From Measurements to Models:

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

Forrest M. Hoffman, William W. Hargrove**, Richard T. Mills*, A. D. Del Genio***, and Jasna Pittman*****

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

Introduction

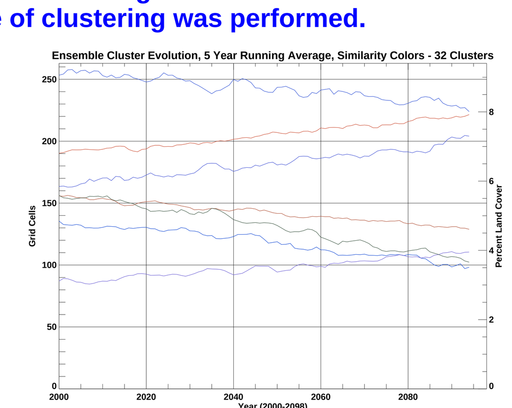
A statistical clustering technique was used to analyze output from the Parallel Climate Model (PCM) (Washington, et al.). Five 100−year ‘‘business as usual’’ scenario simulations were clustered, individually and then in combination, into 32 groups or climate regimes. Three PCM output fields were considered 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 is useful for analyzing observations to find atmospheric regimes or states that may be highly dynamic. CRYSTAL−FACE measurements were clustered to infer atmospheric transport mechanisms from combinations of six trace gases. The discrete groups defined by clustering proved to match well with backtrace experiments focused on determining the origin of air parcels during the observational period.

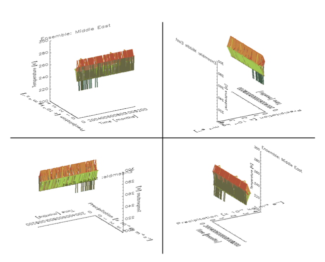
Multivariate Spatio−Temporal Clustering (MSTC) also offers a mechanism for comparing model results with measurements having different spatial and temporal scales. The goal of this project is to exploit such statistical techniques to better understand cloud processes and climate feedbacks, and to provide detailed information about when and where atmospheric models do not match observations.

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.



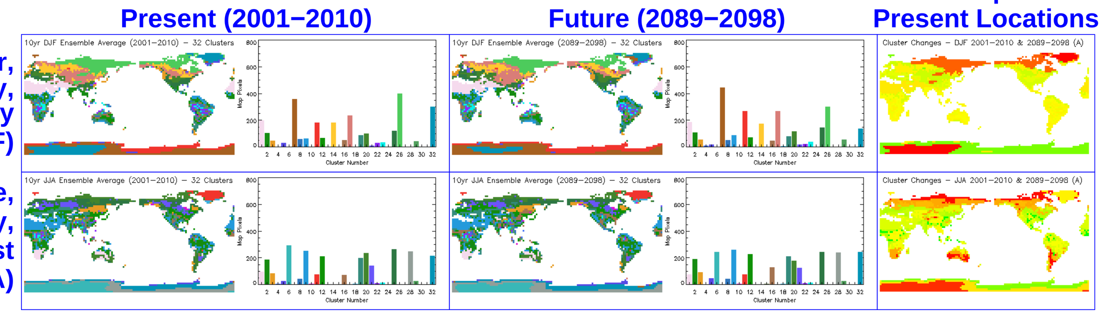
The Ensemble Average Regime Area Change graph at right shows the climate regimes which undergo a significant global area change. These curves are directly comparable to the individual Regime Area Change graphs for the individual BAU runs shown at left because they are in terms of the same single common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

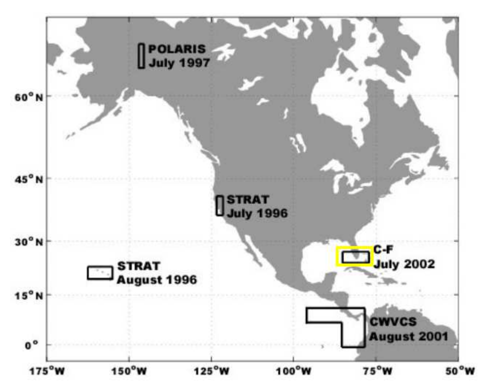
	Present (2001−2010)	Future (2089−2098)	Difference with respect to Present Locations
December, January, February (DJF)			
June, July, August (JJA)			

