Using Multivariate Spatio−Temporal Clustering to Establish Climate Regimes from Parallel Climate Model (PCM) Results

*Forrest M. Hoffman, William W. Hargrove, David J. Erickson, and Robert J. Oglesby**
*Oak Ridge National Laboratory and *NASA Marshall Space Flight Center*

Introduction

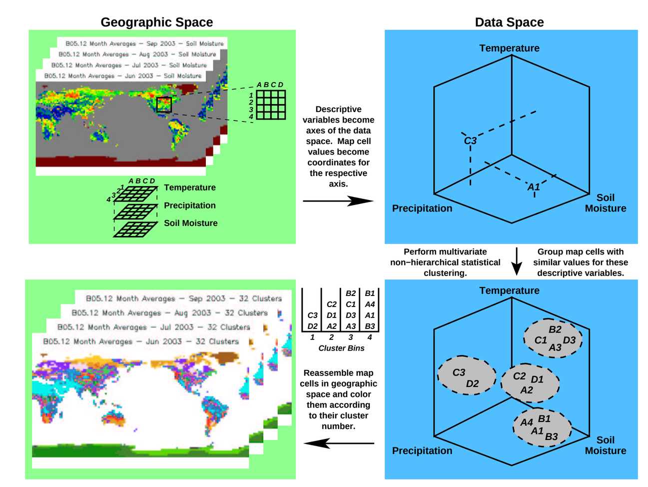
A statistical clustering technique was used to analyze output from the Parallel Climate Model (PCM) (Washington, et al.). Five 100−year ‘‘business as usual’’ scenario simulations were clustered individually and then in combination into 32 groups or climate regimes. Three PCM output fields were considered for this initial work: surface temperature, precipitation, and soil moisture (root zone soil water). Only land cells were considered in the analysis. The clustered climate regimes can be thought of as climate states in an N−dimensional phase or state space. These states provide a context for understanding the multivariate behavior of the climate system. This technique also makes it easy to see the long−term climatic trend in the copious output (about 1200 monthly maps per run) that is otherwise masked by the magnitude of the seasonal cycle. The same approach may be useful for comparing various model results with long time series observations to better understand cloud processes and climate feedbacks.

Multivariate Spatio−Temporal Clustering

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

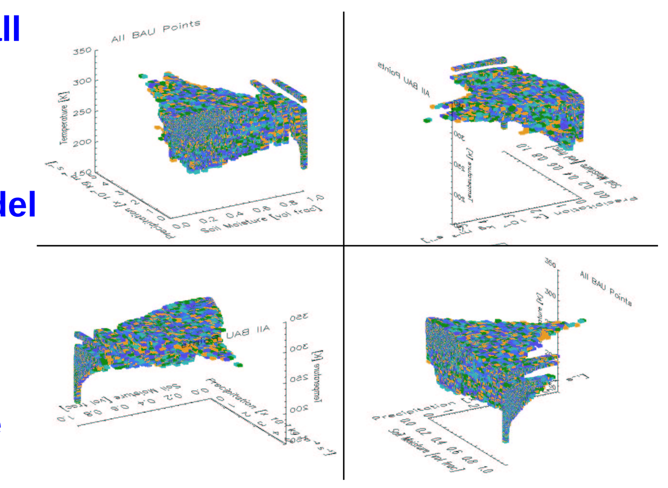
Non−hierarchical clustering produces a single level of division of objects into some specified number of groups.

Multivariate Geographic Clustering employs non−hierarchical clustering to the classification of geographic areas.



Multivariate Spatio−Temporal Clustering is an application of Multivariate Geographic Clustering across space and through time.

All BAU Points Plotted in a Climate State Space



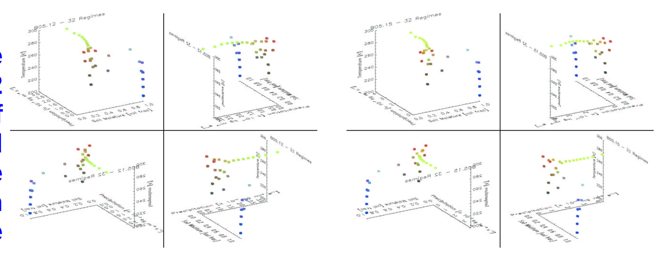
When every monthly data point from all 5 Business−As−Usual (BAU) runs is plotted in this three−dimensional climate phase space, we can see the portion of this space occupied by model predictions. In this phase space, we see that the majority of points (land grid cells) reside in a region of warm temperatures, low precipitation, and low soil moisture (near the front in the upper left frame). Discrete values of high soil moisture (in polar and tropical regions) result in planes of points. Points are colored by BAU model run, and the manifolds formed by each run overlap since the same model was used for each run.

