Using Clustering to Establish Climate Regimes from PCM Output

Forrest Hoffman*, Robert Oglesby**, William W. Hargrove*, David Erickson*

*Oak Ridge National Laboratory  **NASA Marshall Space Flight Center

Introduction
A new statistical clustering technique was used to analyze output from the Parallel Climate Model (PCM, Warren Washington). Five 100-year simulations of “business-as-usual” (BAU) scenarios were compared. This analysis considers three PCM output variables: surface temperature, precipitation, root-zone soil water. Only land was considered in the clustering analysis.

The copious output (about 1200 monthly maps per run) is difficult to decipher. The long-term climatic trend of interest is masked by the magnitude of the seasonal cycles and large inter-annual variability.

Multivariate Clustering
Multivariate clustering is the division or classification of objects into groups or categories based on the similarity of their properties.

Non-hierarchical clustering produces a single level of division of objects into some (often pre-determined) number of groups.

Multivariate Geographic Clustering (MGC) employs non-hierarchical statistical clustering to the classification of geographic areas.

Multivariate Spatio-Temporal Clustering (MSTC) is an application of Multivariate Geographic Clustering across space and through time.
On the left panel, each of the 5 BAU runs was analyzed individually and separately. On this panel, however, all 5 BAU runs were taken together and divided into a single, common set of climate regimes in the same clustering analysis. All 5 BAU runs are thus placed into a common set of climate regimes which permit the direct comparison of the individual runs within the same climate space.

Thus, a parallel set of results can be shown on this panel as on the opposite one, but these results are now shown in TERMS OF ALL BAU RUNS COMBINED FOR DIRECT COMPARISON.

**Conclusions**

Cluster analysis is a powerful data mining tool which provides a common basis for comparison across space and time for multiple climate simulation runs. Because it runs efficiently on a parallel supercomputer, it can be used to reveal long-term patterns in large multivariate data sets. The clustering technique statistically establishes a common and exhaustive set of occupied climate regimes within climate space. Cluster regimes are defined in terms of the original measurement units for each variable used.

The regimes defined by the clustering process were very similar, both in the analysis of individual runs and in a composite analysis of multiple runs taken together. Cluster analysis can be used to compare an entire ensemble of multiple runs of the same scenario (as done here), two ensembles of different scenarios, and even results from two different climate models.