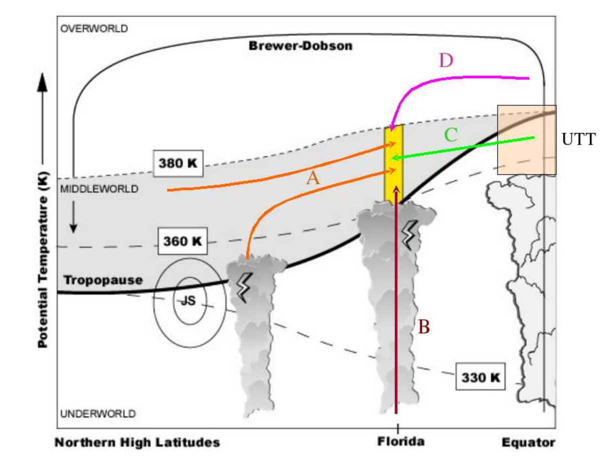


All of these results, along with maps and animations, are available in:

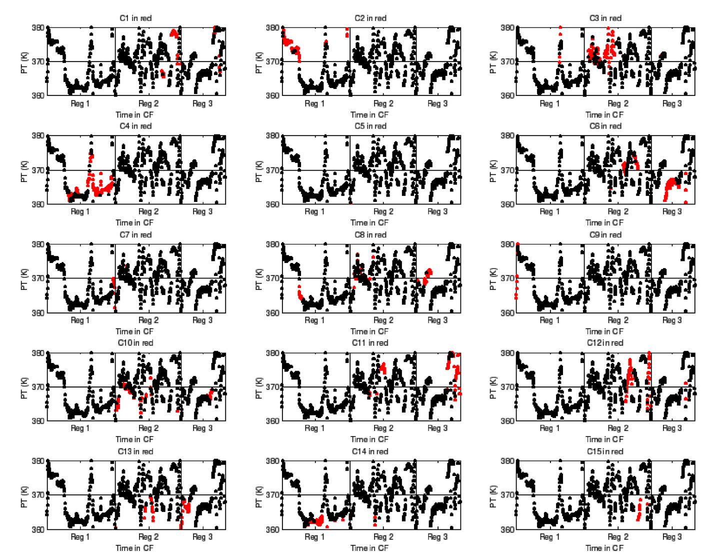
Hoffman, Forrest M., William W. Hargrove, David J. Erickson, and Robert J. Oglesby. August 3, 2005. "Using Clustered Climate Regimes to Analyze and Compare Predictions from Fully Coupled General Circulation Models." *Earth Interactions,* 9(10): 1−27, doi:10.1175/EI110.1.

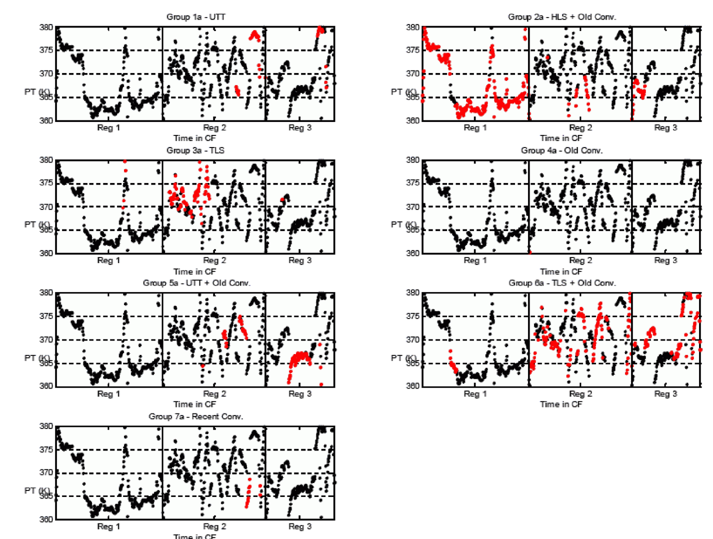
Clustering CRYSTAL−FACE Measurements into Atmospheric Regimes





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.

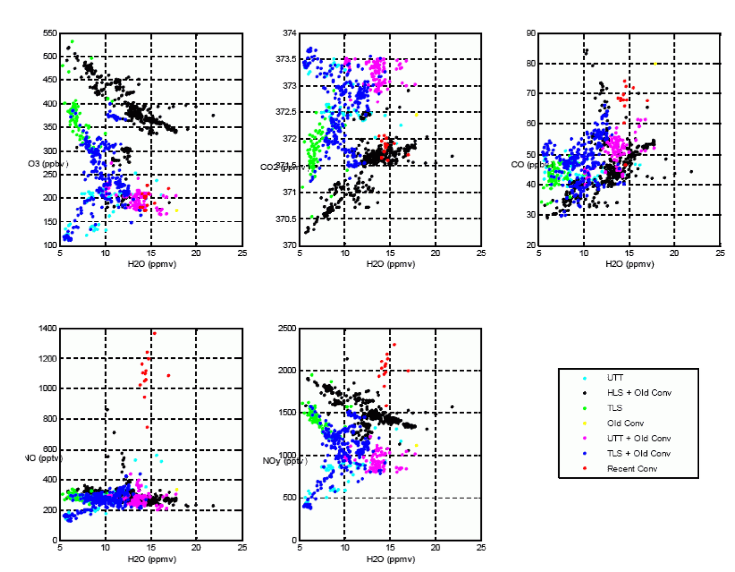




Four prospective atmospheric transport mechanisms within the "middleworld" troposphere were hypothesized: (A) High Latitude Stratosphere (HLS), (B) Old Convective (convective air with low NO lost to 1−week old atmospheric chemistry), (C) Upper Tropical Troposphere (UTT), and (D) Tropical Lower Stratosphere (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.