Independent Cluster Analysis

B05.12 B05.15

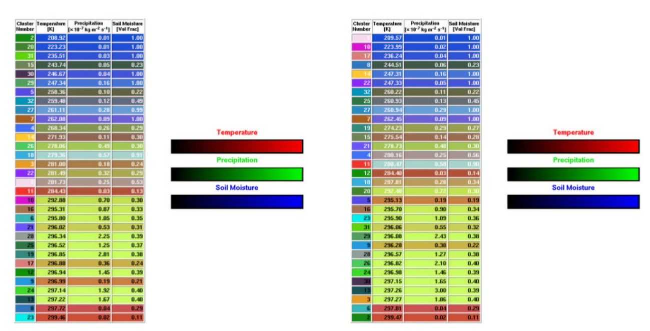
Clustered Climate Regimes

The clustering process establishes an exhaustive set of occupied climate regimes (i.e., the 32 cluster centroids) which define the subset of phase space occupied by the simulated atmosphere/land surface at all points in space and time. Any geographic location will exist in only one of these climate regions at any single point in time.

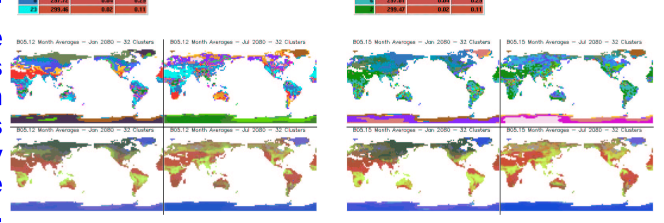


Climate Regime Definitions & Maps

The centroid coordinates of each of the clusters represent the synoptic conditions of that climate regime in the original measurement units. The first column of the table shows the random colors for each regime used in the top row of maps below. The remaining columns are shown in similarity colors, where each of the 3 variables contributes a red, green, or blue component.

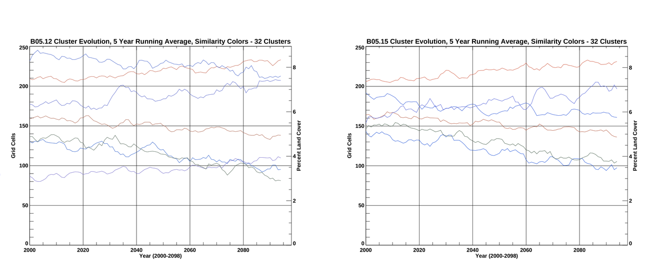


The top row of maps is colored randomly while bottom row depicts the same climate regimes colored using similarity colors. The first column of maps is January 2080; the second column is July 2080. Randomly colored maps differ widely across the independently−analyzed runs, but the similarity colored maps appear very similar.



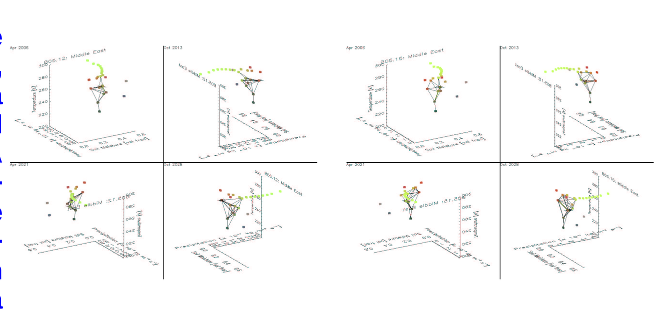
Regime Area Changes

Because the same clustered sets of conditions are identified through time, we can plot changes in geographic area of the globe in a particular climate regime as it evolves. Many of the 32 regimes remain relatively constant in area throughout the model run. These constant regimes are not shown; only climate regimes experiencing large global area changes in each run are shown.



