Authors: Ikaro J S Beraldo, Renan A V Mendes, Cleiton Lopes-Aguiar (2022)

Sleep-Wake Cycle Classification Toolbox

The toolbox was implemented in MATLAB and the graphical interface was developed on MATLAB's App Designer. The codes were developed using as input data local field potential recordings from CA1 and electromyographic recordings (EMG) of mice and rats. The algorithm separates the recordings in epochs of 10 seconds, extracts the Theta/Delta ratio from the CA1 LFP, and computes the RMS from the EMG. Both elements are inputs to a trained Gaussian Mixture Model algorithm, which computes the posterior probability of each epoch regarding each one of the three possible clusters. Later, the algorithm uses the visual inspection executed by the user to select the best posterior probability threshold for each state.

Installation

1 - Prerequisites and basic information

Our software was implemented on MATLAB 2020b on Windows 10. Consequently, the execution of some of the routines may present errors when running in MATLAB later versions or on different operational systems.

Pacotes oficiais do MATLAB necessários:
- MATLAB 9.9
- Signal Processing Toolbox 8.5
- Statistics and Machine Learning Toolbox 12.0
- Curve Fitting Toolbox 3.5.12

It is possible to download them separately on: Get and Manage Add-Ons - MATLAB & Simulink

2 - Installation itself

1 - Download all the files from the master branch:

https://github.com/ikaro-beraldo/sleep-stages-classification

2 - Add all scripts, functions, and variable files on MATLAB's 'Search Path' (More details: Change Folders on Search Path - MATLAB & Simulink)

Usage

Run the following command on MATLAB's command window:

```
RMS_pwelch_integrate
```

1 - Main Window

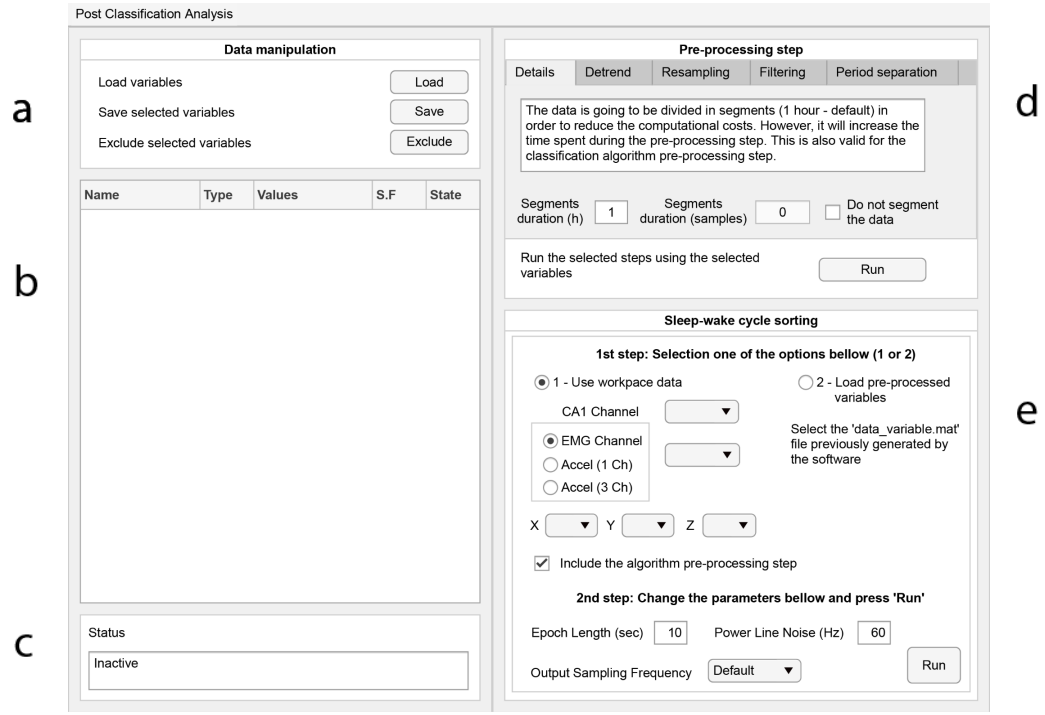


Figure 1 . **Primary interface.** The first window opened during algorithm initialization. **a -** *Data manipulation* panel. It houses options to load (*Load*), save (*Save*), and exclude (*Exclude*) variables. **b -** *Workspace* Panel. It consists of a simulation of the MATLAB workspace, containing basic information about the variables loaded by the user. The information presented is the name (*Name*), class (*Class*), values (*Values*) or dimension, if it is a vector, associated sampling rate (*SF*), and the selection state (*State*) of the variables. **c -** Information about the *status* of the classifier. **d -** *Pre-processing step* panel; **e -** *Sleep-wake cycle sorting* panel. It consists of options for the initialization of the sleep-wake cycle stages classification algorithm.

