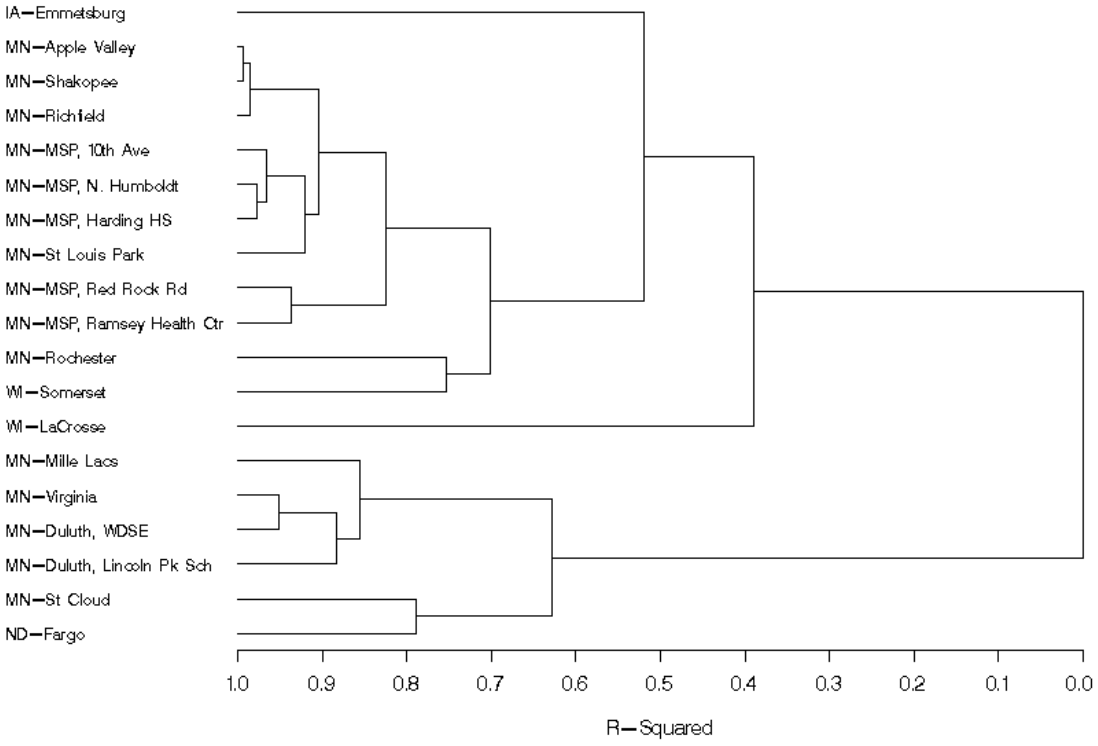
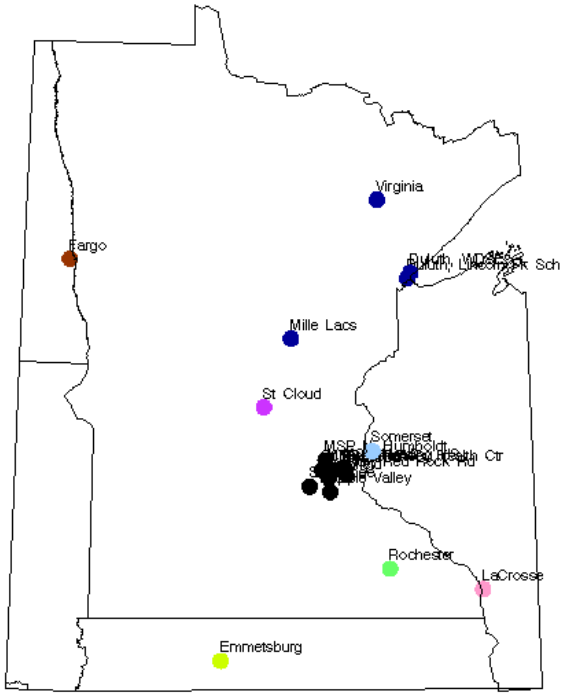
Cluster Analysis