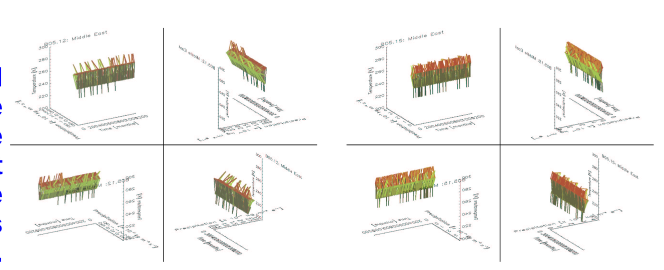
Climate Trajectories

A geographic location exists in only one climate regime at any point in time. By incrementing time, any single geographic location will trace out a trajectory or orbit among successively occupied climate regimes in climate phase space. A ‘‘spider’’ representing the simulated atmsophere−land surface sequentially moves among the climate regimes leaving a thickening ‘‘web’’ out−lining the trajectory. When a geographic location adopts a regime it never previously occupied, a climatic change has occurred at that location.



Climate Manifolds

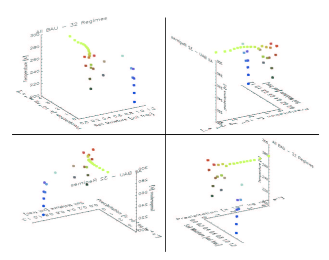
Tracing out the entire seasonal and annual trajectory for a single location yields a climate ‘‘manifold’’ in state space representing the shape of the predicted climate occupancy for that location. The predicted climate extremes and the frequencies of occupation are easily seen in this representation.



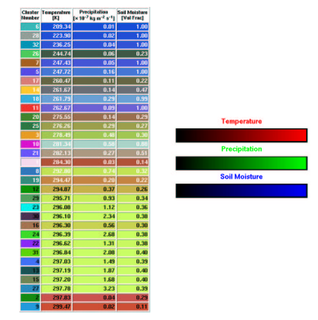
Combined Cluster Analysis

B05.12 B05.15 B05.18 B06.06 B06.09

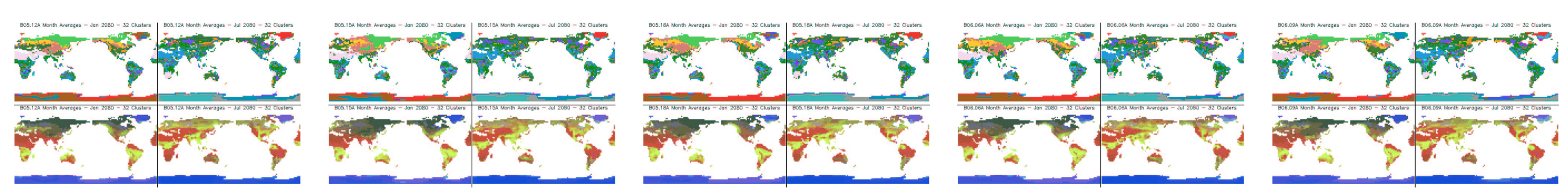
All 5 BAU Runs Clustered into a Single Common Set of Climate Regimes for Direct Comparison



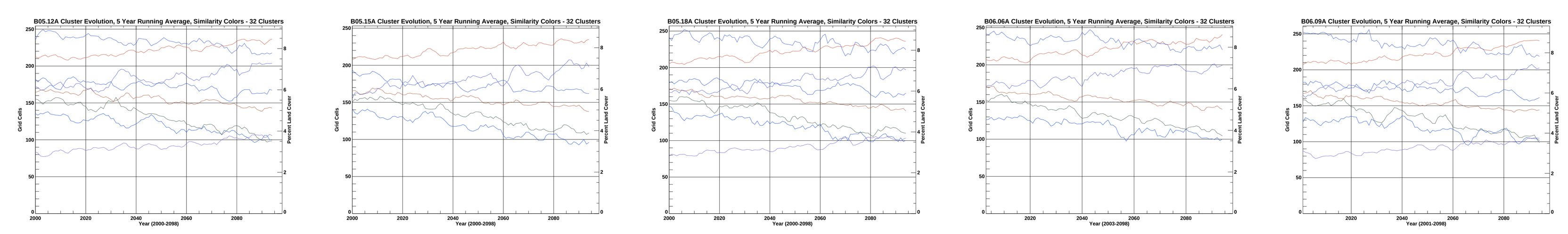
These 32 cluster centroids are a new set of climate regimes resulting from the cluster analysis of the output from all 5 BAU runs taken together. The visualizations below are in terms of this common state space.