The first window that opens is the primary interface (**Figure 1**), which consists of separate panels for each of the software's main functions. The *Status* text box (**Figure 1.c**) contains information about the progress of processes launched from the main interface. The texts *Busy* and *Inactive* will be displayed in the box if any process is running or not, respectively.

2 - Workspace

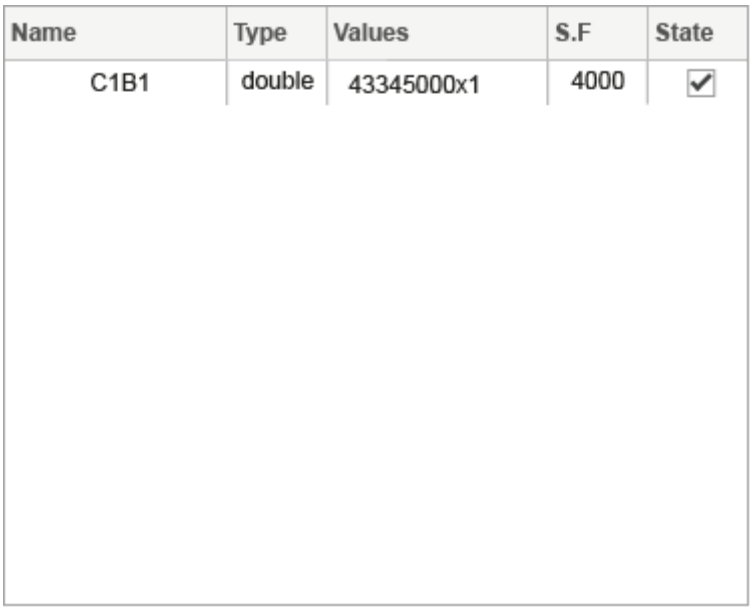


Figure 2. ***Workspace*** - contains information about the loaded variables, such as name (*Name*), class (*Class*), values (*Values*) or dimension if it is a vector, associated sampling rate (*SF*), and the selection state (*State)*.

In the main interface, below the *Data Manipulation panel,* there is the *Workspace* panel (**Figure 1.b** and **Figure 2**), to simulate the MATLAB *Workspace*, being a space to display, in columns, information about the loaded variables, such as name (*Name*), type (*Type*), values (*Values*) or dimension if it is a vector, associated sample rate (*SF*), and the selection state (*State)*. The last column defines which variables will be considered (check box checked) or not (check box unchecked) in the delete, save, or preprocessing processes.

3 - Data manipulation

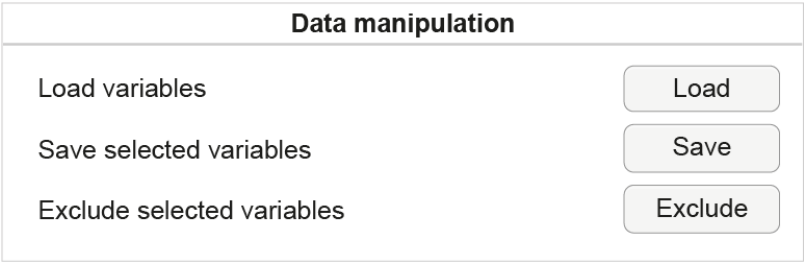


Figure 3. Data manipulation - contains *Load*, *Save*, *Exclude buttons* to perform basic functions.

The *Data manipulation* panel (**Figure 1.a** and **Figure 3**) groups together functions that allow data to be loaded, saved or deleted.

3.1 - Loading variables

By clicking on the *Load button*, a file selection window opens, allowing the user to select files with the MAT extension.