A cluster analysis of the PM2.5 monitors was performed to examine which monitors group together, based on the 2006-2008 every 6th day FRM data. Clustering uses distances between objects (monitors) to build up clusters. Ward's method was used to link clusters. It minimizes the distance of any two hypothetical clusters that can be formed at each step, in a sort of reverse ANOVA. Distance in this case is the sum of squares $\{\Sigma(x_i - y_i)^2\}^{0.5}$ where x_i and y_i are concentrations measured at monitors x and y and the index i represents the daily measurements. Ward's method was chosen for its efficiency and because it produced clusters that conformed to known characteristics of the monitoring networks. That is, monitors tended to be clustered in a natural geographic pattern and monitors that were strongly influenced by local sources tended to cluster separately from other nearby monitors. The figures that follow show the clusters first as hierarchical trees. The length of each 'branch' is proportional to R-square. Without any clustering (at the far right edge), R square is zero because there are no clusters and no variability has been explained. Using the Minnesota tree below as an example, if the data are divided into 2 clusters (one with IA-Emmetsburg through WI-LaCrosse, and the second with MN-MilleLacs through ND-Fargo), about 40% of the variability is explained. Next, the first cluster can be broken down into 2 clusters, one consisting of just WI-LaCrosse and one with all the other monitors. This three-cluster tree accounts for about 52% of the variability. Each tree continues in that manner until the far left axis, which shows each monitor in a cluster by itself. The monitors with the shortest branches are those that provide the least additional explanation of variability in the data – similarly to a correlation analysis. The user can choose the degree of clustering that makes most sense for a particular purpose. The maps accompanying each state tree show the data clustered into eight groups of monitors in each state. These plots give a better sense of the geographic nature of this clustering and the occasional single-monitor cluster that represents a source-oriented monitor.

