A Cluster Analysis Approach to Comparing Atmospheric Radiation Measurement (ARM) Data and Global Climate Model (GCM) Results

Salil Mahajan[1](salilmahajan@tamu.edu), Forrest M. Hoffman[2], William W. Hargrove[3], Sigurd W. Christensen[2], Richard T. Mills[2]
[1]Department of Atmospheric Sciences, Texas A&M University, College Station, TX; [2]Oak Ridge National Laboratory, Oak Ridge, TN; [3]USDA Forest Service, Asheville, NC

1. Introduction:

Cluster analysis is a statistical technique to classify multivariate data into distinct regimes by grouping together similar data based on Euclidean distance in a multivariate state-space. Atmospheric column conditions can be classified into different groups or regimes based on their states in the multivariate space comprised of temperature, humidity, wind-speed, etc. Classification of complex atmospheric column conditions into various groups provides a systematic basis for comparison of regimes discerned in observations to those achieved in GCM simulations. Here, we investigate cluster analysis as an approach to compare abundantly available high temporal resolution multivariate atmospheric data from the Southern Great Plains (SGP) Atmospheric Radiation Measurement (ARM) site in Oklahoma to those simulated by the NCAR Community Climate System Model (CCSM) corresponding to that location.

2. Data:

The ARM Archive maintains a database of lower-atmospheric measurements for each of its facilities starting in 1992. The location of SGP for an ARM site was carefully chosen during the design phase of ARM to be free of topography such that it has a homogeneous climate over a large surrounding area of a size typical to that of a GCM grid box. In our study, we use readily available Value Added Products (VAPs) from the Archive for the SGP site for the time period April 2002 - April 2007. We use point data for temperature and water vapor mixing ratio vertical profiles at 48 height levels derived from the Atmospherically Emitted Radiance Interferometer (AERI) measurements at the central SGP facility with a temporal resolution of 8 minutes, and hourly wind speed profiles at 62 height levels derived from the NOAA wind profiler from the Lamont, OK site. The derived VAPs from the two instruments are called AERIPROF3FELTZ and WPDNMET.X1.b1 respectively. WPDNMET.X1.b1 also provides the surface pressure data used here.

