Geophysical Research Abstracts
Vol. 12, EGU2010-14814, 2010
EGU General Assembly 2010

Pseudo Random Classification of Circulation Patterns - Comparison to Deliberate Methods

Andreas Philipp
Institute for Geography, University of Augsburg, Augsburg, Germany

Classification of circulations patterns, e.g. of sea level pressure patterns, can be done by many different methods, e.g. by cluster analysis, methods based on eigenvalues or those based on the leader algorithm like the Lund classification. However none of these methods can give clear advice on the problem of appropriate numbers of classes and even though the number is decided different methods lead to different results. High efforts are made to find methods leading to indisbutable results. However, doubts on the classifiability of tropospheric circulation states have been raised recently and the existence of natural groups of similar patterns within the circulation data, which might be caused by circulation regimes, are questionable.
If those groups or clusters exist, methods which are designed to find them, in particular cluster analysis, should be superior to classification schemes based on pseudo random definition of classes. In order to prove this assumption, a classification method called "random centroids" has been designed, for each class choosing one single circulation pattern using a random number generator and assigning all remaining patterns to them according to the minimum Euclidean distance. Evaluation metrics like the "explained cluster variance" for pressure, temperature and precipitation are calculated in order to compare those pseudo random classifications to classifications provided by the cost733cat dataset including many different classification catalogs for various methods (COST Action 733 "Harmonisation and Applications of Weather Type Classifications for European regions"). By running the randomcent method 1000 times the empirical probability density function of the evaluation metrics can be established and provides information about the probability for the established deliberate methods to be better than random classifications.
The results show that most of the classifications fail to succeed the 95th percentile of the empirical probability density functions indicating that the respective methods cannot find intrinsic groups of similar patterns in the data. However, since some few methods do show a clear tendency for being significantly better it is concluded that any structure in the data exists and can be detected.