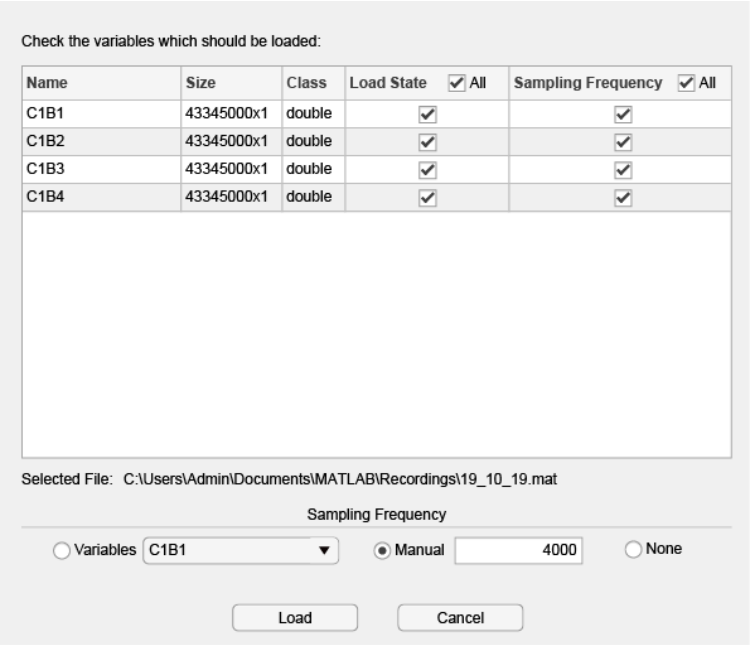


Figure 4. File loading interfaces. Interface for loading variables existing in the file selected by the user. There is a table with basic information about the variables present in the file, as well as a menu of options regarding how the variables will receive a sample rate value.

After selecting the file, a new window opens (**Figure 4**), allowing the user to view all variables stored in the selected file. Also, basic information about these variables, such as name (*Name)*, dimensions (*Size*), and class (*Class)* are displayed in a table, allowing the user to select which variables will be loaded by checking/unchecking the checkboxes present in the column *Load State*. Finally, the *Sampling Frequency column* contains checkboxes that indicate which variables have

an associated sampling rate. The set of radio buttons (*Radio button)* defines how the sample rate value will be associated with the variables, where:

1 - *Variables* indicates that one of the variables in the list contains the sample rate value and the list box on the side allows the selection of the variable in question;

2 - *Manual* indicates that the sample rate value will be manually entered in the edit box on the side;

3 - *None* indicates that no sample rate values will be associated with the variables.

Later, when you click on *Load*, the selected variables will be loaded, whether or not they contain the associated sample rate. On the other hand, by clicking *Cancel*, the loading is canceled and the user will be directed to the main interface.

3.2 - Saving and Excluding variables

When clicking the *Save button*, a directory selection window will open, allowing the user to choose the location where the variables will be saved. However, only variables whose selection state is checked (checked variables) will be saved. Similarly, clicking the *Exclude button* will delete the marked variables from the *Workspace panel and will* also stop taking up memory space.

4 - Pre-processing step panel

Pre-processing step panel (**Figure 1.d**) consists of 5 tabs. Each one of them groups together a set of information related to one of the possible functions that make up the pre-processing step.

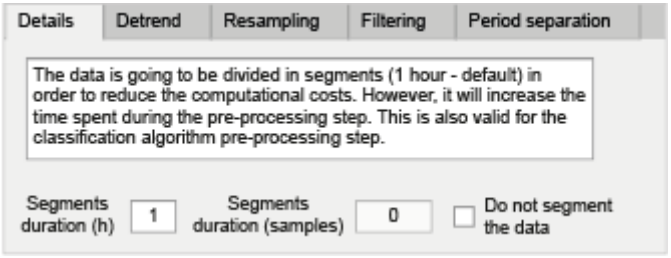


Figure 5. ***Details Panel*** - Opens options for data segmentation

Details panel (**Figure 5**) has the option to pre-segment recordings into blocks of defined duration. This option has the function of reducing computational costs during the pre-processing of recordings. The *Segments duration* boxes define the duration of the segments. The option *Do not segment the data* allows the user to allow or not the segmentation of the data.