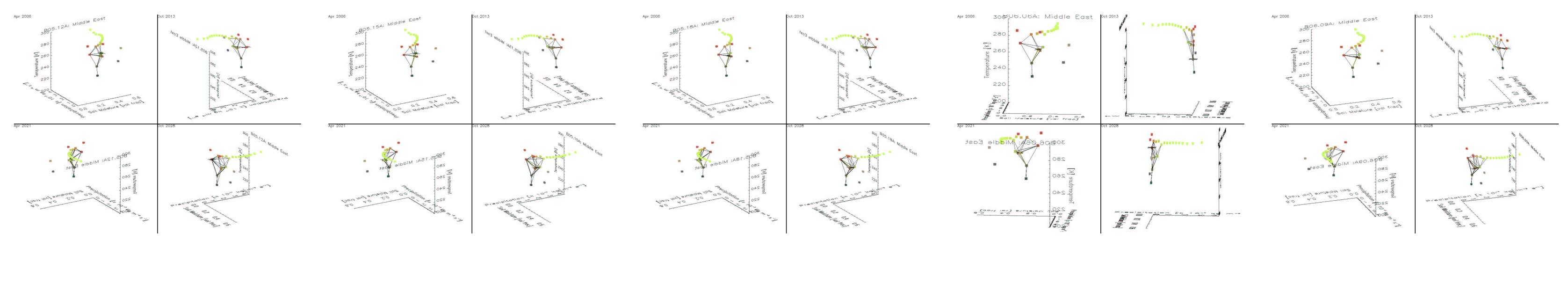
The 32 centroid coordinates represent the synoptic conditions for the 32 new climate regimes. Again, the first column shows the random color associated with each regime while the remaining columns show the similarity color and the mean temperature, precipitation, and soil moisture for each regime.



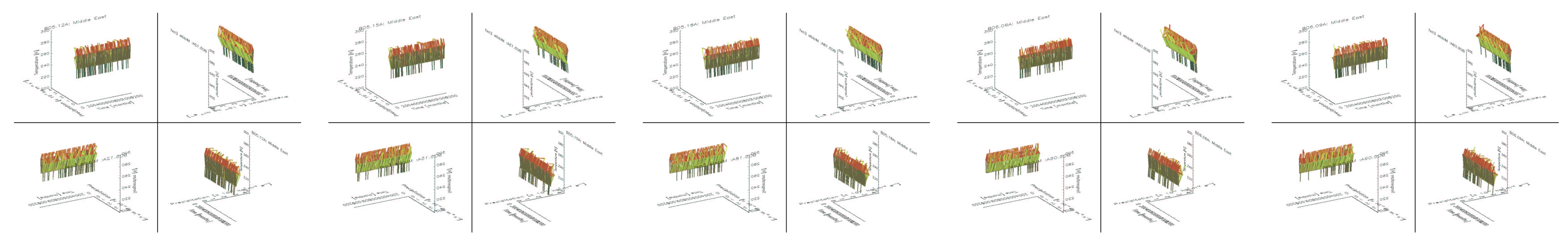
Now the randomly−colored maps in the top row appear very similar across all 5 BAU model runs, and maps from different BAUs may now be directly intercompared. The similarity−colored maps now use the exact same similarity colors. Any differences between BAU maps are due to differences in model predictions.



Now it is easy to identify the same clustered climate regime evolution curve across the graphs for each BAU model run. Differences in global area changes and curve inflections for the same climate regime across BAUs are due to differences in model predictions. All 5 BAUs indicate a growth in the hottest, driest (desert) climatic regime, and decreases in area of the coldest Arctic and Antarctic regimes. These changes are consistent with increased desertification and a general warming of the polar regions.