This study shows that atmospheric regimes produced by cluster analysis formed discrete intervals in the time domain, suggesting different aspects of atmospheric transport.



Pittman, J. V., E. M. Weinstock, R. J. Oglesby, D. S. Sayres, J. B. Smith, J. G. Anderson, O. R. Cooper, S. C. Wofsy, I. Xueref, C. Gerbig, B. C. Daube, E. C. Richard, B. A. Ridley, A. J. Weinheimer, M. Loewenstein, H. J. Jost, J. P. Lopez, M. J. Mahoney, T. L. Thompson, W. W. Hargrove, and F. M. Hoffman. "Dynamics of the Subtropical Lowermost Stratosphere During CRYSTAL−FACE." J. Geophysical Research−Atmospheres, in press.

Global Atmosphere/Cloud Regimes Defined by Multivariate Cluster Analysis from the Community Atmosphere Model (CAM3.0) Show Recurrent Seasonal Convective Regions

One of the primary goals of the ARM Program is to improve atmospheric models by making careful measurements of atmospheric behavior. Clouds are generally acknowledged to be the single biggest challenge to the accuracy of current atmospheric models. This project is applying multivariate statistical techniques to model output and ARM measurements to discern when and where the two disagree.

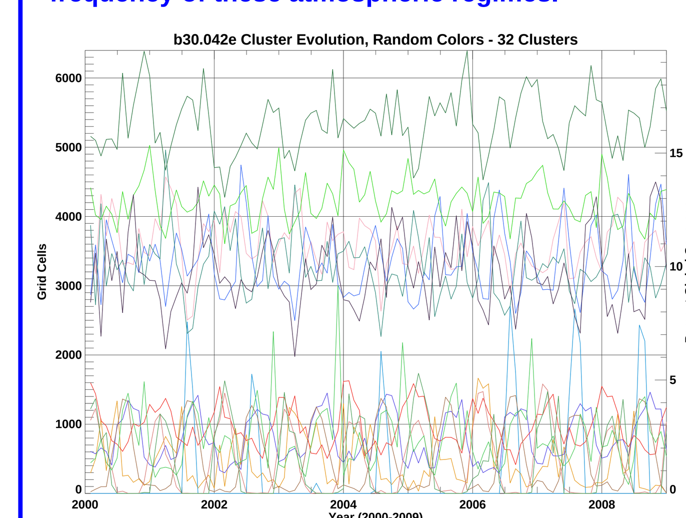
As in previous studies (discussed at left) an intial set of atmospheric regimes were defined using multivariate spatio−temporal clustering. Selected for the first suite of model data was a run from the Community Climate System Model (CCSM3) for the IPCC Fourth Assessment Report, in this case an SRES A2 scenario ensemble member. This model run was performed at T85 resolution (about 2 degrees) for the period 2000−2100. Monthly mean values for six variables that describe the atmospheric state (but not clouds) were chosen for inclusion in the analysis of the first ten years of the run. Five of the variables span all 26 model levels in the vertical of the atmosphere, resulting in 131 total factors. See table below.