Minnesota Monitor Clusters

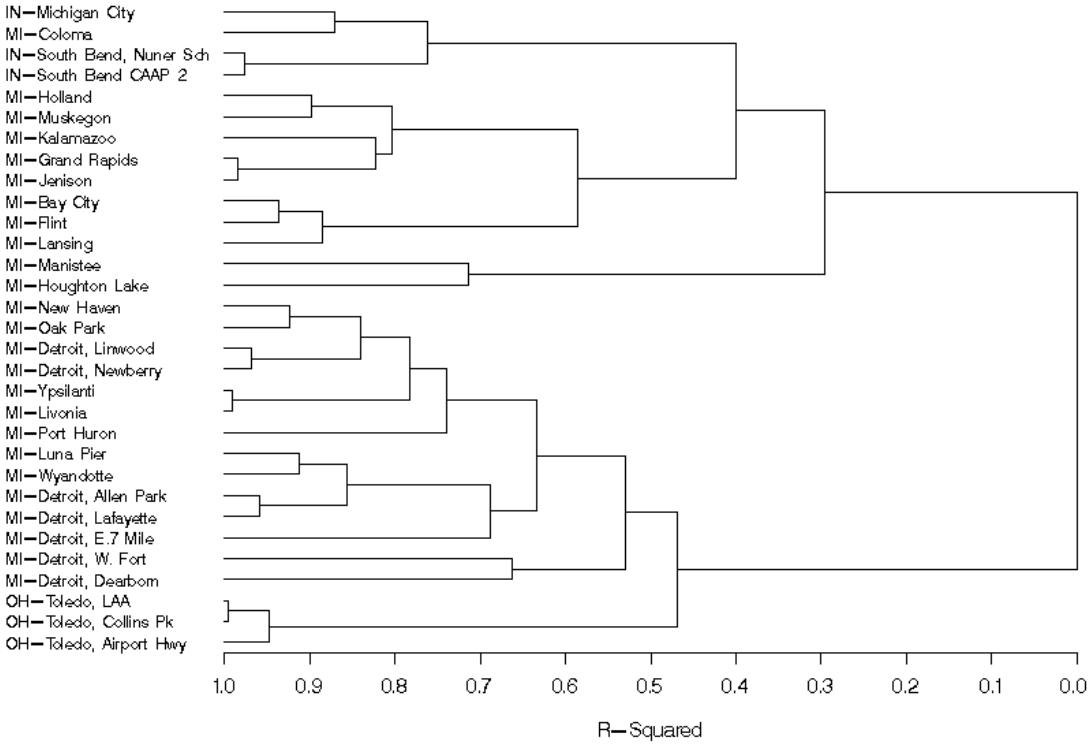


Minnesota Monitor Clusters

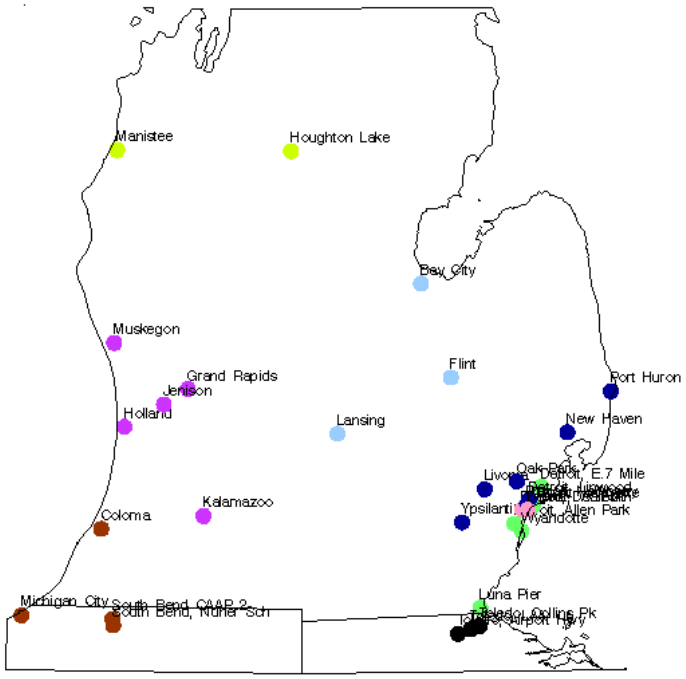


Each color indicates a cluster. Eight are shown.