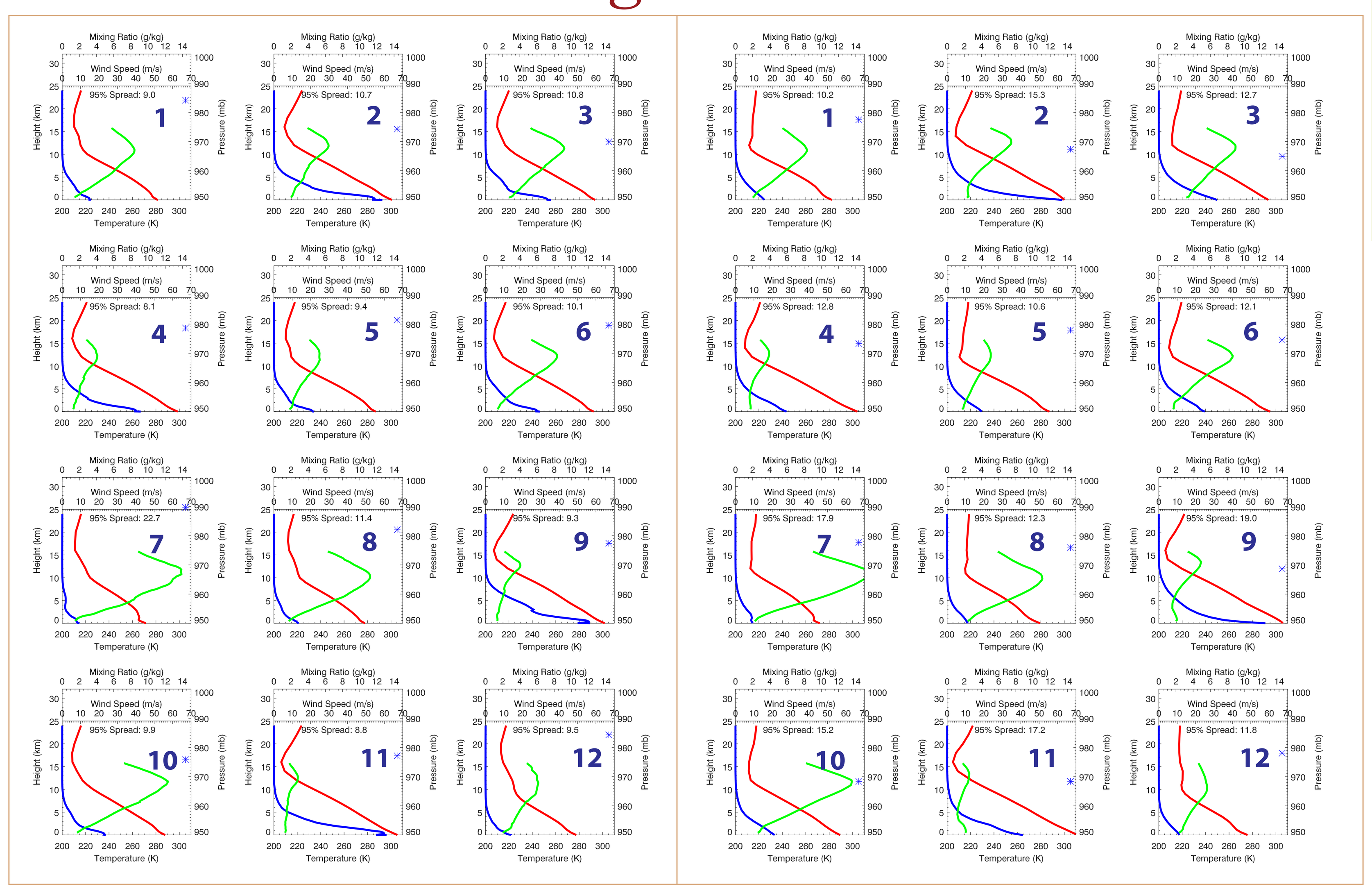
Eight years of output from a SRES A2 scenario 21st century run of the Community Climate System Model (CCSM) corresponding to the SGP site is compared with ARM data. The finest temporal resolution available for CCSM integrations are 6 hourly averages, with a spatial resolution of about 1.4° x 1.4°. ARM data are temporally averaged over 6 hours, and CCSM data are interpolated onto ARM vertical levels to facilitate comparison.

3. Methodology:

A parallel cluster analysis tool, developed at ORNL, employing an iterative k-means clustering algorithm is used to group multivariate atmospheric column data, comprisied of 159 variables, into 12 distinct clusters. We apply a 3 way approach to comparing ARM data with GCM output as follows:

a. ARM data are clustered into 12 regimes. CCSM data are projected onto those regimes.
b. CCSM data are clustered into 12 regimes. ARM data are projected onto those regimes.
c. ARM and CCSM data are combined and then clustered into 12 distinct regimes.

4a. Results: Clustering ARM Data



(a) ARM (b) CCSM projected on ARM Centroids

Figure 1: Vertical **temperature (red)**, **specific humidity (blue)** and **wind speed (green)** profiles and surface pressure (asterisk, right vertical axis) of **(a)** 12 centroids resulting from clustering of ARM data. **(b)** Mean of CCSM output assigned to each of the 12 clusters when projected onto those 12 centroids. Also shown is the radius of the spheroid containing 95% of the members assigned to each cluster in standardized coordinates. Note the stronger jet stream in regime 7 in CCSM output.

