



## **Parallel Clustering Algorithms with Application to Climatology**

**H. Bisgin**, H. N. Dalfes

INFORMATICS INSTITUTE, ISTANBUL TECHNICAL UNIVERSITY,  
(hbisgin@be.itu.edu.tr)

How to determine the ecoregions or climate zones has been a controversial issue. Discussion appears from the debate if the selected method is objective or not. In order to prevent from subjective approaches, one has to utilize some formulations which are independent from such interferences. Cluster analysis, which is one of the famous pattern recognition tools and has hierarchical and non-hierarchical methods, contributes to the objectivity in this sense. Instead of relying on any expertise or personal interpretations, clustering methods provide a mathematical approach with the multivariate data set.

The aim of this work is to implement cluster analysis tools to climatology data in order to obtain climate zones with some other statistical techniques that will make the study more precise. In order to clarify, first we determine how many clusters or regions do we need for valid regionalization by posing a validation criterion on the algorithm.

While acquiring such a number of clusters, we have done experiments with both the high dimensional set where there are from 96 to 109 number of variables and the reduced dimensional data space which was obtained via Principal Component Analysis (PCA). Under the criterion we posed, in the region  $30^{\circ} - 50^{\circ}$  N  $3^{\circ} - 60^{\circ}$  E varying number of clusters obtained as the different variable combinations are used. Nevertheless, in  $34^{\circ} - 43^{\circ}$  N by  $23^{\circ} - 47^{\circ}$  E where Turkey covers almost all the frame, we consistently acquired 4 climate zones. During the cluster analysis (CA), besides the serial k-means algorithm we have also utilized parallel version. According to the time measurements, it is seen that whereas serial code performs better with the reduced

dimensions, parallel version is good at dealing with high dimensional sets.

Consequently, the k-means algorithm suggests another point of view for the climate zones of both regions where it is possible to observe some climatic blocks that are generally stable. More precisely, 4 climate zones appear in all cases concerning the second frame which represents some differences from the preceding climate zone definitions which are based on conventional and hierarchical ideas.