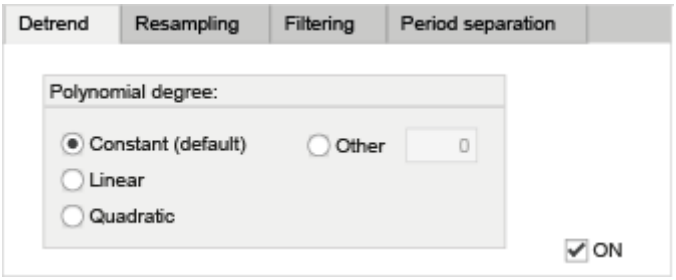


Figure 6. *Detrend* panel. It houses options for selecting the polynomial degree used by the MATLAB *detrend function.*

The *Detrend tab* (**Figure 6**) allows the user to control the polynomial detrending function of the signal. In it, the *Polynomial degree* panel presents selection buttons that modify the polynomial degree used by the function. If the *Constant option* is selected, the polynomial degree is equal to 0; *Linear*, equal to 1; *Quadratic*, equal to 2; case *Other*, allows the user to enter the degree in the text box next to it.

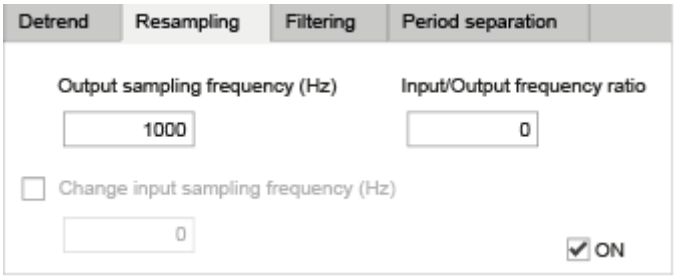
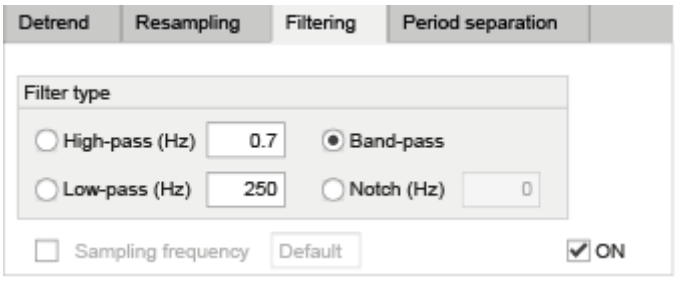


Figure 7. *Resampling panel.* It houses text boxes for defining the final and initial sampling rates that will be used by the *resample function* of MATLAB.

The *Resampling* tab (**Figure 7**) allows the user to control the resampling step. In the *Output sampling frequency text box,* the user can enter the final sampling rate that the data will have after being processed. However, the user can also enter the initial and final sample rate ratio in the *Input/Output frequency ratio text box.* Regardless of the edited text box, the value of the other will be automatically updated. By default, the initial sample rate is the value found associated with the variable in the *SF column of the Workspace* panel. However, the *Change input sampling frequency* checkbox allows the user to enter a different value for the initial sampling rate than the one present in the workspace.



***Filtering* Panel**. It houses filter type options and frequency threshold values for each filter type.

The *Filtering* tab (**Figure 8**) allows the user to control the signal filtering function. In the *Filter type* panel, there are radio buttons that define the type of filter that will be executed: *High-pass*, informs the algorithm that it will be a high-pass filter; *Low-pass*, a low-pass filter; *Band-pass*, a bandpass filter; *Notch*, a notch filter. If the *High-pass options, Low-pass,* or *Notch* are selected, only the text box referring to each one of them will be enabled, allowing the user to inform the cutoff frequency for the filter. In the case of the *Band-pass option*, the text boxes referring to *High-pass* and *Low-pass* will be enabled, containing the lower and upper limit frequencies, respectively. Also, in the same way, the sampling rate of each variable is defined by the *SF column* of the workspace, however, it can be reset by checking the *Sampling Frequency checkbox* and adding a new value in the text box next to it.

Figure 9 - ***Period separation panel****.* Contains options for the final duration of epochs after raw data separation.

Finally, the *Period separation tab* (**Figure 9**) controls the separation of data into epochs of defined duration. The *Period Length* text box is the only one that can be edited, setting the duration of the epochs in seconds. The *Period Length* text box indicates the duration of epochs as the number of samples, the *Number of Periods text box* indicates the total number of epochs, and the *Number of discarded samples text box* indicates the number of discarded samples. The sampling rate of each of the variables is defined by the *SF column of the workspace but can be reset if the user selects the Sampling frequency* checkbox.

