# Appendix B. Initial Site Classification

## Introduction

An initial site classification to identify discrete site types that distinguished variability in the reference diatom samples was attempted using taxa ordination, cluster analysis, and classification and regression tree (CART) analysis.

152 reference sites were identified and included in the analysis. One sample per site was used in the analyses. No duplicate or re-visit samples were used.

Operational Taxonomic Units (OTU) were established to identify common unique taxa that would be appropriately compared among samples. Of the 637 taxa identified in the reference samples, 199 OTUs were established. Taxa were either lumped into these OTUs or eliminated from the reference site classification analysis. OTUs are established to minimize loss of data. Species-level identifications were maintained as OTUs if most species in the genus were common (≥5 occurrences). Species were lumped to genus if most species were uncommon (<5 occurrences).

Natural environmental variables that were tested for associations with ordination axes, cluster analysis, and CART included those related to geology, location, climate, hydrology, and sample date. They included the following, among others: ecoregion / basin, latitude / longitude, slope, stream size, temperature / precipitation, background lithology, lithologic chemistry, and sampling month.

## Results

The NMS ordination diagram was explored to discover patterns in the arrangement of samples, attempting to associate natural environmental variables with the first three axes of the ordination. The diagram suggested some patterns with ecoregions, but there was considerable overlap in the sample taxa space (Figure B-1).



Figure B-1. NMS ordination of taxa presence in 152 reference samples with level 4 ecoregions identified per sample.

Cluster analysis suggested 4 groups of similar taxonomic composition among samples. These were associated with predicted background conductivity, baseflow index (BFI), maximum air temperature, watershed area, and latitude (which was also interpreted as ecoregion) (Figure B-2, Table B-1). CART analysis was used to identify thresholds for identifying the groups (Figure B-3).

Groups 2 and 14 were found to be similar to each other based on sample overlap in taxa and metric ordinations, similar reference metric box plots, and sensitive metrics with similar distributions. In addition, there were few (3) stressed sites in group 2, which would make index development in that group uncertain. Therefore, the two groups were tentatively combined.

The three resulting groups were considered as initial classes. After associating the groups with location and elevation they were renamed to simplify communication (Table B-2). The diatom index was not ultimately calibrated to these initial classes, but they were used to inform metric selection and for interpretation of the final index.

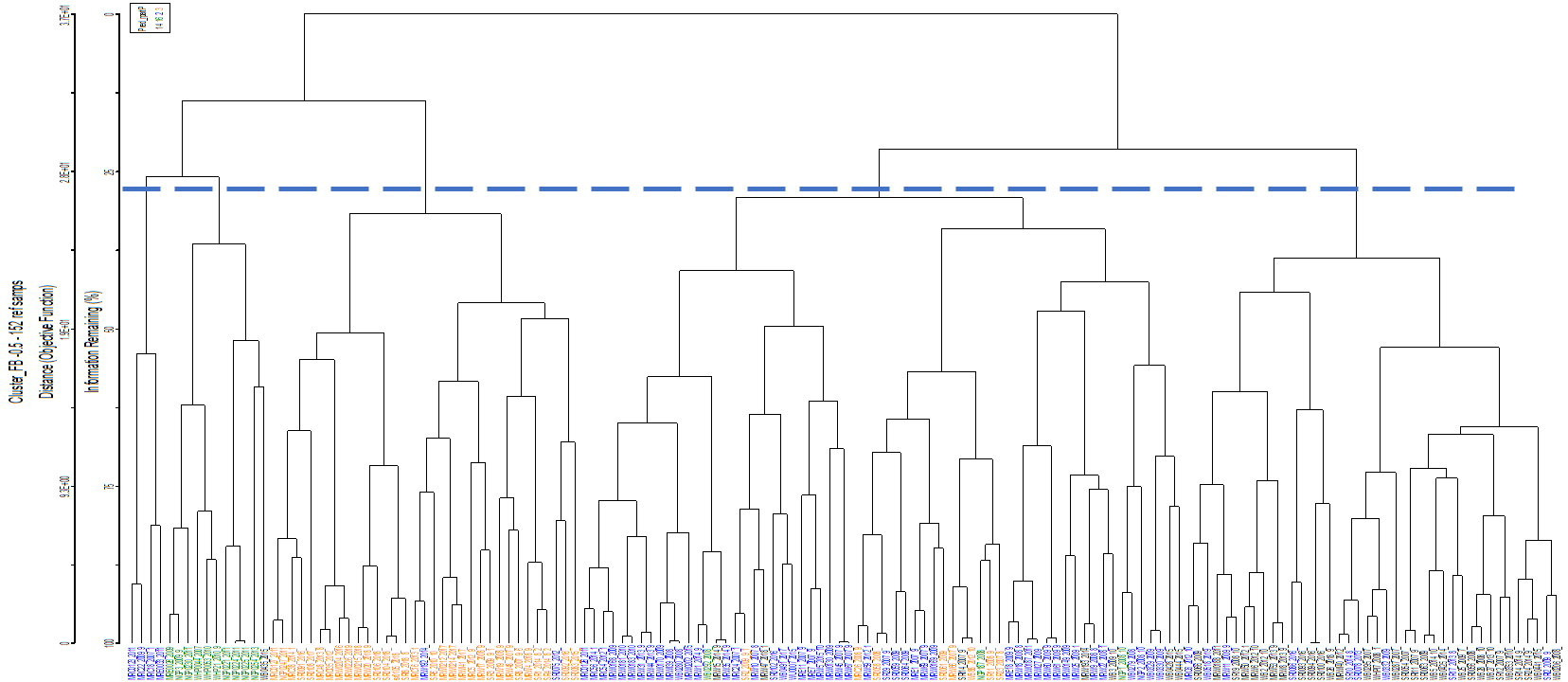


Figure 2. Cluster diagram of taxa similarity among samples, showing distinction of 4 groups.