(a) ARM (b) CCSM projected on ARM

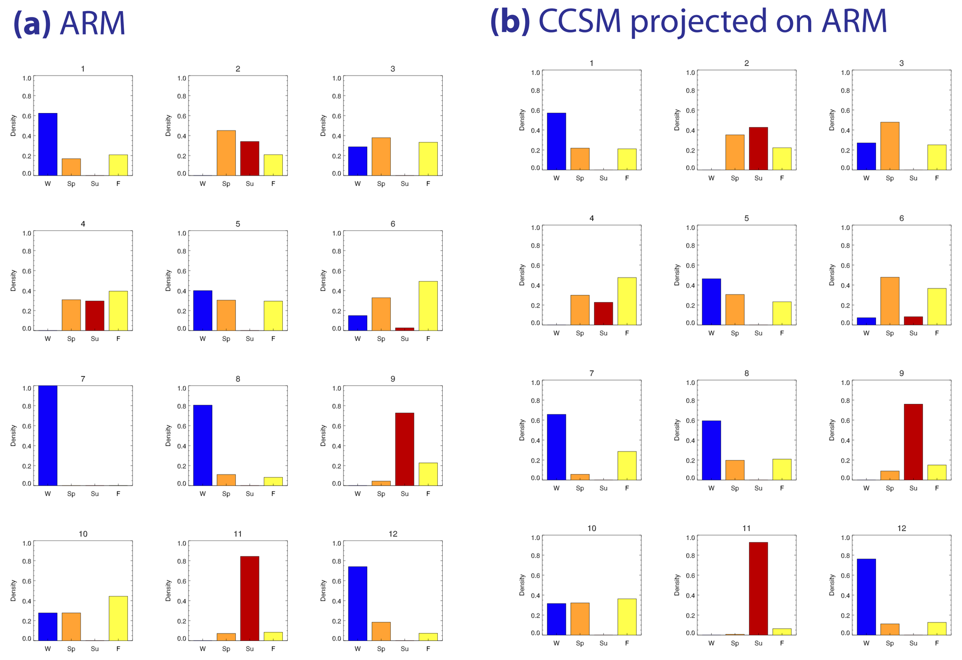


Figure 2: Seasonal frequency distribution of the number of data points in each of the 12 clusters for **(a)** ARM data. **(b)** CCSM output when projected onto