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.

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” scenarios 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.

Climate Trajectories

A geographic location exists in only one climate regime at a single point in time. But as you increment through time, the location can shift among the climate regimes according to changes in the forcing parameters. The regimes map to two–dimensional “climatic” space. The boundary of time–space or state space is a “trajectory” or “web” among the climate regimes. Each regime has a set of descriptive variables. Multivariate Clustering is a powerful tool which can provide a common basis for comparison across space and time for comparing various model results with long time series observations to better understand cloud processes and climate feedbacks.

Clustered Climate Regimes

The clustering process establishes an exhaustive set of occupied climate regimes (i.e., the 32 clusters) in the state space occupied by the simulated atmosphere’s 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.

Climate Regime Definitions & Maps

The centroids of each of the climate regimes represents the climate regime in the original measurement units. The first column of the table shows the range of values 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 given color weighting to the overall regime.

The top row of maps is colored randomly while the bottom row shows the same climate regimes colored using similarity colors. The first column only shows the regime at any single point in time. By incrementing time, the clustered climate regimes can be thought of as climate states in an N-dimensional phase (or state) space. These states are defined in terms of their original measurement units.

Regime Area Changes

Because the same clustered set of climate conditions are present in the same geographic area globally for any climate regime at a single point in time, each of these climate regimes is defined by its unique value or values in the three variables. These conditions are not observed; only climate regimes experiencing similar changes in each run are plotted.

Climate Trajectories

A geographic location exists in only one climate regime at a single point in time. By incrementing through time, the location can shift among the climate regimes according to changes in the forcing parameters. The regimes map to two–dimensional “climatic” space. The boundary of time–space or state space is a “trajectory” or “web” among the climate regimes. Each regime has a set of descriptive variables. Multivariate Clustering is a powerful tool which can provide a common basis for comparison across space and time for comparing various model results with long time series observations to better understand cloud processes and climate feedbacks.

Climate Manifolds

Tracking out the entire seasonal and annual trajectory for a single location yields a climate “manifold” in state space representing the shapes of the predicted climate occupancy for that location. The predicted climate extremes and the frequencies of occupation are easily seen in the graphical representation.

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 a single trajectory in the common state space. These 32 cluster centroids are a new set of points representing the multivariate behavior of the climate model in geographic space globally for any climate regime at a single point in time. Each of these points is defined by the unique value or values in the three variables. These conditions are not observed; only climate regimes experiencing similar changes in each run are plotted.

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 its proper 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 regions of high 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.

Conclusions

Cluster analysis is a powerful tool which can provide a common basis for comparison across space and through time for complex climate simulations. Because it can utilize 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 easily–parsed variables, cluster analysis is a powerful tool that automatically 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 interpret climate simulations, but also to analyze observations and intercompare them with model results. The area changes graphs above could show two climatic and climate feedbacks. The trajectory figures could show multivariate climate behavior. When measurements are used in combination with model predictions, two–dimensional climate manifolds can be seen to diverge when models and observations do not agree. This multivariate approach can be used to identify patterns and trends in time series measurements with model or observations results. The manifold figures could show the occurring in the climate system. Multivariate clustering can be used to interpret climate simulations and observations and to group them into meaningful patterns and trends.