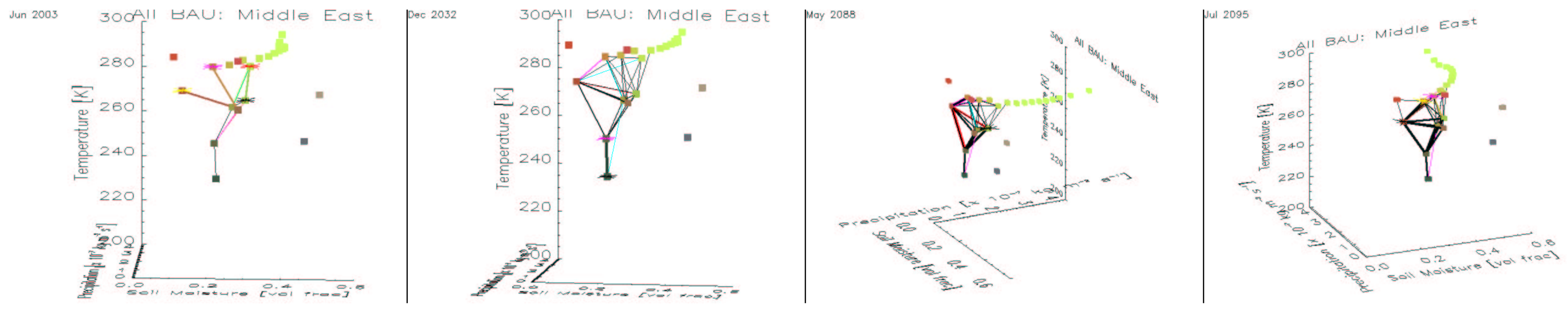


These plots now show each BAU trajectory for a single location in the Middle East in terms of a common set of climate states. Therefore, the ‘‘spider’’ representing the geographic location of interest spins a ‘‘web’’ among the same climate states or regimes. The only differences between webs is due to differences in model predictions. Because output from BAU runs start at different times, some plots are shown at different angles.



Since trajectories are now drawn in a common climate state space, the resulting climate manifolds for a single geographic location may be directly compared. Frequency of visitation for local extremes are easily seen around the edges of the manifold. For this location in the Middle East, all 5 BAU runs show a decrease in visitation frequency for the bottommost regime representing a very cold winter condition. In addition, most of the runs predict that this location will enter a warmer, drier (more arid) desert climate state that it had never previously experienced after 2050. Moreover, visitation of this state increases as the runs near the end of the prediction period.

Five Climate Trajectories in a Common Climate State Space



Now that a common set of clustered states has been obtained, the climate trajectories for a single geographic location can be shown as 5 different ‘‘spiders’’ (one for each BAU run) traversing a single shared set of climate states. Here, each spider, representing a single BAU, has a different color. When two spiders occupy the same climate regime, the overlapping spiders are colored black.

Trajectories are drawn with the similarity color of the climate regime to which spider has just moved, but the links subsequently change to the color of the spider that traversed them most frequently. Line segments between states become thicker with repeated traversal.

The multiple spiders are often co−incident on the same climate state or regime in January and July, the climatic extremes of the year, but spread out across multiple states in spring and fall ‘‘transitional’’ months. Spiders often appear on opposite sides of the diamond−shaped seasonal orbit in both the soil moisture and the precipitation planes, but rejoin at the top and bottom of the diamonds in the summer and winter months. Thus, the BAU run predictions are similar with regard to temperature, but tend to be more variable with respect to soil moisture and precipitation. This variability seems to increase to some degree as the simulation progresses.

Conclusions

Cluster analysis is a powerful tool which can provide a common basis for comparison across space and through time for multiple climate simulations. Because it runs efficiently on a parallel supercomputer, the tool can be used to reveal long−term patterns in very large multivariate data sets. Given an array of equally−sampled variables, the technique statistically establishes a common and exhaustive set of approximately equal−variance regimes or states in an N−dimensional phase (or state) space. These states are defined in terms of their original measurement units for every variable considered in the analysis.

Clustering may be used not only to analyze and intercompare climate simulations, but also to analyze observations and intercompare them with model results. The area change graphs above could show trends in cloud and climate states from long time series measurements. The trajectory figures could show multivariate cloud behavior. When measurements are clustered in combination with model results, two trajectories could be seen to diverge when models and measurements diverge and converge when models and measurements agree. By analyzing long time series measurements with model or reanalysis results, the manifold figures could show the occupancy by a single site in a ‘‘full’’ cloud/climate phase space yielding insights into the representativeness of individual observation sites or an entire observation network.