Michigan Monitor Clusters

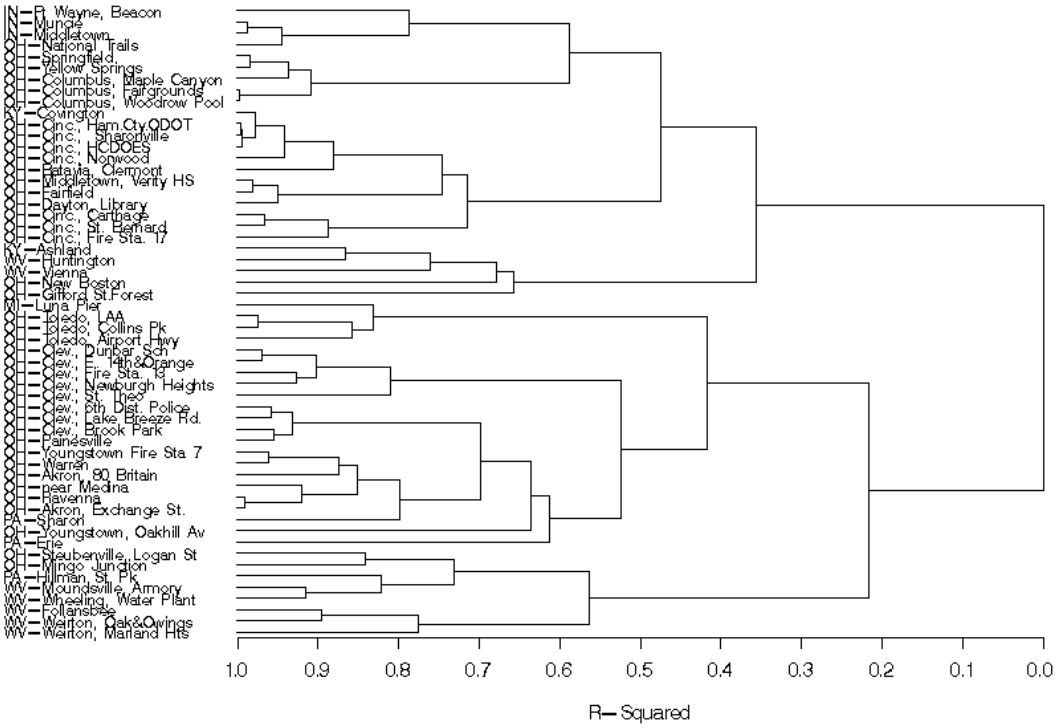


Michigan Monitor Clusters

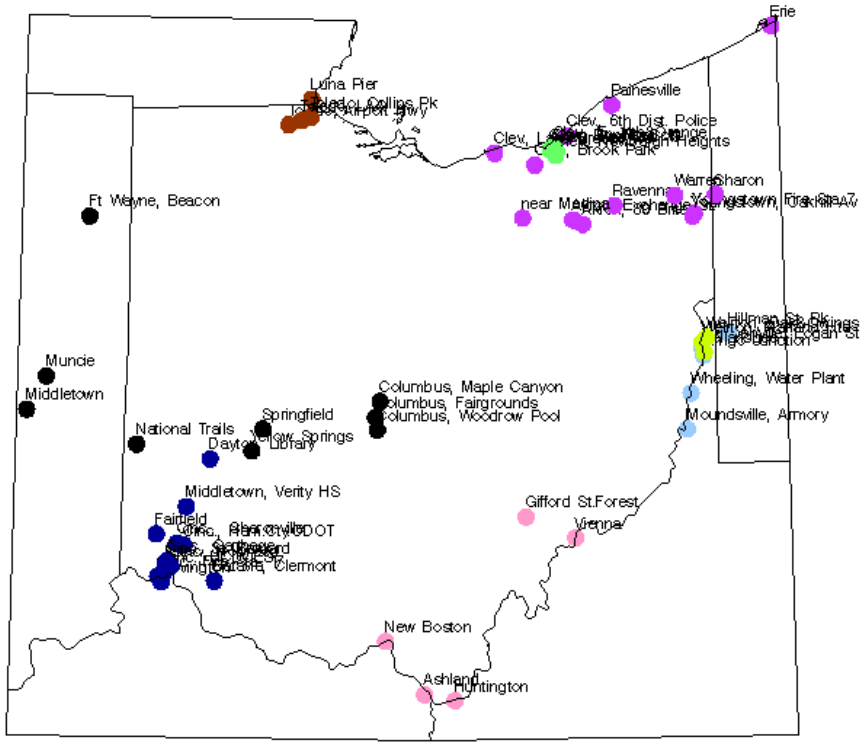


Each color indicates a cluster. Eight are shown.

