GMMsort – Spike sorting with Gaussian mixture models

Instructions for using the graphical user interface (GUI) of the spike sorting described in Souza et al., Spike sorting with Gaussian mixture models. *BioRxiv*, 2018.

First steps to use the GUI

Download the GMMsort files from https://github.com/tortlab/GMM-spike-sorting. Open Matlab and, in the GMMsort directory, run the command *GMMsort()* to open the GUI. Once the GUI is open, click the *LOAD* button and choose the file with the waveforms (e.g., sample_waveforms.mat). The waveforms of the selected channel (Ch id #) will be shown in the *Unsorted waveforms* panel (Figure 1).

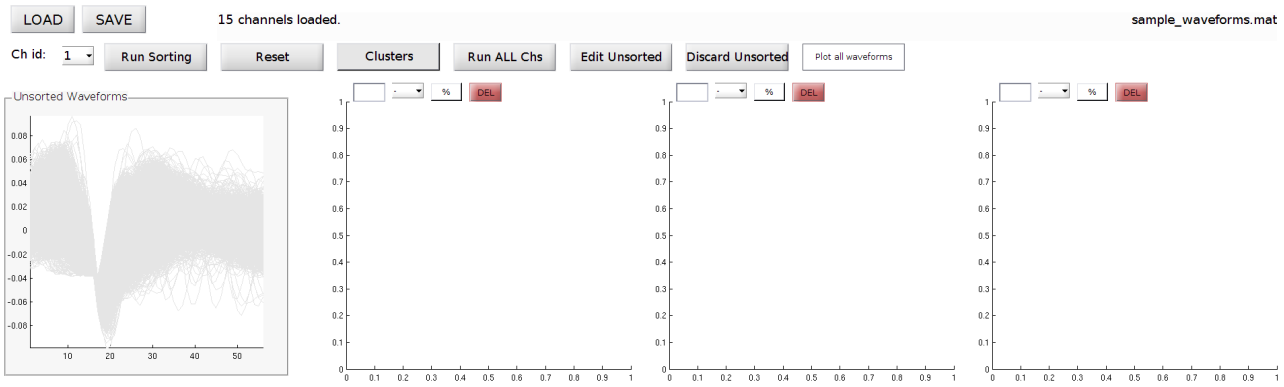

Figure 1. Initial screen of the GUI showing the unsorted waveforms from channel 1 of the loaded file.

Sorting the waveforms

The multiple functionalities of the GUI can be applied in varying order, depending on the results of the sorting. In Figure 2 we present a general pipeline to perform sorting (numbers correspond to structures shown in Figures 3 and 4), explained below:

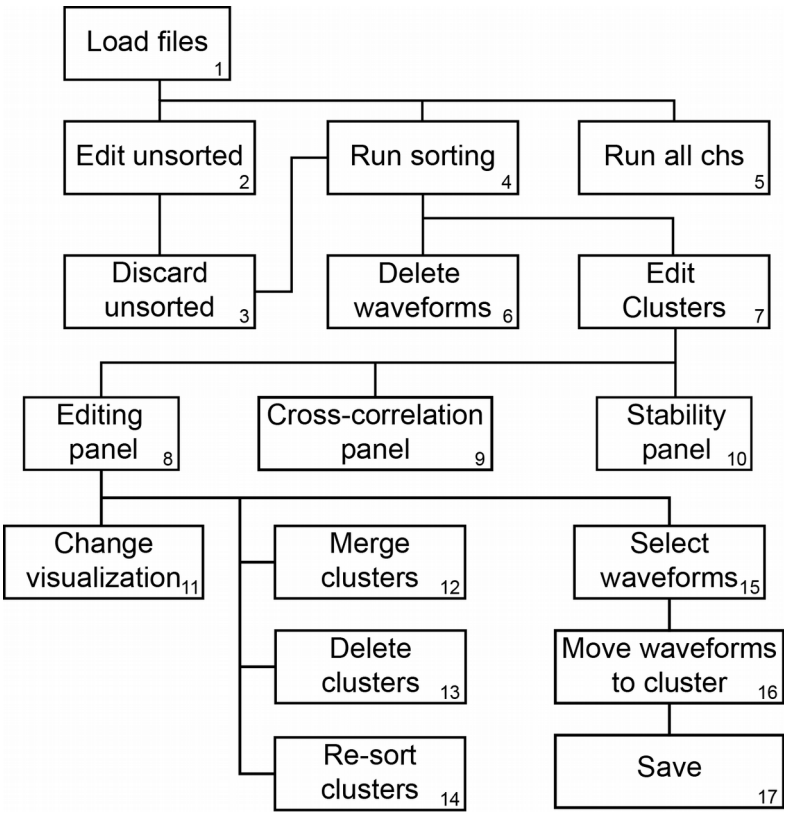

Figure 2. General pipeline for using the GUI. Each of the steps is explained below. Numbers link each step with their corresponding structure in the GUI (see Figures 3 and 4).

