FLAME: an automated model-based approach to the high-dimensional analysis of flow cytometry data

Saumyadipta Pyne[1], Kui Wang[2], and Geoff McLachlan[2, 3]
[1]Broad Institute of MIT and Harvard University, 7 Cambridge Centrer, Cambridge MA 02142, USA, [2]Department of Mathematics, University of Queensland 4072, Australia, and [3]Institute for Molecular Bioscience, University of Queensland 4072, Australia.

FLAME (FLow analysis with Automated Multivariate Estimation) is an automated procedure for the analysis of high-dimensional flow cytometry (FCM) data [Pyne 2010]. With FLAME, a model-based approach is provided that directly fits multivariate finite mixtures, using skew and heavy-tailed distributions, to address the complexities of flow cytometric analysis and to deal with high-dimensional cytometric data without the need for projection or transformation. It has been demonstrated to have the capacity to detect rare populations, to model robustly in the presence of outliers and skew, and to perform the critical task of matching cell populations across samples that enables downstream analysis. The approach facilitates the application of flow cytometry to new, complex biological and clinical problems. FLAME is available in the publically available *GenePattern* software developed at the Broad Institute of MIT and Harvard (http://www.broadinstitute.org/cancer/software/genepattern/) for the analysis of genomic data.   FLAME has been applied to many FCM datasets, including all five FlowCAP datasets in all four types of challenges with encouraging results. The fitting of the mixtures of skew normal and *t*-densities in FLAME is undertaken using the EMMIX-skew R package available at http://maths.uq.edu.au/~gjm/. In the flowCAP competition, the method of partitioning around medoids [Kaufman 1990] was used to decide on the number of clusters. The choice of starting points for the iterative fitting of the skew densities included a random sampling strategy.   No transformation and variable selection methods were needed to be employed on the data sets in this competition.

References:
Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T.-I., Maier, L.M., Baecher-Allan,C., McLachlan, G.J., Tamayo, P., Hafler, D.A., De Jager, P.L., and Mesirov, J.P. (2009),   Automated high-dimensional flow cytometric data analysis. Proceedings of the National Academy of Sciences USA, 106, 8519-8524.

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