Geophysical Research Abstracts,
Vol. 11, EGU2009-8740-4, 2009
EGU General Assembly 2009

Spatio-Temporal Clustering of Monitoring Network

I. Hussain (1,3) and J. Pilz (2,4)

(1) ijaz.hussain.kherani@gmail.com , (3) PhD student at Department of Statistics University of Klagenfurt, Universitätsstasse 65-67, 9020 Klagenfurt, Austria, (2) juergen.pilz@uni-klu.ac.at, (4) Department of Statistics University of Klagenfurt, Universitätsstasse 65-67, 9020 Klagenfurt, Austria

Pakistan has much diversity in seasonal variation of different locations. Some areas are in desserts and remain very hot and waterless, for example coastal areas are situated along the Arabian Sea and have very warm season and a little rainfall. Some areas are covered with mountains, have very low temperature and heavy rainfall; for instance Karakoram ranges. The most important variables that have an impact on the climate are temperature, precipitation, humidity, wind speed and elevation. Furthermore, it is hard to find homogeneous regions in Pakistan with respect to climate variation.

Identification of homogeneous regions in Pakistan can be useful in many aspects. It can be helpful for prediction of the climate in the sub-regions and for optimizing the number of monitoring sites. In the earlier literature no one tried to identify homogeneous regions of Pakistan with respect to climate variation. There are only a few papers about spatio-temporal clustering of monitoring network.

Steinhaus (1956) presented the well-known K-means clustering method. It can identify a predefined number of clusters by iteratively assigning centriods to clusters based.

Castro et al. (1997) developed a genetic heuristic algorithm to solve medoids based clustering. Their method is based on genetic recombination upon random assorting recombination. The suggested method is appropriate for clustering the attributes which have genetic characteristics.

Sap and Awan (2005) presented a robust weighted kernel K-*means* algorithm incorporating spatial constraints for clustering climate data. The proposed algorithm can effectively handle noise, outliers and auto-correlation in the spatial data, for effective and efficient data analysis by exploring patterns and structures in the data.

Soltani and Modarres (2006) used hierarchical and divisive cluster analysis to categorize patterns of rainfall in Iran. They only considered rainfall at twenty-eight monitoring sites and concluded that eight clusters existed. Soltani and Modarres (2006) classified the sites by using only average rainfall of sites, they did not consider time replications and spatial coordinates.

Kerby et.al (2007) purposed spatial clustering method based on likelihood. They took account of the geographic locations through the variance covariance matrix. Their purposed method works like hierarchical clustering methods. Moreovere, it is inappropiriate for time replication data and could not perform well for large number of sites.

Tuia.et.al (2008) used scan statistics for identifying spatio-temporal clusters for fire sequences in the Tuscany region in Italy. The scan statistics clustering method was developed by Kulldorff et al. (1997) to detect spatio-temporal clusters in epidemiology and assessing their significance. The purposed scan statistics method is used only for univariate discrete stochastic random variables.

In this paper we make use of a very simple approach for spatio-temporal clustering which can create separable and homogeneous clusters. Most of the clustering methods are based on Euclidean distances. It is well known that geographic coordinates are spherical coordinates and estimating Euclidean distances from spherical coordinates is inappropriate. As a transformation from geographic coordinates to rectangular (D-plane) coordinates we use the Lambert projection method. The partition around medoids clustering method is incorporated on the data including D-plane coordinates. Ordinary kriging is taken as validity measure for the precipitation data. The kriging results for clusters are more accurate and have less variation compared to complete monitoring network precipitation data.

References

Casto.V.E and Murray.A.T (1997). Spatial Clustering with Data Mining with Genetic Algorithms. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.56.8573

Kaufman.L and Rousseeuw.P.J (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley series of Probability and Mathematical Statistics, New York.

Kulldorf.M (1997). A spatial scan statistic. Commun. Stat.-Theor. Math. 26(6), 1481-1496

Kerby. A , Marx. D, Samal. A and Adamchuck. V. (2007). Spatial Clustering Using the Likelihood Function. Seventh IEEE International Conference on Data Mining – Workshops

Steinhaus.H (1956). Sur la division des corp materiels en parties. Bull. Acad. Polon. Sci., C1. III vol IV:801– 804

Snyder, J. P. (1987). *Map Projection: A Working Manual.* U. S. Geological Survey Professional Paper 1395. Washington, DC: U. S. Government Printing Office, pp. 104-110

Sap.M.N and Awan. A.M (2005). Finding Spatio-Temporal Patterns in Climate Data Using Clustering. Proceedings of the International Conference on Cyberworlds (CW'05)

Soltani.S and Modarres.R (2006). Classification of Spatio -Temporal Pattern of Rainfall in Iran: Using Hierarchical and Divisive Cluster Analysis. *Journal of Spatial Hydrology Vol.6, No.2*

Tuia.D, Ratle.F, Lasaponara.R, Telesca.L and Kanevski.M (2008). Scan Statistics Analysis for Forest Fire Clusters. *Commun. in Nonlinear science and numerical simulation 13,1689-1694.*