Ohio Monitor Clusters

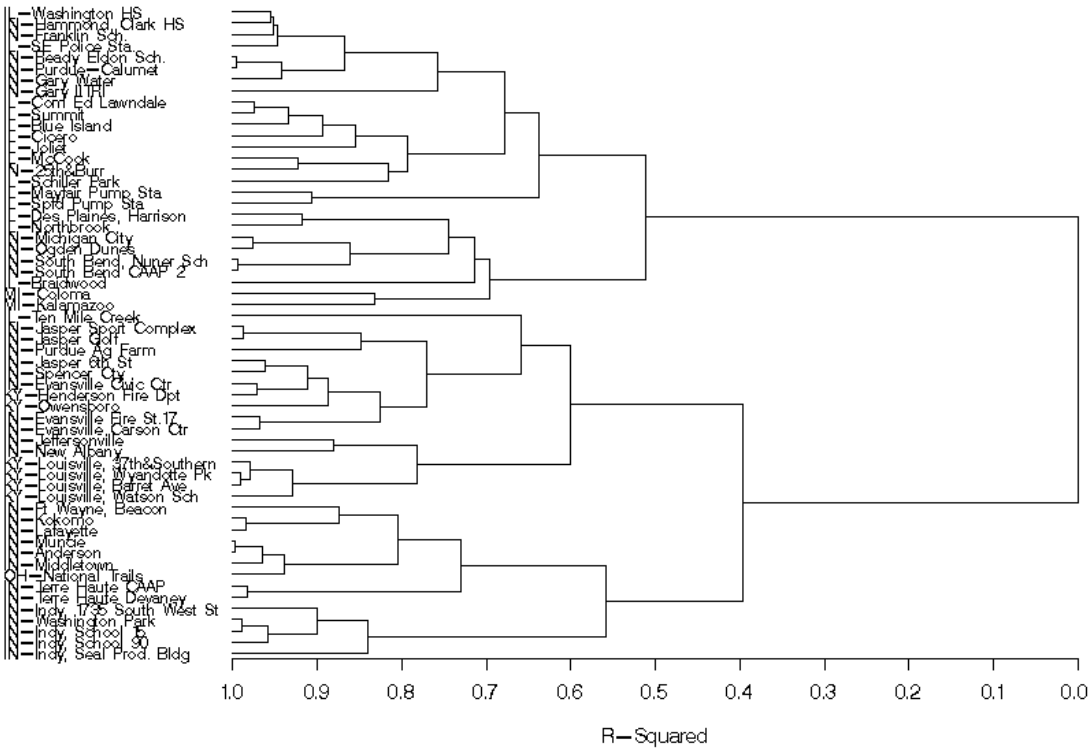


Ohio Monitor Clusters

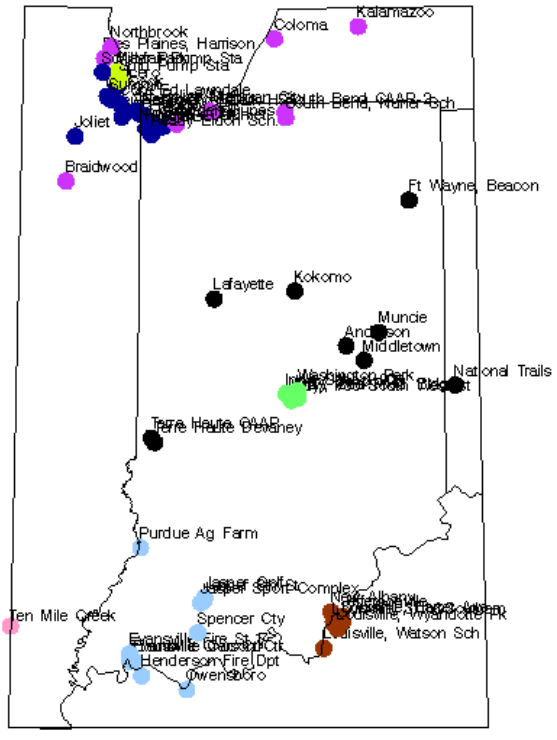


Each color indicates a cluster. Eight are shown.

