Comparison

In this simulation study, we consider comparing RC with MCLUST, which employs the finite Gaussian mixture model to fit the data and applies the BIC criterion to decide the optimal number of clusters. We are interested in the case when the data has outliers and the resulting mean profile deviates from its true value. We extract a subset of the data in Simulation Study 2 of the paper. The subset data is the expression data of genes in clusters 1, 3, 4. In this scenario, $N = 150, J = 10, R = 4, K = 3$. Then we add noise to the expression data to turn some of them into "outliers". For each gene in each condition, with probability 0.10, its maximum expression data (in terms of magnitude) will be doubled. The following table is the clustering results of RC and MCLUST in the cases with outliers and without outliers. Each case is repeated for 20 times. We can see that RC is more robust than MCLUST when outliers exist, by taking advantage of the full replicate data.

TABLE 1
ARI measurement

	RC	MCLUST	RC (no outlier)	MCLUST (no outlier)
Case 1	0.997(0.007)	0.666(0.074)	0.996(0.008)	0.865(0.195)
Case 2	0.982(0.023)	0.933(0.145)	0.982(0.021)	0.987(0.016)
Case 3	0.996(0.010)	0.759(0.204)	1.000(0.000)	0.931(0.169)
Case 4	0.983(0.013)	0.964(0.015)	0.987(0.013)	0.990(0.011)
Case 5	0.997(0.007)	0.998(0.006)	0.996(0.008)	1.000(0.000)

TABLE 2
Number of inferred clusters

	RC	MCLUST	RC (no outlier)	MCLUST (no outlier)
Case 1	3.0(0.0)	5.7(0.7)	3.0(0.0)	4.0(1.5)
Case 2	3.0(0.0)	3.4(1.3)	3.0(0.0)	3.0(0.0)
Case 3	3.0(0.0)	4.9(1.7)	3.0(0.0)	3.5(1.2)
Case 4	3.0(0.0)	3.0(0.0)	3.0(0.0)	3.0(0.0)
Case 5	3.0(0.0)	3.0(0.0)	3.0(0.0)	3.0(0.0)