Multivariate Spatio−Temporal Clustering of Time−Series Data:
A Method for Diagnosing Cloud Properties and Understanding ARM Site Representativeness

*Forrest M. Hoffman, William W. Hargrove, A. D. Del Genio**
*Oak Ridge National Laboratory and *NASA Goddard Institute for Space Studies*

Introduction

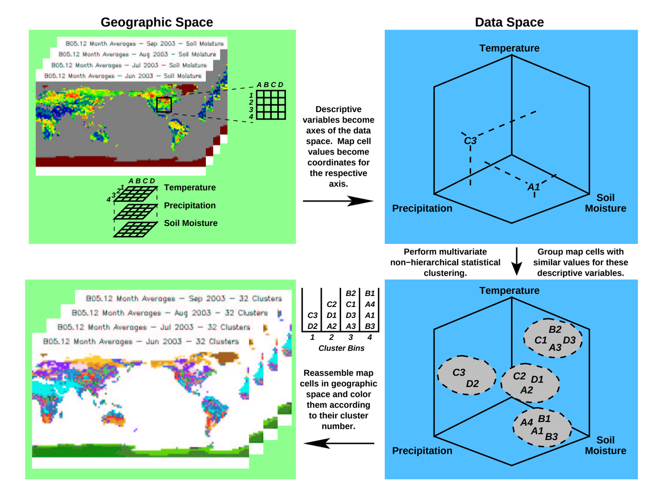
A statistical clustering technique was used to analyze output from the Parallel Climate Model (PCM) (Washington, et al.). Five 100−year ''business as usual'' scenario simulations were clustered individually and then in combination into 32 groups or climate regimes. Three PCM output fields were considered 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 climate 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.

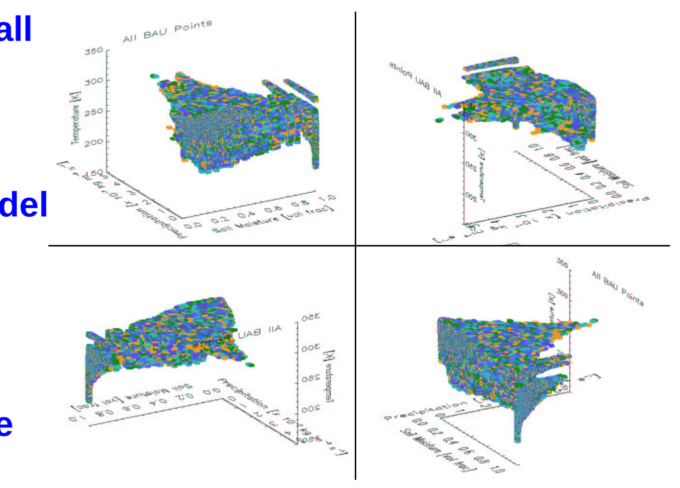
Multivariate Geographic Clustering employs non−hierarchical clustering to the classification of geographic areas.



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.

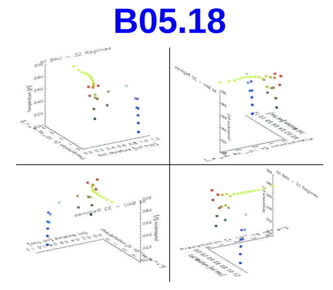


Ensemble Cluster Analysis

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

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 regimes at any single point in time.



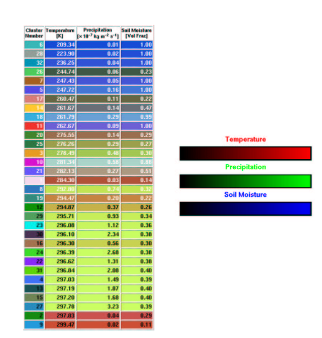
These 32 cluster centroids are a new set of climate regimes resulting from the cluster analysis of the output from all 5 BAU runs taken together. The visualizations below are in terms of this common state space.