Indiana Monitor Clusters

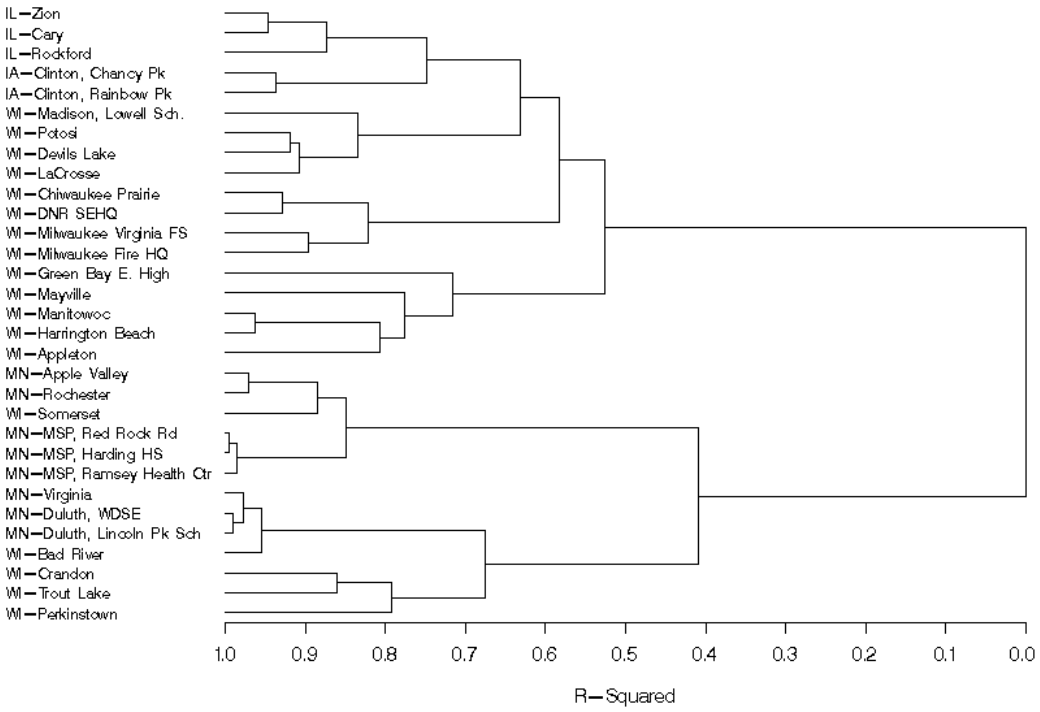


Indiana Monitor Clusters

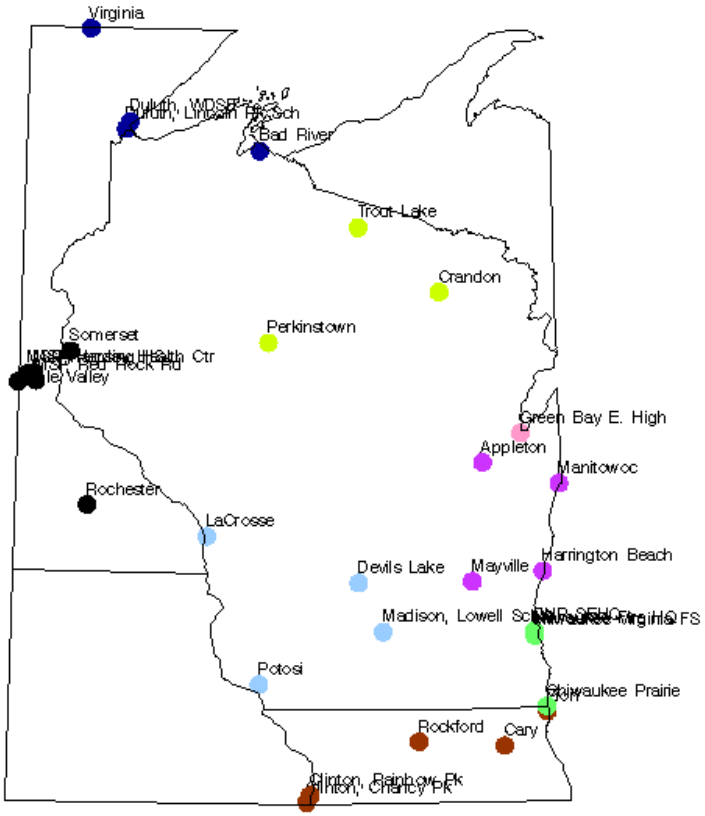


Each color indicates a cluster. Eight are shown.