ON checkbox, which defines the function present in the tab as active and will be executed when the *Run button* is pressed. When you clear the check box, the text *OFF* will appear as the answer. Also, only the variables whose *State* column checkbox is checked will be preprocessed when the *Run button* is pressed. The pre-processing order follows the order of the tabs, and as soon as each of the steps is finished, a warning will be presented to the user.

5 - Sleep-wake cycle sorting panel

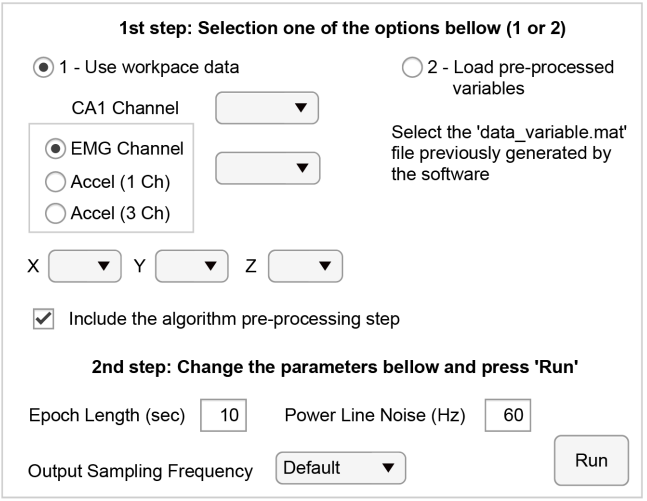


Figure 10 - *Sleep-wake cycle sorting panel*. It consists of options for the initialization of the sleep-wake cycle stages classification algorithm.

In the *Sleep-wake cycle sorting* panel (**Figure 10**), there is a group of radio buttons that define the origin of the data that will be used by the sleep-wake cycle stage sorting algorithm.

- Option 1 - *Use workspace data* defines that the data will come from the *Workspace* area, requiring the user to select the CA1 LFP and electromyogram (or accelerometer) recordings through the *CA1 Channel* and *EMG Channel list* (or *Accel 1*) *boxes*, respectively.
    - Obs: If the user is using 3 separate channels for the accelerometer, it is necessary to select the option *Accel 3* and select each one of the channels in the *X, Y,* and *Z* list boxes.

  All of the list boxes are automatically updated with the names of the variables present in the workspace. Furthermore, the check box *Include the algorithm pre-processing step* defines whether pre-processing will be (checked) or not (unchecked) performed automatically according to the patterns that best fit the classification algorithm. In this case, the pre-processing panel Pre-processing step doesn't need to be executed.

- Option *2 - Load pre-processed variables* defines that the classification algorithm will use a file named *data_variables.mat*, which contains the data originated by the pre-processing performed by the classification algorithm. Therefore, this option is only valid if the classification algorithm has already generated the file (*data_variables.mat*) for the dataset in question. Finally, by clicking the *Run button*, if option *1* has been selected, the variables selected in the list boxes will be transferred to the classification algorithm. If option *2 has been selected, a file selection window will open, asking the user to select the corresponding data_variables.mat* file.

The value of the *Power Line Noise box* defines the frequency of noise coming from the power line (in Hertz). The *Output Sampling Frequency list* defines the output sampling rate of the data if the user selects to include the default pre-processing step of the algorithm. Finally, the *Epoch Length edit field* defines the epoch length.

