) were analyzed in 1987 and two others (Ag and Ba) had so few samples above detection that they can not be dealt with statistically. This leaves a total of 22 elements (plus pH) which can be used to produce a geochemical classification of the waters.

A common way of grouping data is to graph it and pick out obvious clusters. However, when dealing with more than just a few variables they become difficult to represent in two dimensions and it becomes increasingly more difficult as the number of variables increases. One method of dealing with this problem is to produce an individual plot for each case (sample), the shape of which is dependent on the data (elements). The plots can then be grouped on the basis of their shapes. There have been many variations on this theme which have produced a number of different types of representative shapes.

One of the more successful shapes is that of a tree. A computer is used to generate the plots and samples are then grouped by pattern recognition. Similar shaped trees represent waters with similar chemistry. It is only a computer-aided method because the decisions as to what constitutes a group and which samples fit into that group are still left up to the investigator. A program to plot "Kleiner-Hartigan trees" was made available to Hamilton by R.W. Garrett at the Geological Survey of Canada.

A-4.3.2. Kleiner-Hartigan Methodology

Before developing tree plots, Kleiner and Hartigan (1981) reviewed most of the existing pattern recognition systems such as castles, boxes, glyphs, star plots, profiles and faces. The principle behind all of these systems is similar to that of the tree system, as follows. Correlation coefficients are calculated for the variables, followed by cluster analysis (explained in detail in section 5.4.4) on the correlation matrix. The variables, which vary together, are placed on the same "branch" of the tree. Therefore, the configuration of the tree is determined by how the variables vary with respect to each other throughout the whole data set. The variables represent "leaves", the lengths of which are proportional to concentration. The length of the supporting branch is proportional to the combined concentrations of all the leaves on the branch. This produces a tree with the same configuration of variables, but with a different shape for each case, provided that the chemistry of the water is different. The 22 variables, already mentioned, were used to produce the "K-H trees"; however, pH was not included. Details of the steps employed to produce the trees are given in Hamilton (1990).

A Kliener-Hartigen "Tree" plot (Figure 5.3, Section 5.4.3.) shows the resulting configuration. The tree has 6 major branches or sub-branches, each having a number of co-variant elements on them as "leaves". Because elements which vary together throughout the sample set were positioned close together, groups of similar waters tend to have one or several enlarged branches. The six sub-branches discussed below contain trace element associations which are significant with respect to the genesis of the waters. Resulting groups are listed in Table 5.2.

A4.3.3. Results and discussion of Kleiner-Hartigan Tree Groupings

Kleiner-Hartigan tree classification plots for the Ragged Ranges (text, Figure 5.4) and the Nahanni Karst study areas (Figure 5.5) show the results of visual pattern recognition on the tree-plots. Ungrouped samples are shown on Figure 5.6. It can be seen from the figures that the variation in shape of the plots makes some water types easily identifiable.

On Branch A (Fig. 5.3), Ca^{2+}, Mg^{2+}, $SO4^{2-}$, HCO_3^- and U surprisingly vary together. Uranium forms soluble carbonate complexes such as $UO_2(CO_3)_3^{4-}$ which may be the reason for its observed co-variability with HCO_3^-. The most important feature of groups T1 to T3 (Figs. 5.4 and 5.5) is that this branch is enlarged and the relative size of the branch determines which of the three groups in which the sample belongs. The groups represent Ca-HCO_3 dominated waters but from three different environments. A

characteristic of T3 is that several of the other branches are also enlarged.

Branches B and C (Fig. 5.3) contain only trace metals. Branch B contains Al, Cd, Co, Ni and Zn, while Branch C contains Fe, Mn and Cu. Group T4 on Branch C is enlarged, mostly due to elevated Fe and Mn. Group T5 on Branches B and C is enlarged due to elevated levels of trace metals. The major difference between the two groups (although not visible from Fig. 5.3) is that waters in T5 are acidic to highly acidic, whereas most waters in T4 are near neutral. High SO_4^{2-} indicates that the characteristics of both groups may be due to contact with sulphides. T4 may be a result of mixing with T1 and T5 type waters.

Branch D contains the ions Sr^{2+}, K^+ and Cl^-. Na^+ is not included in this branch, as might be expected, because it contains Cl^-, for reasons discussed below. In T7, several of the branches are enlarged but the group's most important feature is the great enlargement of Branch D due to the high salt content. Branch D is somewhat enlarged in T3, also indicating a significant salt component in those waters.