Wisconsin Monitor Clusters

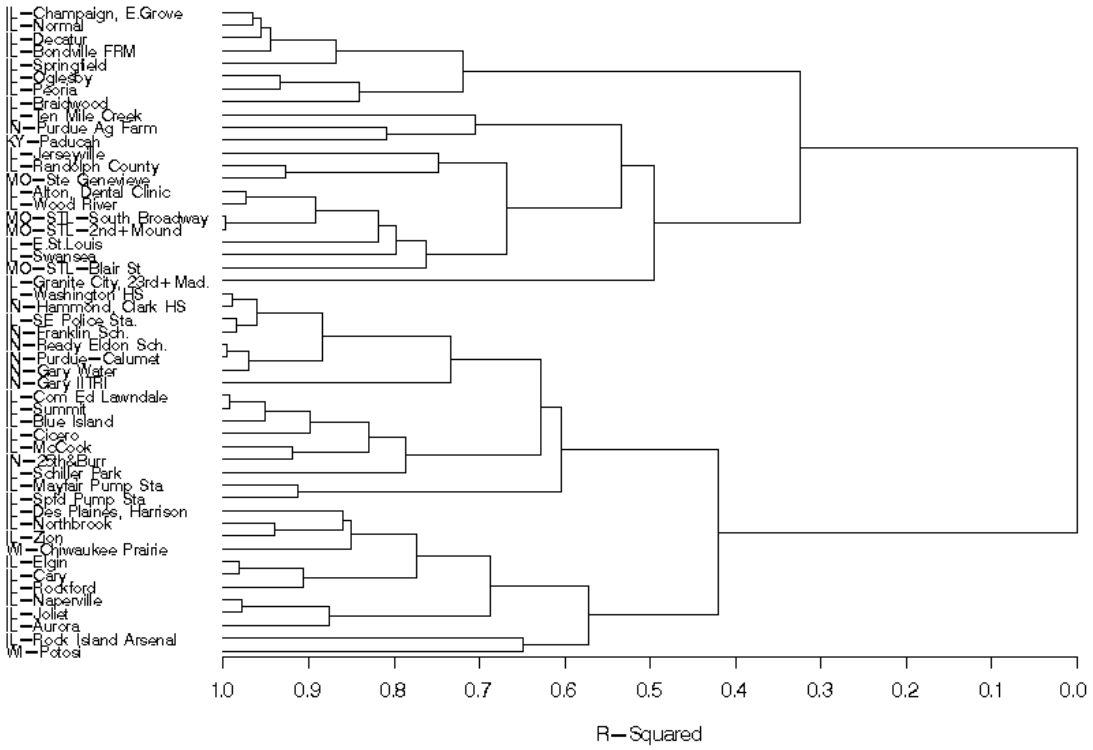


Wisconsin Monitor Clusters

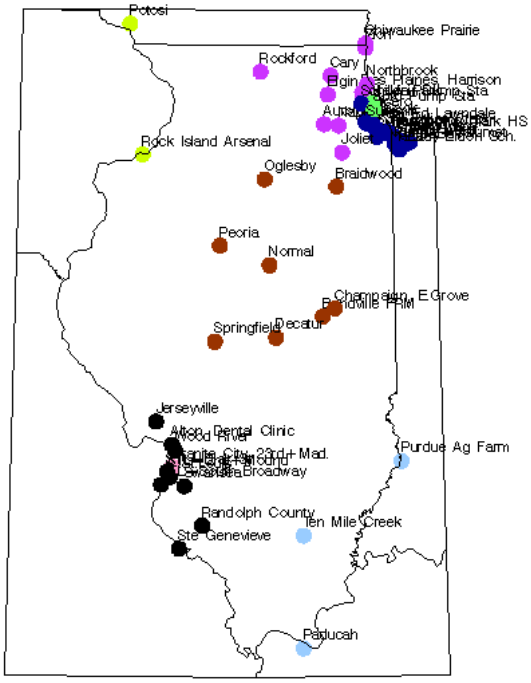


Each color indicates a cluster. Eight are shown.

Illinois Monitor Clusters



Illinois Monitor Clusters



Each color indicates a cluster. Eight are shown.