6 - Classification parameters

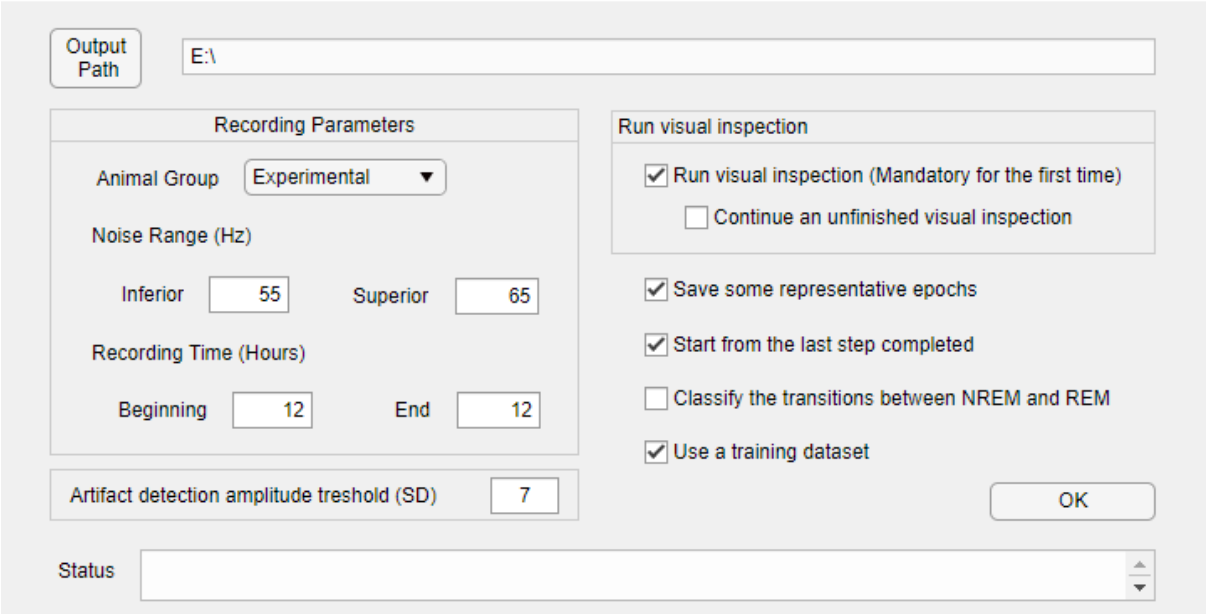


Figure 11 - Information input interface for the classification algorithm. There is a set of options for obtaining information about the electrophysiological recording used, as well as information about the output directory for figures and final results. There is also a set of options that define the behavior

of the algorithm, modifying the number of figures produced and the execution of the visual inspection. Finally, the *Status bar* indicates to the user the sequence of actions performed by the algorithm.

Subsequently, after the pre-processing of the data is completed, a new window will be opened (**Figure 11**). It is the main form of communication between the classification algorithm and the user since it is through it that the execution of the algorithm will be controlled. As well as the main interface, there is also a *Status text box*, which contains texts indicating processes being executed by the algorithm.

In addition, there is the *Output Path button*, which when pressed, opens a directory selection window. In this case, the selected directory will be the destination for all data and figures produced by the classification algorithm. The output path is presented in the text box next to the button. The *Recording Parameters* panel includes information about the recording provided by the user: the *Animal Group list box* presents the *Control* and *Experimental options*, defining whether the recordings were performed on control or experimental animals, respectively; the text boxes in the *Noise Range group* define the minimum (*Lower*) and maximum (*Upper*) threshold, in Hertz, for the possible network noise present in the signal; and the *Recording Time box set* define the start time (*Beginning*) and end time (*End*), in hours, of the recordings.

The *Artifact detection amplitude threshold option* sets the cut-off threshold for artifact detection as the number of standard deviations. The default value is 7.

The *Save some representative epochs check box* defines whether, after the classification is finished, the algorithm will save (checked) or not (unchecked) figures with representative periods of each of the states of the sleep-wake cycle. In the *Run visual inspection* panel, the Run visual inspection checkbox defines whether a *visual inspection* step by the user will be (checked) or not (unchecked) performed. Finally, the *Continue an unfinished visual inspection checkbox* defines whether the user wants (checked) or not (unchecked) to continue an unfinished visual inspection.

The *Start from the last step completed option* allows sorting to resume from the last step performed.

The *Use a training dataset option* enables the GMM algorithm to be trained when clustering the dataset given by the user. It is recommended to enable this option.

The sorting algorithm is idle until the *OK button* is pressed, causing all buttons, text boxes, checkboxes, and list boxes to be locked, preventing the user from modifying this information. Afterward, the algorithm will continue its execution, updating the *Status text box* as needed.

All figures generated by the algorithm are created in the background and automatically saved in the directory specified as the destination when pressing the *Output Path button*.

