

McMASTER UNIVERSITY  
GRADUATE PROGRAM IN STATISTICS

---

STATISTICS SEMINAR

- Speaker:** Rong Zhu  
Department of Mathematics & Statistics  
McMaster University
- Title:** Model Clustering and its Application to E-coli Classification.  
**Day:** Tuesday, February 26, 2008
- Time:** 3:30 - 4:30 PM
- Place:** HH/217 - Deloitte Colloquium Room  
(refreshments in HH/216 at 3:00 PM)

SUMMARY

The classification of objects into groups such that within a group the objects have a set of common characteristics is an important environmental problem, particularly in zoning. We propose a general procedure for forming the groups when a data set of measurements is available for each object. A parametric model of the same family is fitted for individual data set, thus representing the associated object. We develop a special test called model linking to compare several models regarding the equality of a subset of coefficients. The likelihood ratio statistic is used in the test and the p-value is utilized as a similarity measure among them. The larger the p-value, the more similar the models, so the more similar the associated objects. Several strategies are proposed to form the groupings such as cluster peeling, pairwise combining, as well as a speeding technique called splitting-and-binding. A small simulation study demonstrates the utility of these methods. Finally, we applied model clustering to an environmental study where the interest is to classify E-coli bacteria according to their response to different antibiotic treatments. These E-coli bacteria were sampled biweekly from twenty locations in three Canadian watersheds. Metric closeness in parametric space used by conventional method and likelihood closeness in model space employed by model clustering are discussed in this application.

This is a joint work with Dr. Abdel H. El-Shaarawi.

## REFERENCES

- El-Shaarawi, A.H. and Kwiatkowski, R.E. (1977). A Model to Describe the Inherent Spatial and Temporal Variability of Parameters in Lake Ontario 1974. *Journal of Great Lakes Research*, 3(3-4), 177-183.
- El-Shaarawi, A.H. and Shah, K.R. (1978). Statistical Procedures for Classification of a Lake. *Inland Waters Directorate, Environment Canada, Scientific Series No. 86*.
- Lehmann, E.L. (1986). *Testing Statistical Hypotheses (second edition)*. John Wiley and Sons, New York.
- Liu W., Jamshidian M. and Zhang Y. (2004). Multiple Comparison of Several Linear Regression Models *JASA*, 99, 395-403.
- Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26, 354-359.

## ABOUT THE SPEAKER



Dr. Rong Zhu is an Assistant Professor of Statistics in the Department of Mathematics and Statistics at McMaster University. He obtained his BS and MS in Statistics from USTC (China), and Ph.D in Statistics from UBC. Dr. Zhu's research interests include non-normal time series, mixture model and generalized clustering analysis.

## MORE SEMINAR INFORMATION

Please contact Angelo Canty at 905-525-9140 ext. 27079, email: [cantya@mcmaster.ca](mailto:cantya@mcmaster.ca).