Branch E, which contains the elements Mo, W, As, Si, Na^+ and F, is made up of two sub-branches, but as they vary so closely together they can be considered to be one. Group T6 results from increased concentrations of the elements in these two sub-branches which gives the tree a right skew. All the waters in T6 are Na-HCO_3 dominated and have a high pH resulting from dissolution of felsic intrusive rocks.

In Figure 5.6 (Section 5.4.3. Kleiner-Hartigan Trees Classification) and Table 5.2, a number of samples have been left ungrouped. Some of these are examples of surface waters (022, 024, 077, 117, 118, 124 and 128), which could have been left out. Two samples (10 and 36) had incomplete analyses and could not be included in the calculations. Sample 067 is water from the Prairie Creek Mine and had such unnaturally high concentrations of certain trace metals that it had to be removed from the calculations. Sample 026, although taken farther from the vent, is from the same spring as Sample 046, and therefore was not included. The rest of the unclassified samples (102, 114 and 126) are unusual waters which would not fit into the groups.

Similarities between the above groups and those of the physical grouping are apparent. They are discussed in text section 5.4.5.

Grouping by pattern recognition has a number of advantages. It allows one to see immediately what is causing the groupings. For instance, the differences in elemental associations of groups T2, T6 and T7 are obvious in Figures 5.4 & 5.5, but would not be obvious using methods such as cluster analysis. Another advantage is that when producing the trees, the process correlates and then clusters variables and not samples. This results in the use of far less computer time and memory than would be necessary for other computer methods that correlate and cluster samples.

A third advantage is the flexibility that results from the subjectivity of the method. During the selection of samples, one has the ability to use criteria which may be known to be important from other data. One can decide what will constitute a group and can use judgement as to what fits into that group based on factors such as temperature or spring precipitates. The latter point can be important in cases where mixing of water types is known or suspected. In this case, the sample can be included in the group which has the water type of greater interest to the investigator. A non-subjective computer method, such as cluster analysis, might place it into one of the two groups, but would be more likely to put it into neither as only half the criteria used to determine either group would have been met. Being able to decide what constitutes a group can also be important when the criteria for determining that group are minor relative to other components of the tree. Group T4 is an example of this because the branch containing Fe and Mn has been used as one criteria for choosing waters in that group.

Group T8 in Figure 5.6 is made up of waters which appear to have undergone cation exchange in shales or clay units. Na^+ from the rock has been exchanged for Ca^{2+} in the water. Only a few water samples were found to have been affected by this process and were recognised because they differ in shape from the other, larger groups. The differences are subtle enough that a method such as cluster analysis may not have picked them out and would have forced the waters into a group (probably one like T6) to which they are genetically different.

Subjectivity is a very important factor in this method. It can be argued that samples

could fit into more than one group and that some of the groups picked may have little geochemical significance. However, if enough true variation exists in the water types sampled (as in these waters) then this problem is greatly reduced. Cases where subjectivity can be a criticism are where samples have unusual chemistry or in those cases where dilution or mixing has taken place. A more important argument is that the investigator could establish groupings based on pre-conceived ideas or models. Admittedly this did play a part in establishing groups T4 and T8. They do not have very distinctive shapes, but they do represent important geochemical processes, which are not obvious by looking at the shape alone. It required subjectivity to create the groups and discern which waters have been involved in these processes.

The pattern recognition method invokes several other more technical problems. Different types of waters may have the same elements but have the elements vary in different ways. For example, Group T6 has Na^+ and HCO_3^- varying together but in T7, Na^+ and Cl^- vary together. Because the elements are assigned to a branch once, on the basis of the entire data set, the shape of the tree cannot properly represent the co-variability of elements for all cases. This may or may not prove to be a problem. It is a problem in T6 where the waters are Na-HCO_3 dominated and Na^+ is on one branch while HCO_3^- is on another. This also explains why Ca^{2+} and Mg^{2+} do not vary most closely with HCO_3^- and Na^+ does not vary with Cl^- as would be expected. Pattern recognition systems are also hampered by the difficulty of visually detecting differing shapes with increasing numbers.

Although the method was found to be satisfactory for the data herein, a non-subjective method would serve, at the very least, as verification. Cluster Analysis is presented as a non-subjective method in text Section 5.4.4 and Appendix 4.4