Variables Used for Atmospheric Regimes

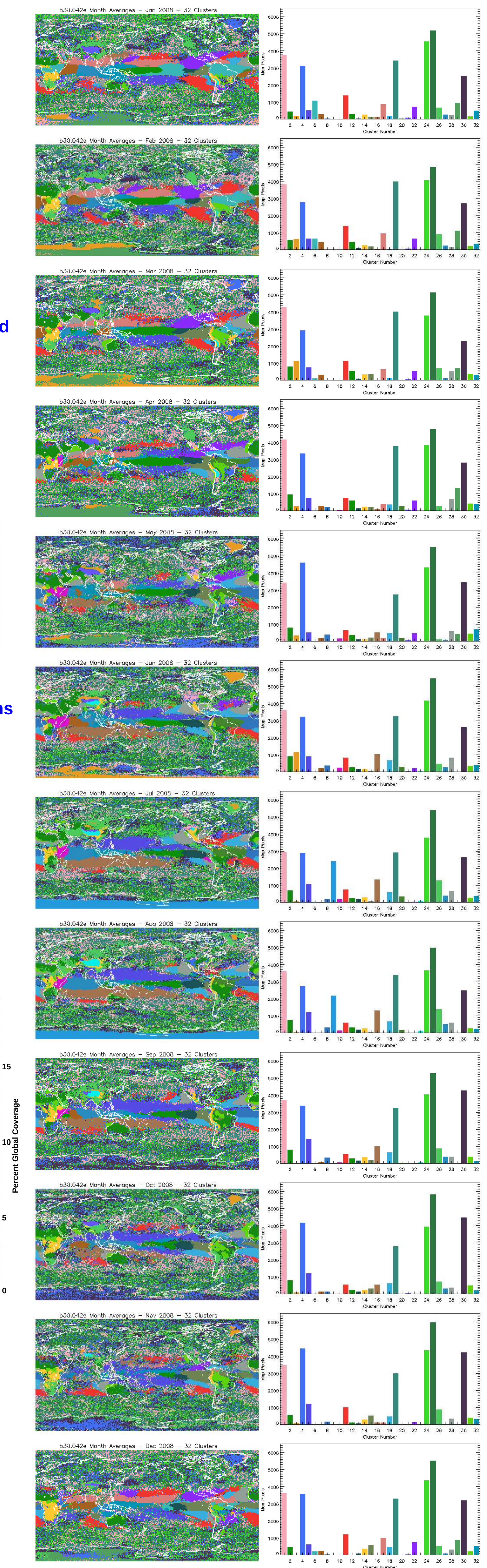
Variable	Description	Levels
PS	Surface Pressure	1
T	Temperature	26
RH	Relative Humidity	26
U	Zonal Wind	26
V	Meridional Wind	26
Z3	Geopotential Height	26

131 Total Factors

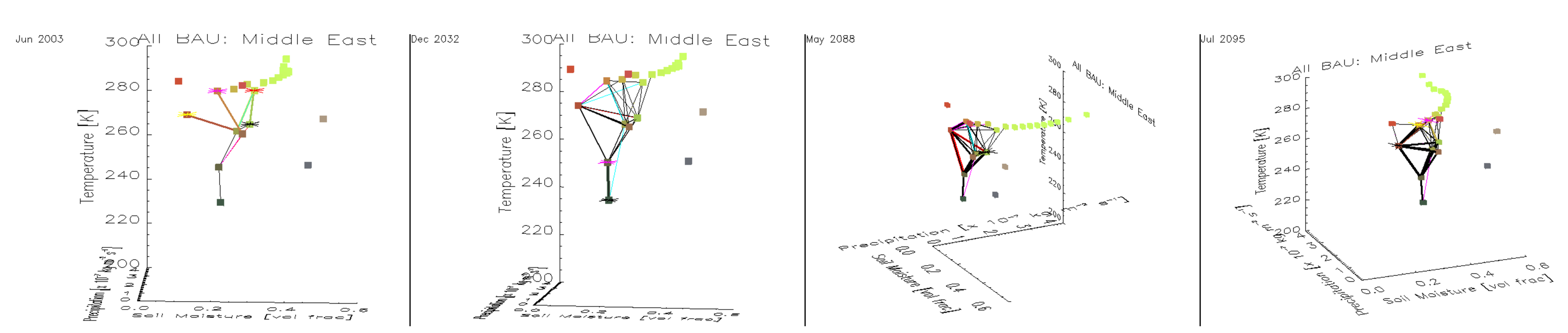
In the first analysis, 32 clusters/atmospheric regimes were requested. The maps and histograms at right show the distribution of those regimes (colored randomly) for each month of a single year (2008). Strong, coherent regimes appear in the tropics, suggesting that these indicate wide areas of strong convection. Coherent patterns appear over Greenland and Antarctica, as well as over large orographic features. The cluster evolution plot (below) shows the areal frequency of the largest regimes (using the same random colors as the maps) over the ten year period included in the analysis. A recurrent seasonal cycle is evident in the frequency of these atmospheric regimes.



Short−term dynamics are not represented well by the model’s monthly means; however, this run was chosen because it also includes six−hourly output of the important atmospheric fields. This enormous dataset will be used in a subsequent analysis to better match the temporal frequency of cloud dynamics and ARM measurements. In addition, a thorough factor analysis (too large to present here) has shown that little to no infor−mation is added by the inclusion of all model layers. As a result, only a few vertical levels will be included in subsequent analyses.



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 that 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.

Cluster analysis is useful for analyzing and intercomparing model results with measurements as well. The initial definition of atmospheric states from fundamental geophysical variables (shown above) demonstrates that cloud type and function can likely be determined from these model fields. The next steps in this project are to similarly determine atmospheric/cloud states from the long time series of measurements at ARM sites, and then to cluster model output and measurements together to determine where the modeled and observeds trajectories among states diverge and converge. The combination of states for each ARM site will yield insights into the representativeness of the entire ARM observation network, suggesting when and where the mobile facility should make additional measurements to maximize coverage or confirm model results.