A Single Common Set of Climate Regimes for Direct Comparison

A Single Common Set of Climate Regimes for Direct Comparison

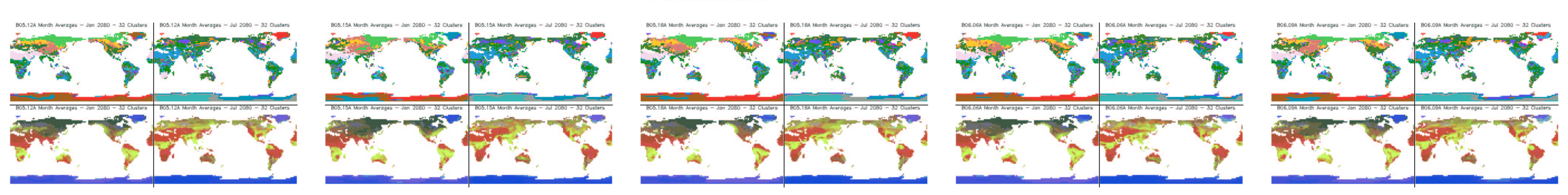
Climate Regime Definitions & Maps

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 32 centroid coordinates represent the synoptic conditions for the 32 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.

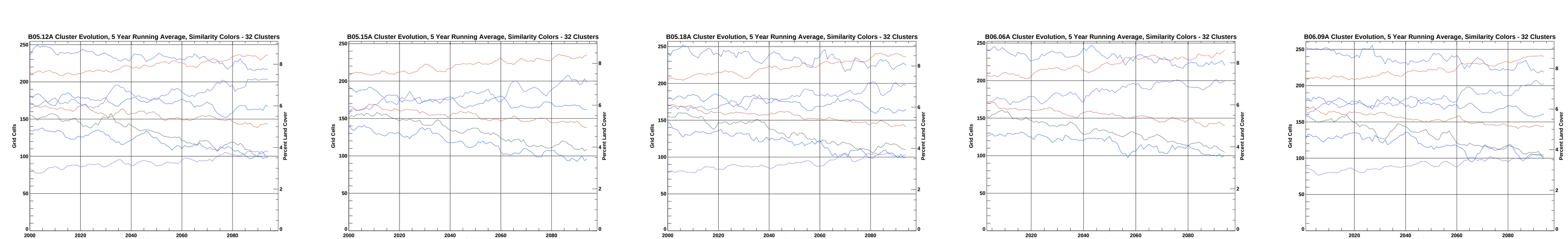
The top row of maps is colored randomly while the bottom row depicts the same climate regimes colored using similarity colors. The first column of maps is January 2080; the second column is July 2080.



The maps appear very similar across all 5 BAU runs. Any differences between the January maps or between the July maps are due to climate variability in the model predictions.

Regime Area Changes

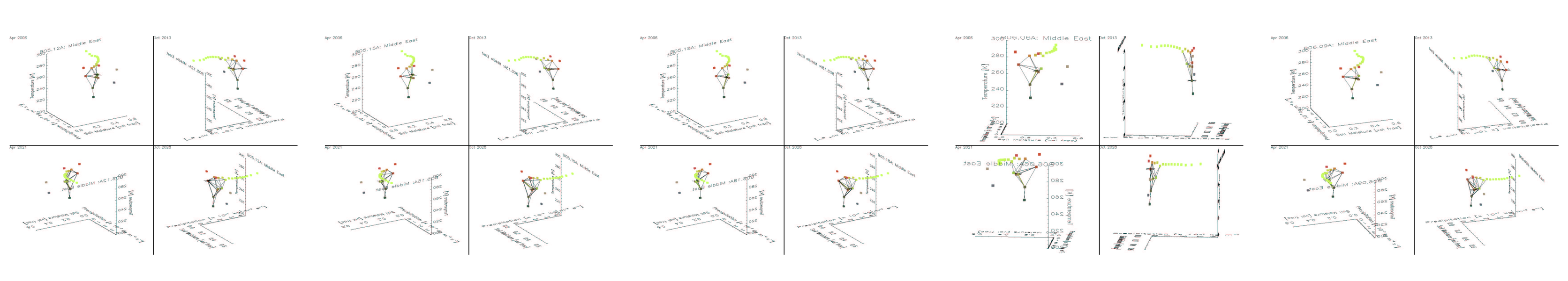
Because the same clustered sets of conditions are identified through time, we can plot changes in geographic area globally for any climate regime as it evolves. Many of the 32 regimes remain relatively constant in area throughout each model run. These constant regimes are not shown; only climate regimes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



From these graphs it is easy to identify which climte regimes experience large changes in area for each BAU run. Differences in the curves across runs is due to predicted variability. All 5 BAUs indicate a growth in the hottest, driest (desert) regime, and a shrinking of the coldest Arctic and Antarctic regimes. These changes are consistent with increased desertification and a general warming in polar regions.

