From Measurements to Models:

Cross−Comparison of Measured and Simulated Behavioral States of the Atmosphere

Forrest M. Hoffman, William W. Hargrove, A. D. Del Genio, Jasna Pittman***

*Oak Ridge National Laboratory and *NASA Goddard Institute for Space Studies, **NASA Marshall Space Flight Center*

Introduction

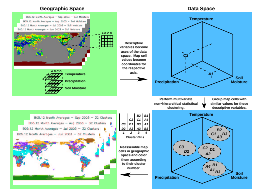
A statistical clustering technique was used to analyze output from the Parallel Climate Model (PCM) (Washington, et al.). Five 100−year ''business as usual'' scenario simulations were clustered individually and then in combination into 32 groups or climate regimes. Three PCM output fields were considered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

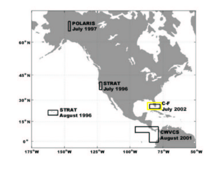
Non−hierarchical clustering produces a single level of division of objects into some specified number of groups.

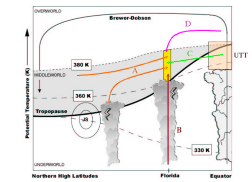
Multivariate Geographic Clustering employs non−hierarchical clustering to the classification of geographic areas.



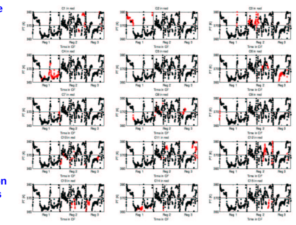
Multivariate Spatio−Temporal Clustering is an application of Multivariate Geographic Clustering across space and through time.

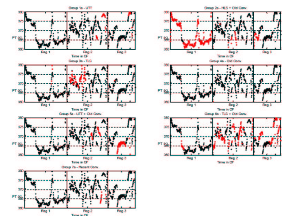
Clustering CRYSTAL−FACE Measurements into Atmospheric Regimes





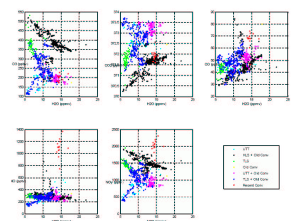
Multivariate clustering was used to infer atmospheric transport mechanisms from the combination of six trace gases in samples taken in the Crystal−FACE experiment performed over southern Florida in July 2002. The concentration mixtures of O_3, CO_2, CO, NO, NO_y, and H_2O vapor were used as "fingerprints" to identify sources of sampled air masses using multivariate clustering to delineate 15 concentration signature groupings. Shown at the right are 15 altitude x time maps, one for each tracer signature group, with samples assigned to that group highlighted in red. Although the tracer groupings were based on concentration data alone, the cluster groups segregate well into discrete regions of the spatio−temporal map, possibly suggesting discrete functional modes.





Four prospective atmospheric transport mechanisms within the "middleworld" troposphere were hypothesized: (A) High Latitude Stratosphere (HLS), (B) Old Convective (convective air with low NO lost to 1−week old atmospheric chemistry), (C) Upper Tropical Troposphere (UTT), and (D) Tropical Lower Stratosphere (TLS). When the 15 tracer concentration clusters were remapped to the 4 prospective atmospheric transport mechanisms (and 3 combinations of those transport mechanisms) on the basis of the tracer fingerprint concentrations, the 7 new altitude x time maps show even more discretization according to the new hypothesized functional assignments.

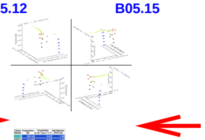
In these five bivariate plots of tracer pairs, points are color−coded by their ostensible atmospheric transport mechanisms. The functionally coded clusters are well−segregated in tracer concentration space. Such separation is consistent with an hypothesis that these air masses resulted from different discrete atmospheric transport mechanisms.



Atmospheric regimes formed by clustering formed discrete intervals in the time domain, suggesting different aspects of atmospheric function. Groups formed by clustering multivariate atmospheric measurements mapped reasonably well to atmospheric regimes associated with different hypothesized atmospheric transport functions.

Ensemble Cluster Analysis

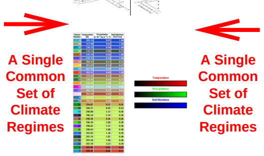
B05.12 B05.15



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 regimes at any single point in time.

A Single Common Set of Climate Regimes

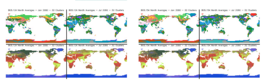


A Single Common Set of Climate Regimes

Climate Regime Definitions & Maps

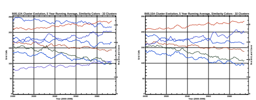
The centroid coordinates of each of the clusters represent the synoptic conditions of that climate regime in the original measurement units. The first column of the table shows the random colors for each regime used in the top row of maps below. The remaining columns are shown in similarity colors, where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

The top row of maps is colored randomly while 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.



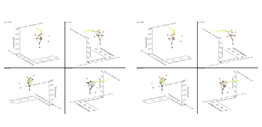
Regime Area Changes

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.



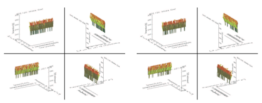
Climate Trajectories

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 atmosphere−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.



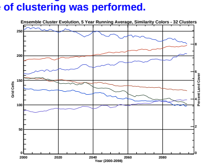
Climate Manifolds

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.

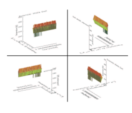


Ensemble Average Cluster Analysis

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.

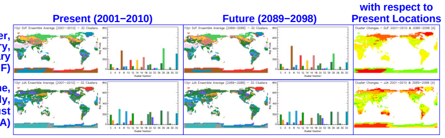


The Ensemble Average Regime Area Change graph at right shows the climate regimes which undergo a significant global area change. These curves are directly comparable to the individual Regime Area Change graphs for the individual BAU runs shown at left because they are in terms of the same single common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

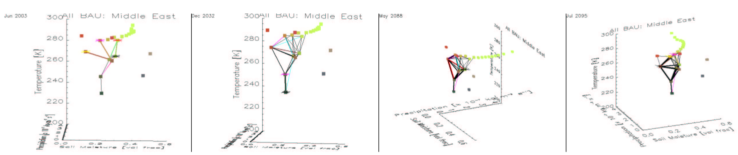
Ten year time interval averages for the present (2001−2010) and the future (2089−2098) for two seasons were created from the Ensemble Average time series. These four snapshots were then similarly classified using a one−pass clustering in conjunction with the previously−defined climate states. The resulting 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.



All of these results, along with maps and animations, are available in:

Hoffman, Forrest M., William W. Hargrove, David J. Erickson, and Robert J. Oglesby. August 3, 2005. "Using Clustered Climate Regimes to Analyze and Compare Predictions from Fully Coupled General Circulation Models." *Earth Interactions,* 9(10): 1−27, doi:10.1175/EI110.1.

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 will 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.