ARM centroids. Note that while regime 7 is a characteristic winter regime in ARM data, it is also seen in the fall and spring seasons in CCSM output .

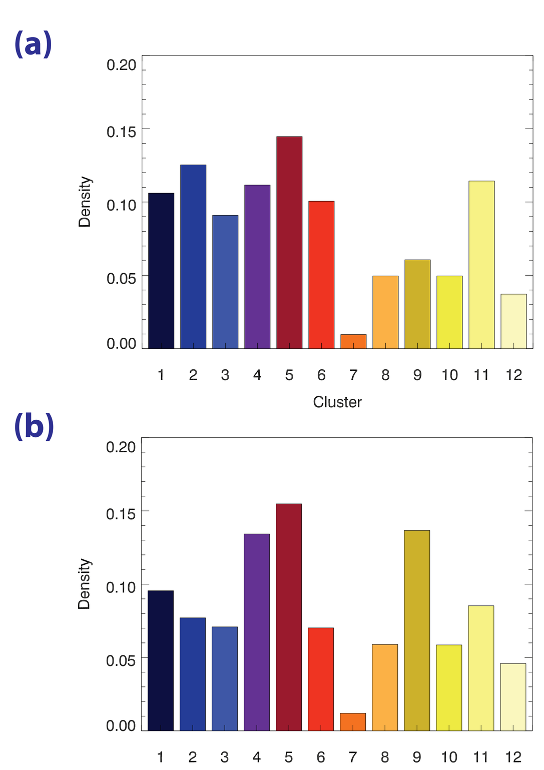
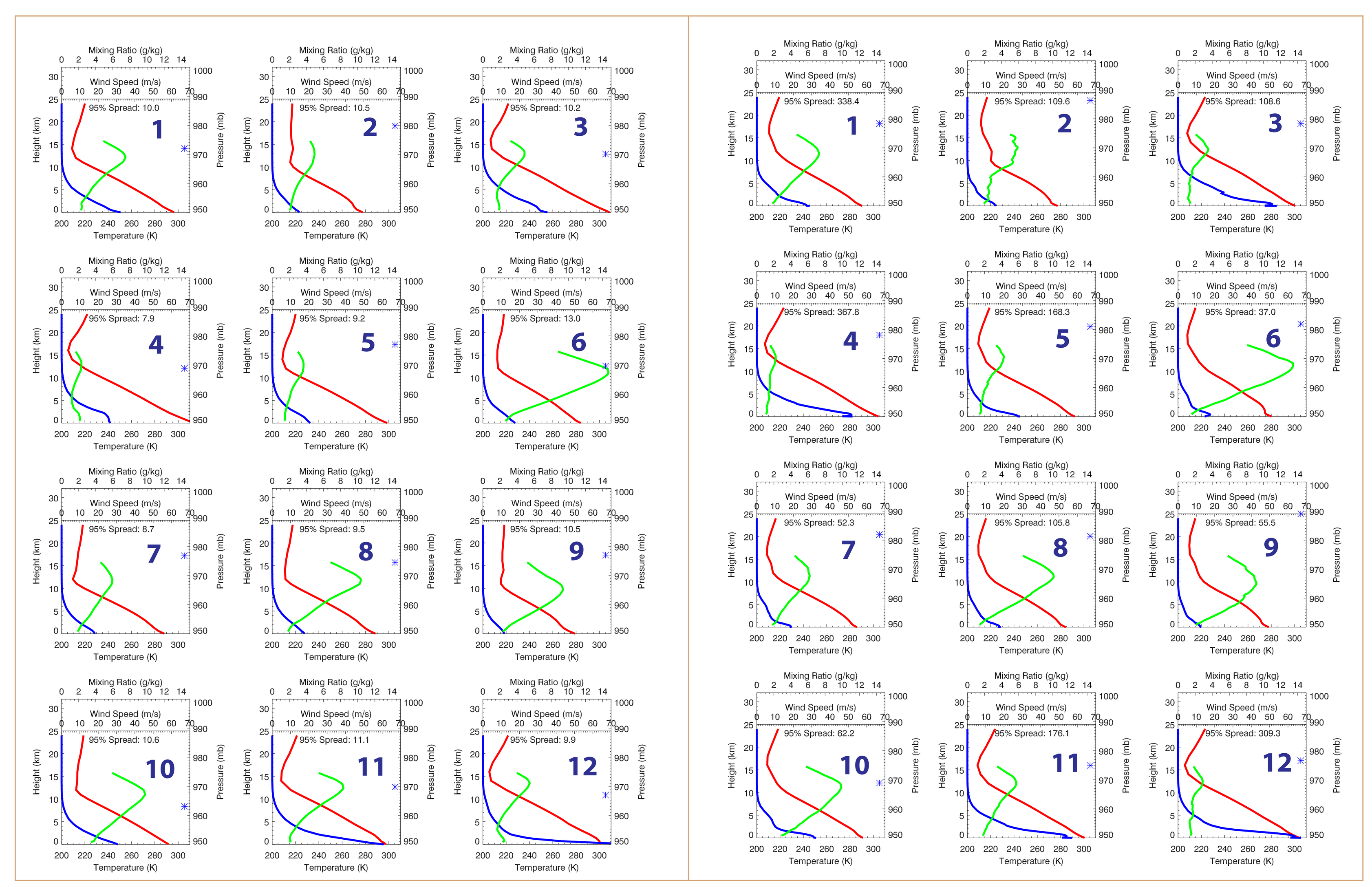