Climate Trajectories

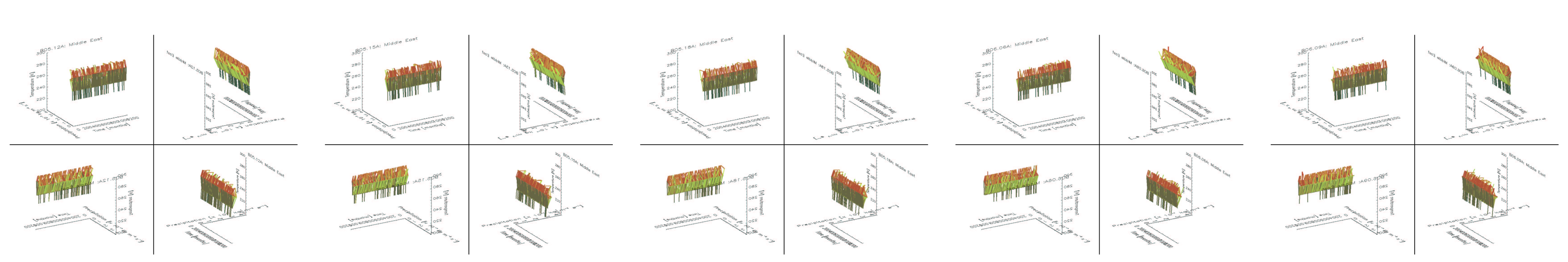
A geographic location exists in only one climate regime at any point in time. By incrementing time, any single geographic location will trace out a trajectory or orbit among successively occupied climate regimes in the climate phase space. A ''spider'' representing the simulated atmsophere−land surface 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 for that location.



These plots show the predicted climate trajectories for a single location in the Middle East from each BAU run in terms of the common set of climate states. The ''spider'' representing the Middle East location spins a ''web'' among the climate states or regimes. Differences across runs is due to predicted variability. Because output from runs start at different times, some plots are shown at different angles.

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 graphical representation.



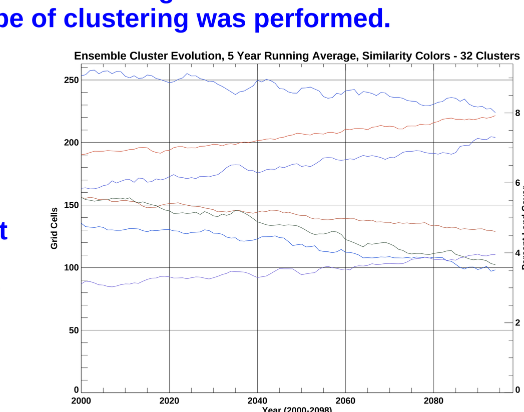
Frequency of visitation for extremes are easily seen around the edges of the manifold. For this location, all 5 BAUs predict a decrease in visitation frequency of the bottommost regime representing very cold winter conditions. In addition, two of the runs predict significant−enough warming and drying to push this location into the hottest and driest regime near the end of the simulations.

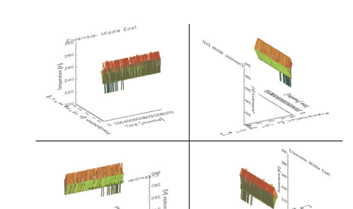
Ensemble Average Cluster Analysis

An Ensemble Average time series was generated using all 5 BAU model runs by averaging all runs at each time interval for each grid cell. To make the analysis of this single time series comparable to the Ensemble Analysis results, a special type of clustering was performed.

A One−pass Clustering was used to classify the Ensemble Average time series into the single common set of climate regimes already defined. Once classified, the Ensemble Average results were analyzed and displayed just like the time series from the individual runs.

The Ensemble Average Regime Area Change graph at right shows the climate regimes which undergo a significant global area change. These curves are directly comparable to the individual Regime Area Change graphs for the individual BAU runs shown at left because they are in terms of the same single common set of climate states.

