1056-Z1-1500 **Brian Bies**, **Kathryn Dabbs** and **Hao Zou*** (hzou@macalester.edu), 1600 Grand Avenue, St Paul, MN 55105. *On Determining the Number of Clusters–An Empirical Study of Different Algorithms.*

In this paper, we perform the first empirical tests comparing several existing algorithms for determining the number of clusters in a data set (the gap statistic, X-means, G-means, data spectroscopic clustering and self-tuning spectral clustering). We use a large number of data sets randomly generated with varying distributions (normal and uniform distributions) and parameters (dimensions, number of clusters, number of data points per cluster, and degree of separation between points). The results show that G-means and X-means perform best on the majority of test cases. In addition, the gap statistic returns good estimates for fewer dimensions and number of clusters, but is less accurate and much slower when the number of clusters and dimensions increases. We therefore explore ways to improve the gap statistic, and formulate the problem in the simplified continuous context to consider its theoretical basis. (Received September 22, 2009)