Figure 3: Frequency distribution of the number of data points in each of the 12 clusters for **(a)** ARM data. **(b)** CCSM output when projected onto ARM centroids. Note the similarity in the distribution.

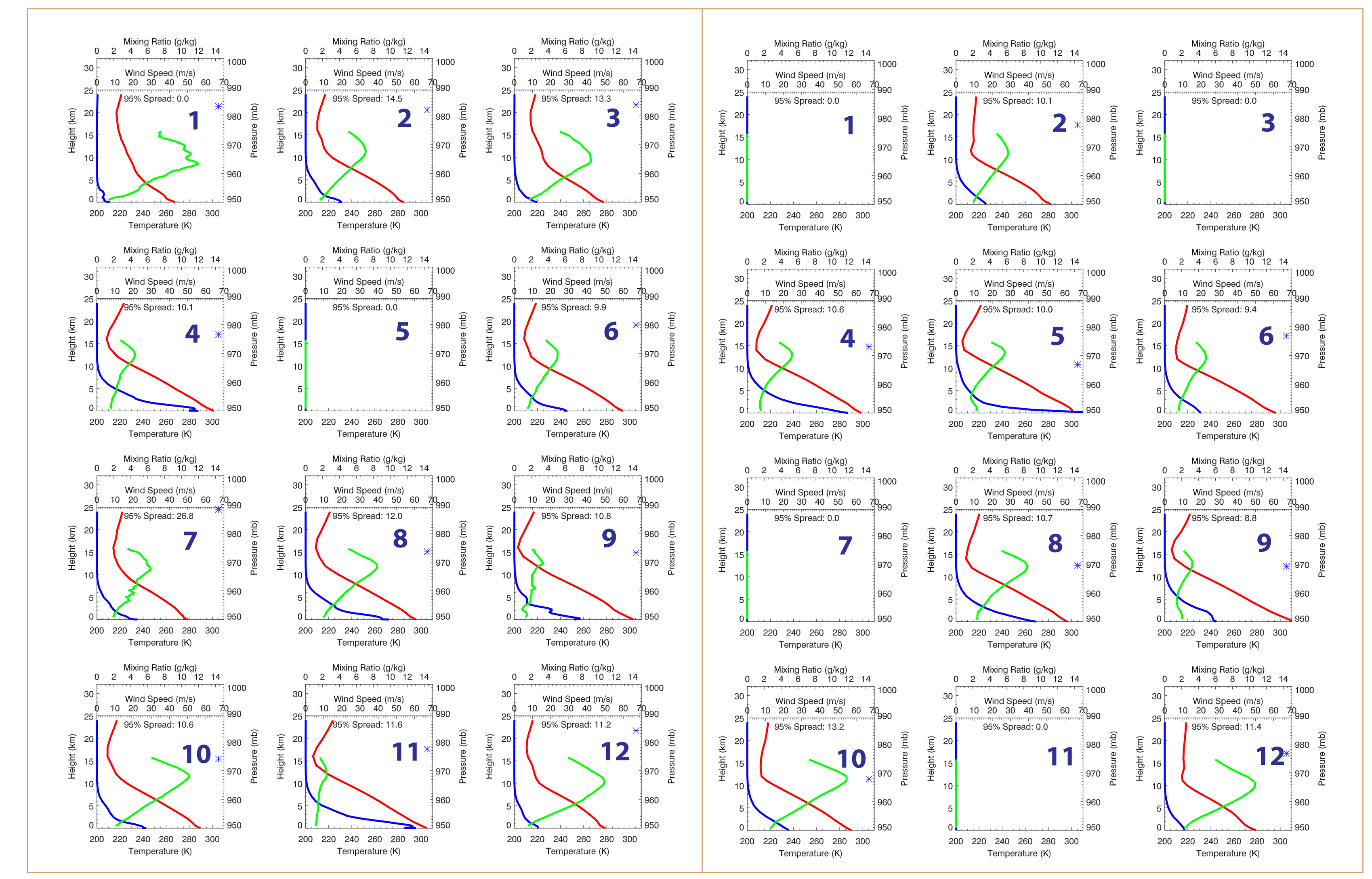
4b. Results: Clustering CCSM Output



(a) CCSM (b) ARM projected on CCSM Centroids

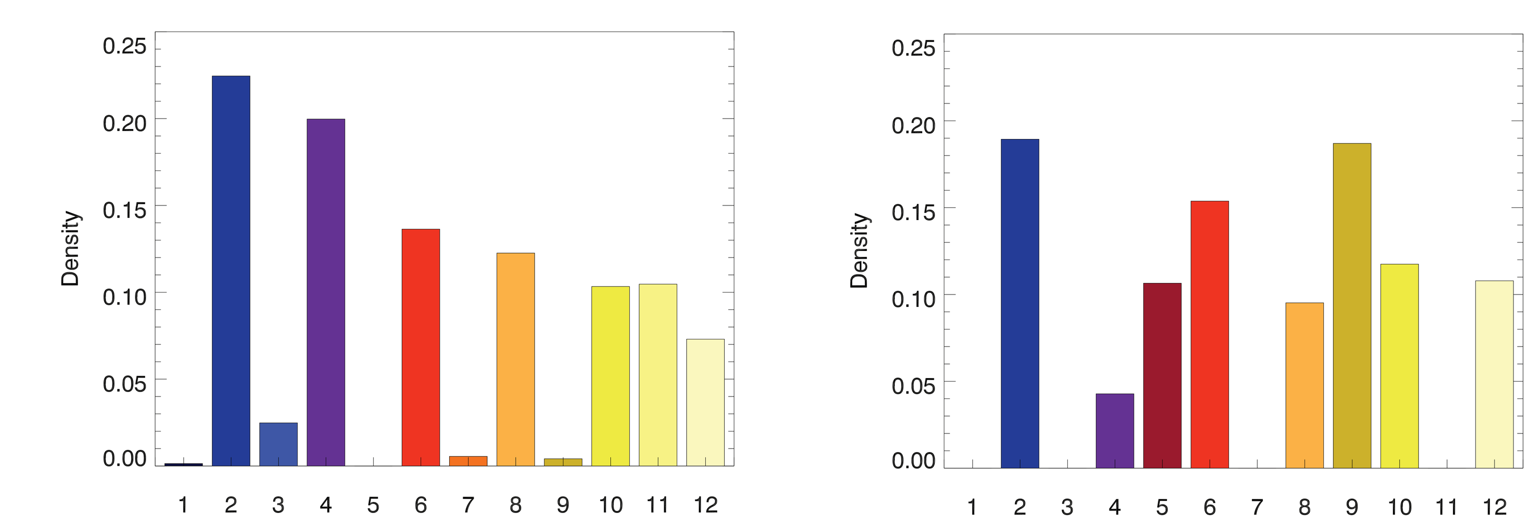
Figure 4: Vertical **temperature (red)**, **specific humidity (blue)** and **wind speed (green)** profiles and surface pressure (asterisk, right vertical axis) of **(a)** 12 centroids resulting from clustering of CCSM output. **(b)** Mean of ARM data assigned to each of the 12 clusters when projected onto those 12 centroids. Also shown is the radius of the spheroid containing 95% of the members assigned to each cluster in standardized coordinates. Note the large radius for ARM data when assigned to CCSM clusters implying the large multi-variance of ARM data in state-space.