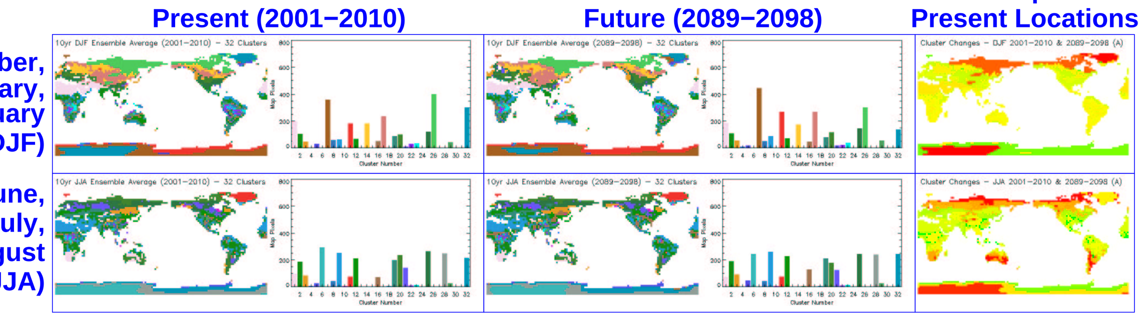




Tracing out the entire seasonal and annual trajectory from the Ensemble Average time series for the usual location in the Middle East, we see that averaging the model results reduces the frequency of visitation to extreme climate states. Because of the predicted climate change and the variability among the runs, the very cold winter state is never visited by the Ensemble Average after about 25 years even though individual runs predict occasional visitation. Moreover, the desertification predicted by some ensemble members is not strong enough to push the Ensemble Average into this desert climate state.

Ten year time interval averages for the present (2001–2010) and the future (2089–2098) for two seasons were created from the Ensemble Average time series. These four snapshots were then similarly classified using a one−pass clustering in conjunction with the previously−defined climate states. The resulting maps and regime histograms show where regime change has occurred and which regimes experience significant area changes. Stop−light color difference maps show which climate regimes shrank from the present to the future (red), which regimes stayed the same size globally (yellow), and which grew (green). The difference maps show the location of the affected climate regimes with respect to present predicted conditions.

	Present (2001–2010)	Future (2089–2098)	Difference with respect to Present Locations
December, January, February (DJF)			
June, July, August (JJA)			

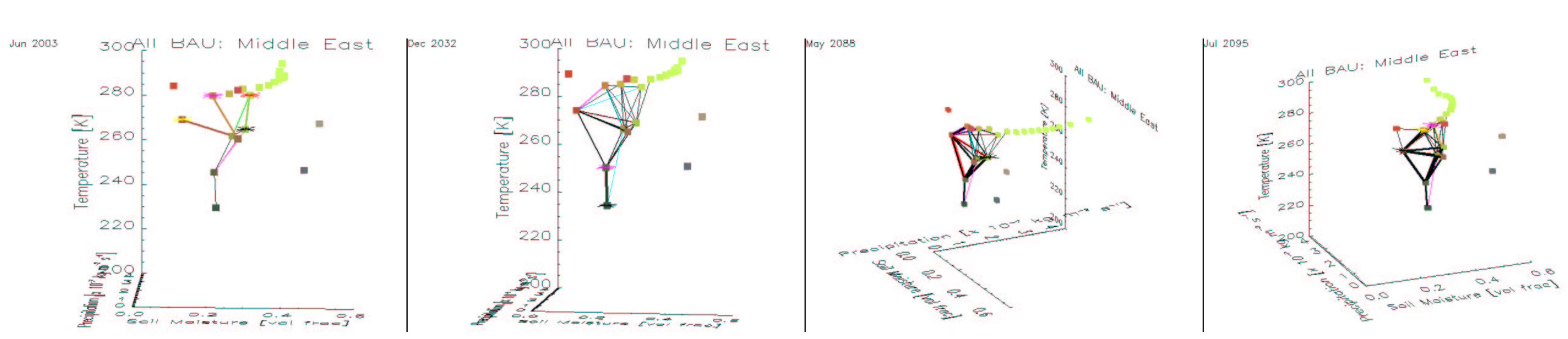


Clustering Methods for Comparison of Time Series Data

Clustering Method	Single Time Series	Multiple Time Series (Ensemble)	Ensemble Average Time Series
Normal Clustering	normal classification and single time series centroids	ensemble classifications and ensemble centroids	normal classification and ensemble average centroids
One−pass Clustering with single time series centroids	normal classification*×	classifications comparable to single time series classification✚	classification comparable to single time series classification✚
One−pass Clustering with ensemble centroids	classification comparable to ensemble classifications×✚	ensemble classifications*	classification comparable to ensemble classifications✚
One−pass Clustering with ensemble average centroids	classification comparable to ensemble average classification ×✚	classifications comparable to ensemble average classification ✚	normal classification*

*Obtained automatically from normal clustering (first row)
×Contained at right if the ensemble contains the single time series
✚Data normalization requires that the input data be transformed to the phase space of the data used to generate the centroids

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 ARM site in a ''full'' cloud/climate phase space yielding insights into the representativeness of individual observation sites or the entire ARM observation network.