Cluster Mouse Activities (MATLAB script)

Last updated: 03.02.19 (NM)

This MATLAB script takes activity data obtained from metabolic cages and clusters (k-mean algorithm) the data according to inherent differences in activity levels exhibited between cages. Output provides cluster memberships and cluster diagnostics and descriptive statistics.

For working example, follow described procedure using allData.xlsx provided in SOP folder.

# Procedure

1. Data obtained from metabolic cages must first be prepared in an appropriate input format prior clustering in MATLAB. Prepare excel sheet with 5 columns labeled: mouseID, time, XTOT, XAMB, ZTOT (**Figure 1**)
   * **mouseID**: numerical value used to identify mouse/cage
   * **time**: time of measurement (*copy from metabolic cage output spreadsheet*)
   * **XTOT**: x movement (*copy from metabolic cage output spreadsheet*)
   * **XAMB**: y movement (*copy from metabolic cage output spreadsheet*)
   * **ZTOT**: z movement (*copy from metabolic cage output spreadsheet*)

|  |
| --- |
|  |
| **Figure 1**. Data input example |

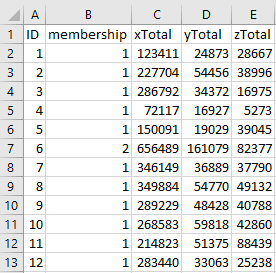
1. Save excel sheet in same folder as ‘*activityCluster.m*’
2. Open ‘*activityCluster.m*’ in MATLAB and specify analysis parameters (**Figure 2**):
   * **File**: name of excel file with input data (e.g., ‘*allData.xlsx*’)
   * **Sheet**: name of excel sheet where input data is stored (e.g., ‘*Sheet1*’)
   * **kClusters**: Cluster number (2-3 recommended)
     + Number of groups expected in data.
   * **nReplicates**: number of clustering replicates (10 recommended)
     + There is a risk of false positive clusters if number of replicates is low.

|  |
| --- |
|  |
| **Figure 2**. Specifying analysis parameters in MATLAB |

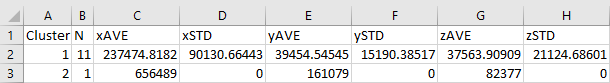
1. Press ‘RUN’ in MATLAB

# Outputs.

1. MATLAB will output graphical results and also save cluster memberships and descriptive statistics in new Excel file
2. Results are saved to new Excel file named [File RESULTS currentTime] and stored in two sheets:
   * **Clusters Sheet**: Contained cluster membership and total activity counts for x, y, z



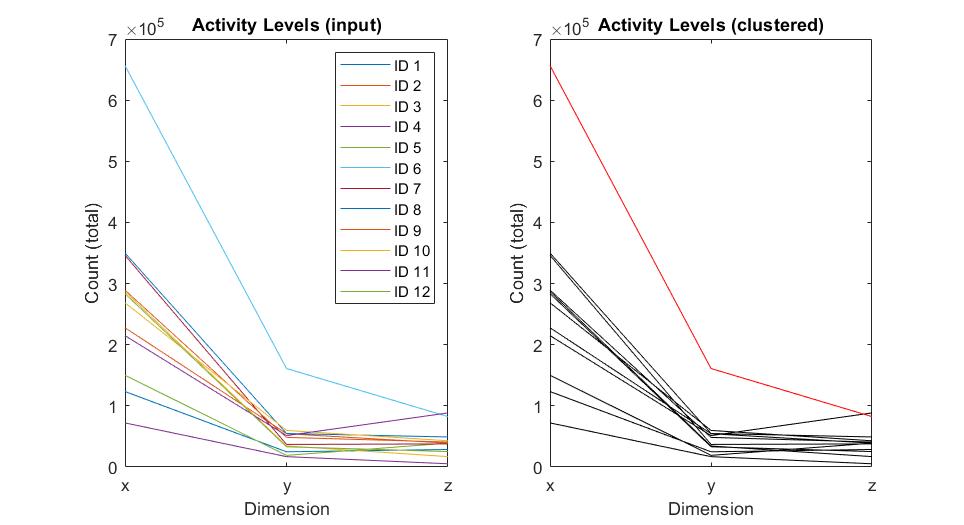
* + **Descriptives Sheet**: cluster averages and standard deviations along with number of members per cluster.



1. Graphical results include 3 figures (cage-specific activity levels, silhouette plot and descriptive statistics for each cluster):

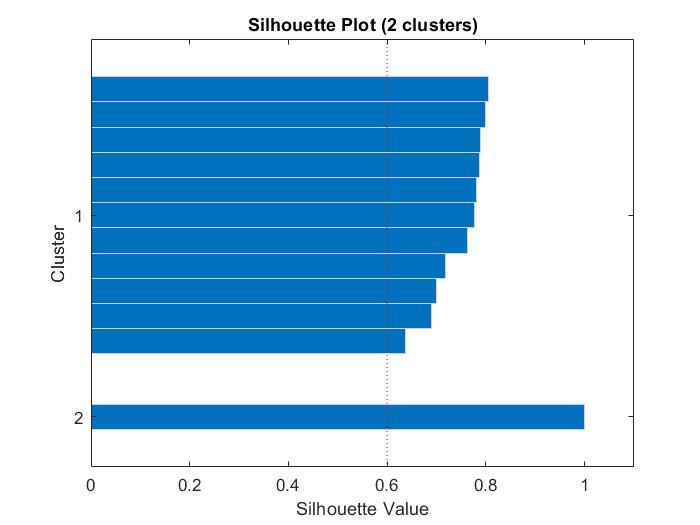
**(i) cage-specific activity levels**

Total activity for each cage (specified by ID) is shown as raw input (*left*) and clustered (*right*). Cluster membership in right plot is specified by color (e.g., cluster 1 = red, cluster 2 = black).



**(i) Silhouette plot**

The silhouette plot is used as a diagnostic tool to determine whether the clustering configuration is appropriate (i.e., do we have true clusters?). Values range from -1 to 1, such that higher values indicating the a given member is similar to members within the same cluster and dissimilar to neighbouring clusters. Silhouette values >0.6 (*red reference line*) are typically interpreted as satisfactory clusters, while values <0.4 suggest that clusters are inappropriate.



**(i) Descriptive statistics for each cluster**

Mean counts (± standard errors) for each activity metric are plotted by cluster. Colors correspond with cluster memberships in cage-specific activity level plots.

