Geophysical Research Abstracts, Vol. 8, 01530, 2006
SRef-ID: 1607-7962/gra/EGU06-A-01530
© European Geosciences Union 2006

Atmospheric circulation regimes: Can cluster analysis provide the number?

B. Christiansen

Danish Meteorological Institute (boc@dmi.dk)

The existence of multiple regimes in the extra-tropical tropospheric circulations is a hypothesis both of theoretical importance and with potential practical consequences. It is also a controversial hypothesis and an abundance of conflicting results regarding both the existence and the number of regimes can be found in the literature.

Studies of atmospheric regime behavior are often based on clustering methods such as k-means and mixture models. In the basic implementation of these methods the number of clusters has to be specified a priory and "How many clusters?" is a highly non-trivial question.

We show that applying the clustering methods to the northern hemisphere winter tropospheric geopotential heights gives conflicting and fragile results. In particular the number of clusters depends both on the clustering algorithm and on the period considered. Furthermore, the clustering methods find multiple regimes when applied to data similar to the atmospheric data but drawn from a unimodal, skewed distribution.

We also show that both clustering methods report multiple clusters for idealized data drawn from distributions that are skewed or platykurtic but otherwise smooth and without bumps or shoulders. In these cases the number of clusters found depends on the sample size.

We conclude that in the atmospheric data set studied the clustering methods provide only weak evidence for multiple regimes although the data is non-Gaussian with high statistical significance. We also conclude that statistical models with basically unknown properties should be approached with utmost care or avoided completely.