Run Sorting (4): performs the sorting of the selected channel or the *Run All Chs* button to sort all the channels of the file. After the algorithm is run, the resulting clusters are shown in the 7 remaining panels of this window (see Figure 3). Note that the GUI is initially set to plot no more than 1000 waveforms for each cluster. This can be changed in the *Plot all waveforms (19)* button.

Each of the panels is composed of:

- an editbox *(20)* to change the label of the current cluster;
- a listbox *(20)* to select the shown cluster;
- the '%' button *(20)* to change the visualization from lines to quantiles (.05, .25, .5, . 75 and .95 quantiles);
- the DEL button *(6)*.

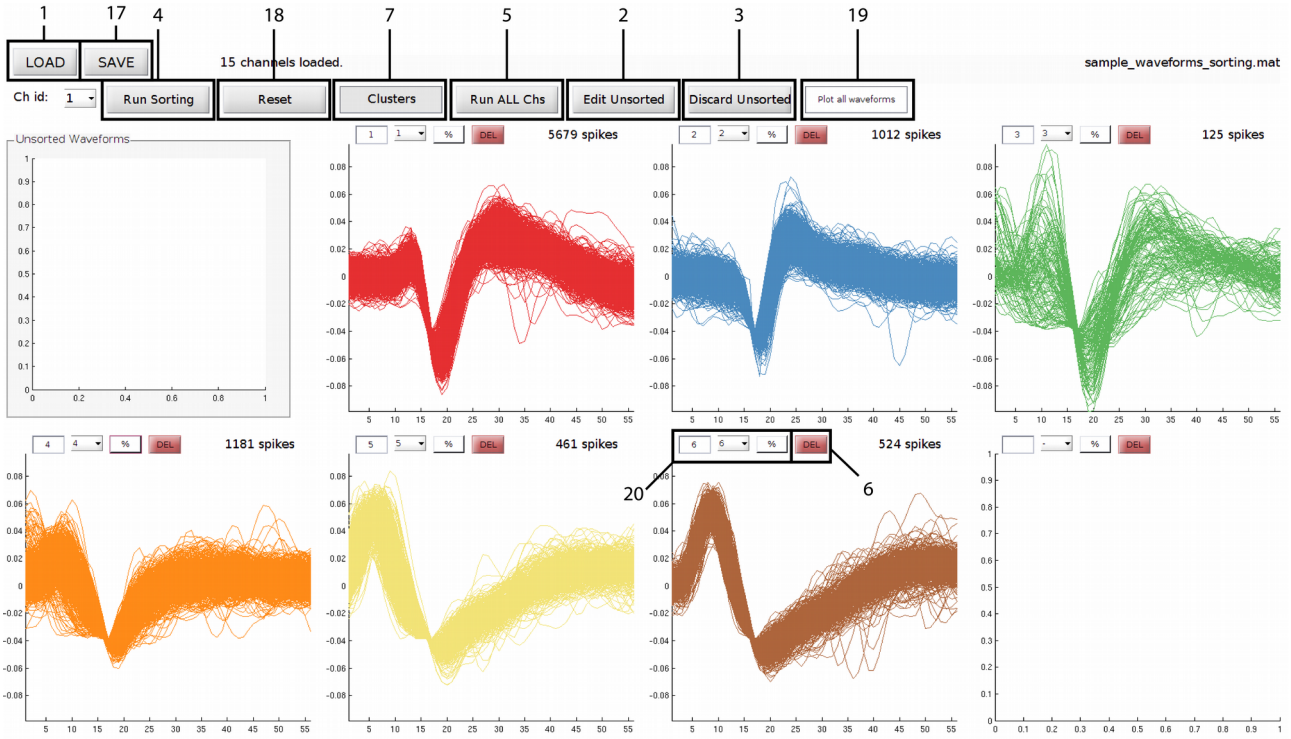

Figure 3. Main GUI window after the sorting of channel 1. In each panel, the cluster to be shown or its label can be defined. Different quantiles can be plotted instead of all the waveforms (0.05; 0.25; 0.5; 0.75 and 0.95 quantiles). Numbers in the boxes match the ones in Figure 2.

Run all channels (5): Iteratively performs the sorting of all the loaded channels.