7 - Adding new frequency bands

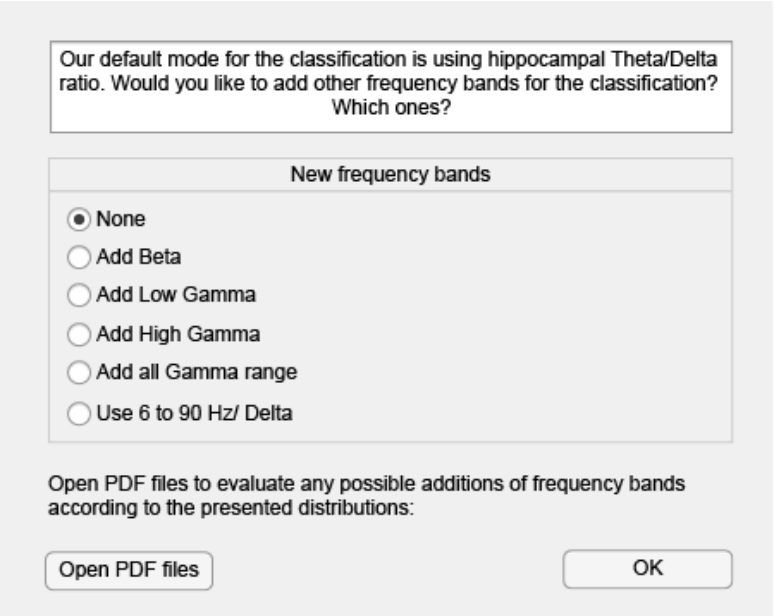
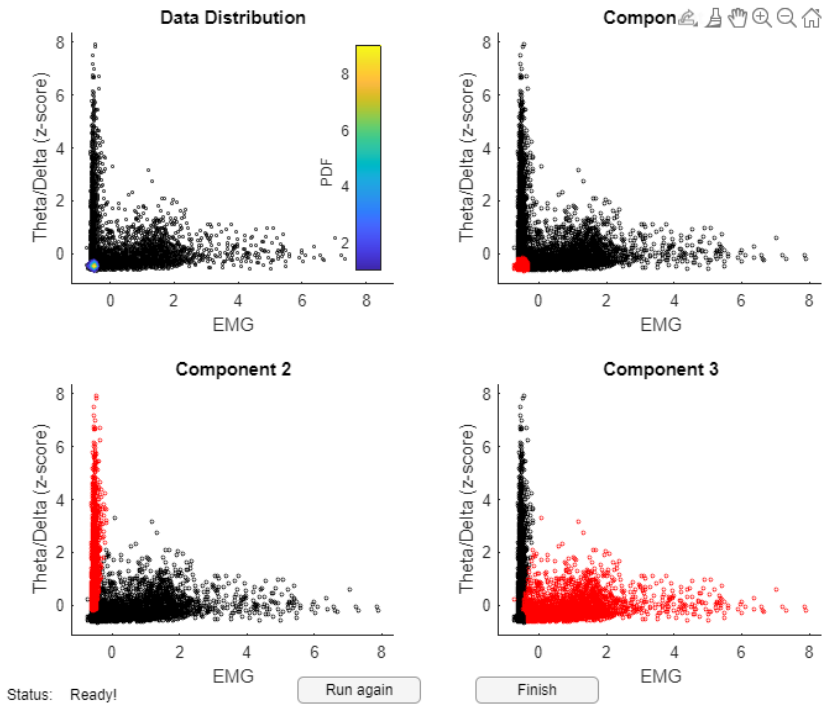


Figure 12 - Frequency band selection interface. It consists of a set of options for selecting the frequency band that will be added to the Theta/Delta ratio. There is a feature for viewing PDF figures to assist the user in his decision, activated by the *Open PDF files button*.

Afterward, after the three first figures have been generated, another window will be opened (**Figure 12**). In it, the user will indicate which frequency band will be added to the Theta/Delta ratio during classification. However, since it is recommended that the user evaluate the distribution and the scatter plots present in the figures *Frequency bands distribution*, *Frequency bands distribution over time,* and *Frequency bands combined*, it is possible to click on the *Open PDF files button*, which will open the figures files using the system's default PDF file reader. Later, when deciding which frequency band is going to be added, in the *New frequency bands* panel, the user can select the radio button for their choice. Finally, the *OK button* will close the window and save the user's selection, adding this information to the execution of the algorithm, which is resumed.

8 - Selection of preliminary clusters



If the user has not enabled the option to use a training dataset, a window presenting preliminary clusters will be opened. The user can run the GMM algorithm again by pressing the *Run again button* until a suitable set of clusters is presented.