4c. Results: Clustering ARM + CCSM



(a) ARM projected on ARM + CCSM Centroids (b) CCSM projected on ARM + CCSM Centroids

Figure 5: Vertical **temperature (red)**, **specific humidity (blue)** and **wind speed (green)** profiles and surface pressure (asterisk, right vertical axis) of **(a)** Mean of ARM data assigned to each of the 12 clusters when projected onto the 12 centroids resulting from clustering of pooled ARM data and CCSM output. **(b)** Mean of ARM data assigned to each of the 12 clusters when projected onto the 12 centroids resulting from clustering of pooled ARM data and CCSM output. Also shown is the radius of the spheroid containing 95% of the members assigned to each cluster in standardized coordinates. Note the missing regimes in ARM data and CCSM output revealing atmospheric regimes reached in one but not in the other. Missing regimes 1, 3 and 7 in CCSM output have low representation even in ARM data.



(a) ARM projected on ARM + CCSM Centroids (b) CCSM projected on ARM + CCSM Centroids

Figure 6: Frequency distribution of the number of data points in each of the 12 clusters resulting from clustering of pooled ARM data and CCSM output for **(a)** ARM data and **(b)** CCSM output. Note the over-representation of regime 9 and 5 associated with high temperatures, moderate/high humidity conditions and high vertical wind-shear and under-represenation of regimes 4 and 11 associated with high temperatures, high humidity but low vertical wind-shear in CCSM output.

5. Conclusions:

a. Cluster analysis reveals that fall and spring atmospheric column conditions are well simulated in CCSM at the SGP site. However, distinct atmospheric regimes are also identified which might impact the simulation of clouds and precipitation and hence affect the local predicted radiation budget.
b. CCSM simulates strong jet streams in the fall and spring not seen in ARM data.
c. ARM observations suggest that hot, humid lower tropospheric conditions are usually associated with low vertical wind-shear conditions. Such conditions in CCSM output are associated with stronger shear. Low shear conditions occur in CCSM usually with hot, but only moderately humid lower tropospheric conditions.
d. ARM data demonstrate larger multi-variance than CCSM output.

References: Hargrove, William W., and Forrest M. Hoffman, 2004: Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions. *Environmental Management*, **34**(5), 39-60.
Hoffman, Forrest M., William W. Hargrove, David J. Erickson, and Robert J. Oglesby, 2005: Using Clustered Climate Regimes to Analyze and Compare Predictions from Fully Coupled General Circulation Models, *Earth Interactions* **9**(10), 1-27.

Acknowledgements: Research sponsored by the Climate Change Research Division (CCRD) of the Office of Biological and Environmental Research (OBER) within the U.S. Department of Energy's Office of Science (SC). Oak Ridge National Laboratory (ORNL) is managed by UT-Battelle, LLC, for the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. ARM data were obtained from the ARM Archive available at www.arm.gov