Reset (18): Discards any modification done after the sorting.

Edit Unsorted (2): Sets the unsorted waveforms as a new, editable cluster. This can be done either before or after running the sorting algorithm.

Discard Unsorted (3): Discards the unsorted waveforms from the file (e.g., discarding artifacts prior to performing sorting). This operation will not be undone by the Reset button.

Delete Waveforms (6) ('DEL' button in each panel): Sets the waveforms of the cluster as unsorted, thus deleting the cluster.

Edit Clusters (7) ('Cluster' button): Opens or close the editing window in which merging and editing clusters can be done (Figure 4).

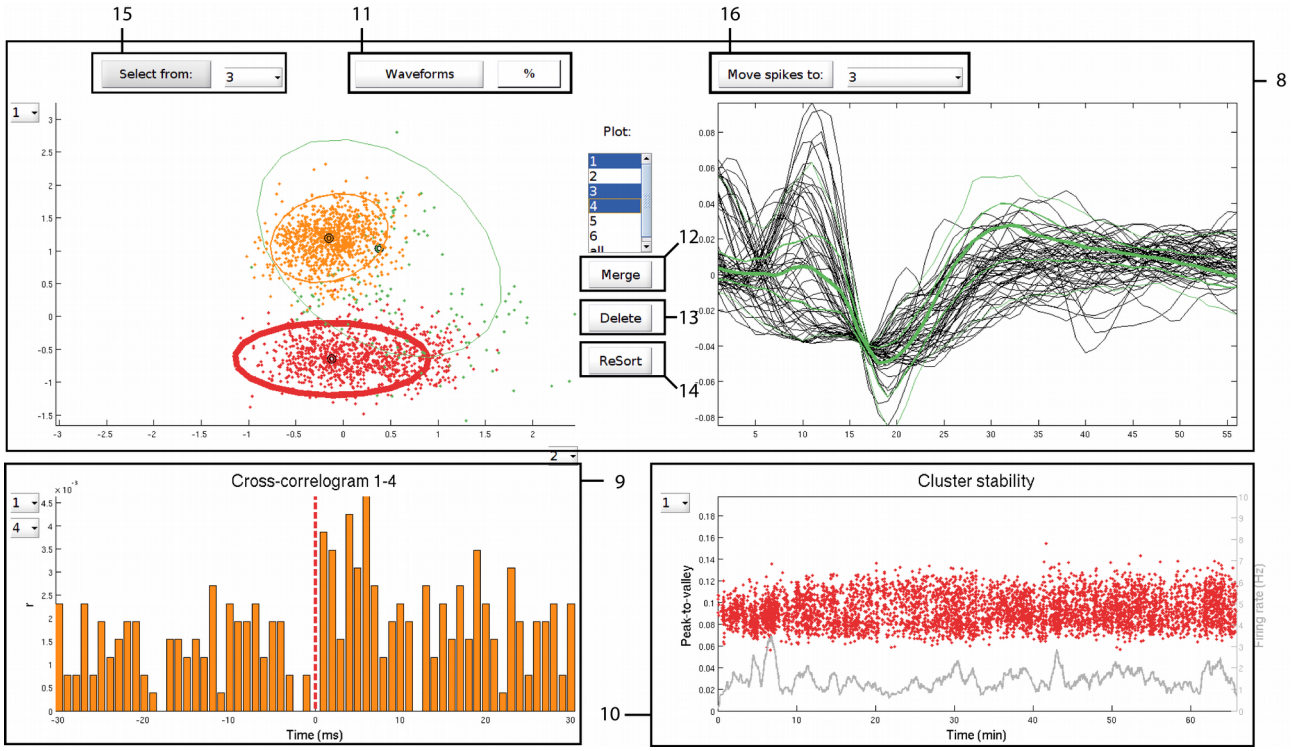

Figure 4. Second GUI window. Top: *Editing panel (8)*, showing a scatter plot of the first 2 dimensions of the clustering space with the centers and standard deviation of 3 clusters (left), and a sample of waveforms selected by the cursor (right). The 5/25/50/75/95 quantiles of waveforms of cluster 3 are also shown (green lines). Bottom: *cross-correlogram panel (9)*, computed for clusters 1 and 4 (left), and the *stability panel (10)* (peak-to-valley distance and firing rate over time) of cluster 1 (right). Numbers in the boxes match the ones in Figure 2.