Table B-2. Groups of taxonomically similar samples identified through cluster analysis and distinguished using CART analysis.

|  |  |
| --- | --- |
| Group Identifier | Group Description |
| Group 2 | (PredCond < 142 and BFIWs > 66) or (PredCond > 142, TmaxWs < 13, and Wsarea[StreamCat] < 73) or (PredCond > 142, TmaxWs < 13, WSarea(StreamCat) > 73, and Ecoregion 17 |
| Group 3 | PredCond < 142 and BFIWs < 66 |
| Group 14 | PredCond > 142, TmaxWs < 13, WSarea(StreamCat) > 73, and not Ecoregion 17 (MR) |
| Group 16 | PredCond > 142 and TmaxWs > 13 |

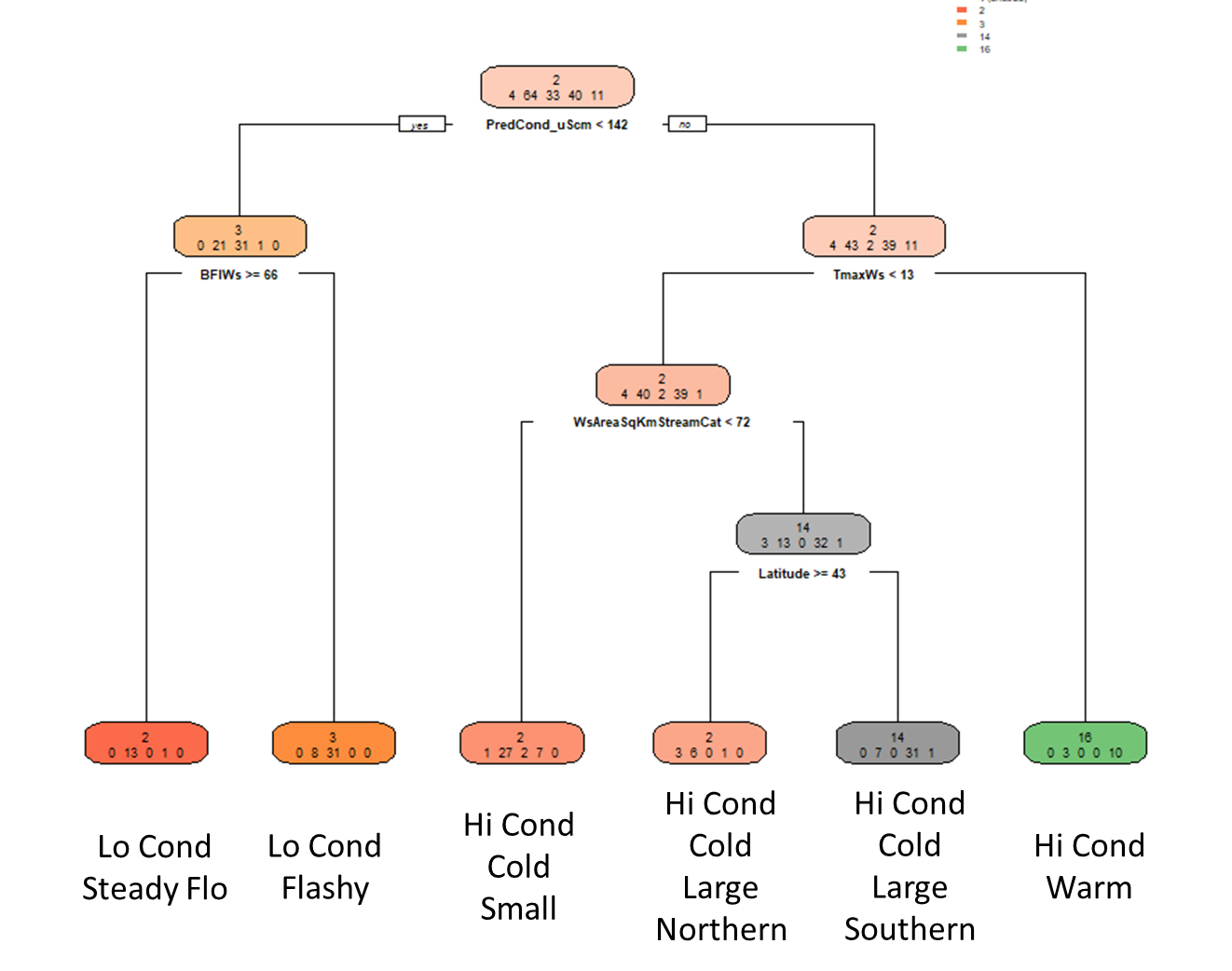


Figure B-3. CART dendrogram for distinguishing among groups of the cluster analysis.

Table B-2. Initial site classes to recognize discrete groups of taxonomically similar reference sites.

|  |  |
| --- | --- |
| Site Class | Description |
| High-Elevation Mountains  (Group 3) | Sites in the Middle Rockies or Southern Rockies ecoregions that are generally above an elevation of 7,500’. Sites generally have smaller drainage basins (<50km2) and less contribution of groundwater (BFI <66). |
| Mid-Elevation Mountains and Basins  (Groups 14 and 2) | Sites in the mountain or basins ecoregions that are generally below 7,500’ elevation. Sites could be in large or small drainage basins and greater contribution of groundwater (BFI >66) |
| Plains  (Group 16) | Sites in the plains ecoregions (High Plains or Northwestern Great Plains) and generally warmer temperature with higher specific conductance. |