Change visualization (11) ('Waveforms' button): Changes the way the clusters are shown in the *Editing panel (8)*, alternating between 3 options: the waveforms of each cluster; a scatter plot of 2 dimensions of the clustering space; or a line plot with all the 5 dimensions of the clustering space (see Figures 4 and 5). The shown clusters can be selected in the *Plot* list. Additionally, line plots can be viewed as quantiles (the % button) and the 2 dimensions of the scatter plot can be chosen from the listboxes near the x- and y-axes.

Merge Cluster (12) ('Merge' button): Merges the shown clusters into a single one while preserving the Gaussian of each cluster in the model.

Delete Clusters (13) ('Delete' button): Deletes the Gaussians of the shown clusters. The waveforms that belonged to them are then reassigned to the remaining Gaussians following the same rule as previously: the Gaussian with highest probability determines the identity of the waveform.

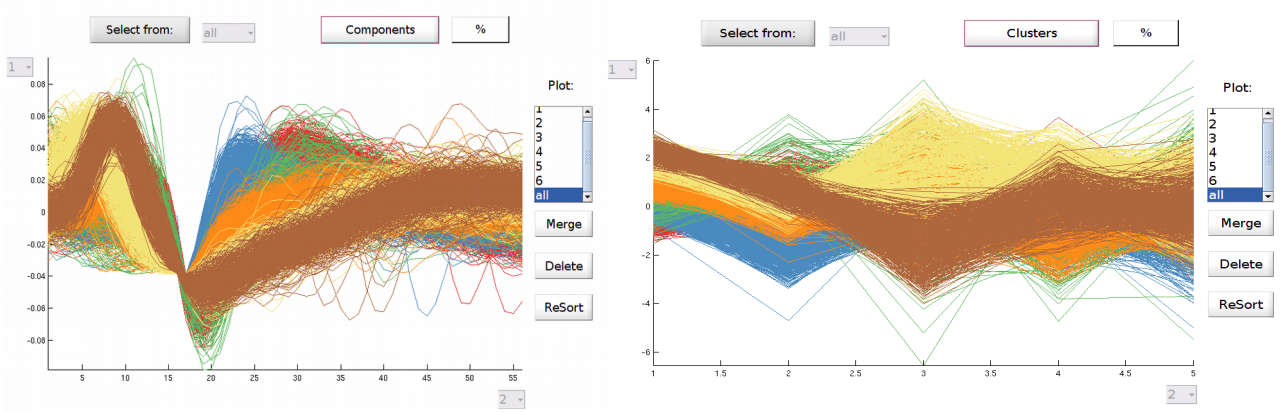

Figure 5. Two other possible views of the *Editing panel*. Shown are the waveforms of each cluster (left) and the 5 dimensions of the clustering space (right).

Re-sort (14) ('ReSort' button): Repeats the Clustering step using only the shown clusters. The resultant model is merged with the previous model so that the unselected clusters remain the same. This allows repeating the sorting without discarding well-separated clusters. Note that the clustering space does not change.

Select waveforms (15) ('Select from #' button): Allows the user to draw polygons and select particular points or lines. The selection is restricted to the shown clusters, or to the cluster selected in the listbox. After the selection is completed, waveforms are plotted in black, where they can be compared to each of the clusters.

Move waveforms to a cluster (16): ('Move spikes to #' button): Allows the user to move the selected waveforms to a particular cluster, a new one, or to the unsorted group.

Cross-correlation panel (9): Shows the (normalized) cross- or auto-correlogram between two clusters defined by the 2 listboxes in the y-axis. Dashed line represents the reference.

Stability panel (10): Shows the firing rate and peak-to-valley distance of the cluster defined in the listbox across the recording session.

Saving the sorting (17) ('Save' button): Opens a dialog window to select the directory and filename of the sorting output and then saves a '.mat' file with the *struct* 'data'.

Input and output formats

The input variable ‘data’ must have at least the two following fields:

- *waveforms*: a 1-by-*Nch* cell vector, where *Nch* is the number of channels. Each cell must have an *Nspike*-by-*N* matrix, with *Nspike* waveforms of length *N*.
- *spiketimes*: a 1-by-*Nch* cell vector. Each cell must have an *Nspike*-by-1 vector with the spiketimes.

The output variable has additionally the fields:

- *class_id*: a 1-by-*Nch* cell vector. Each cell contains an *Nspike*-by-1 vector with the cluster identity of each spike. Unsorted waveforms are classified as NaN.
- *model*: a cell vector similar to *class_id*, with the resultant model of each channel. The model itself is a *struct* with the mean (mu), covariance matrix (S), weight (alpha) of each Gaussian.
- *clustering_space*: a cell vector with the clustering space of each channel. Each clustering space is an *Nspike*-by-5 matrix.