Posteriorly, the user can press the *Finish button*, to accept the current clusters and resume the classification.

9 - Visual inspection

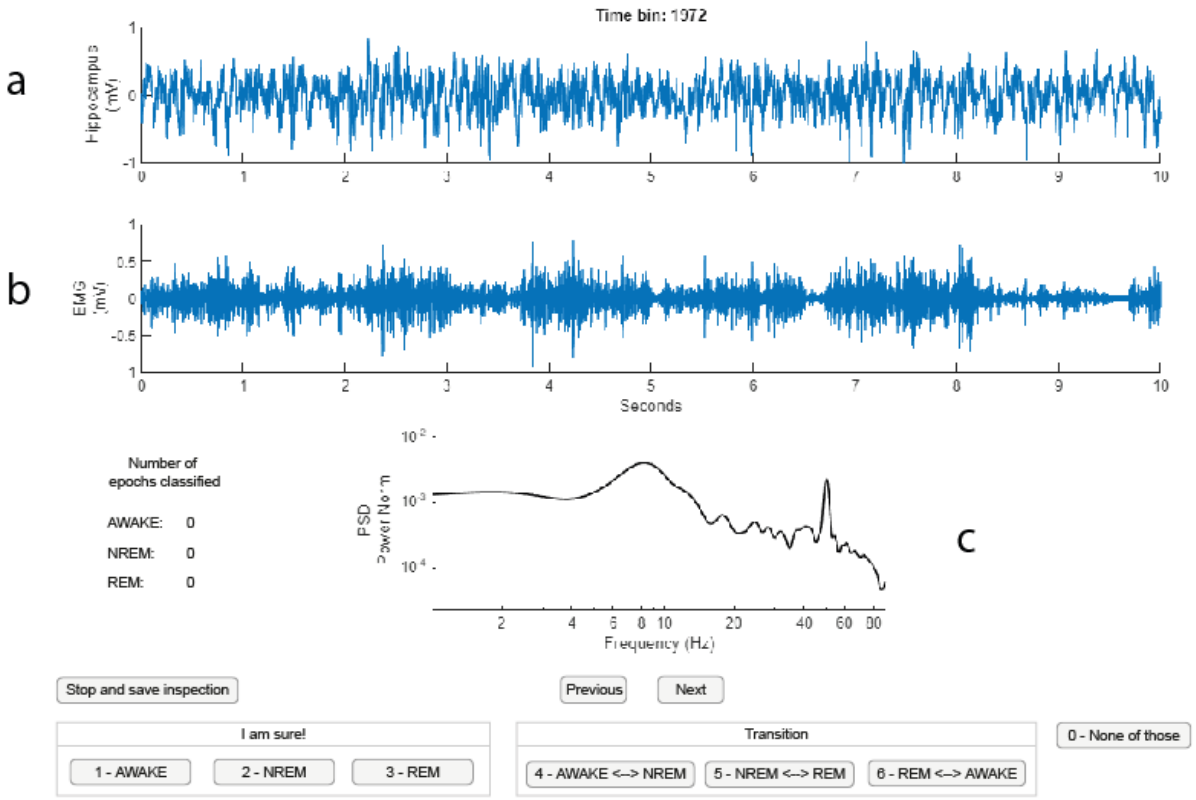


Figure 13 - Visual inspection interface. Interface initialized during the visual inspection phase with the function of presenting information about the 10-second epoch being classified and obtaining the user's classification. The buttons control the algorithm's response to user commands. **a -** LFP of CA1. **b -** Filtered electromyographic record. **c -** Graph of the power spectral density of the LFP signal of CA1.

Run visual inspection checkbox was selected, as soon as the visual inspection is required (as demonstrated by the text in the *Status box*), a new window will be opened (**Figure 13**). This window is the interface that allows the user to classify a set of random n-second epochs from the recording, which will later be used to define the posterior probability thresholds for each of the stages of the cycle. At a given

moment, all the information and graphics present in this window refer to a single n-second epoch of the recording, however, according to the user's action, this epoch can be changed and all information and graphics are updated. The window has a title with the information of the epoch number presented (*Time bin*), a graph of the LFP signal (**Figure 13.a**) of the hippocampus (*Hippocampus*), and a graph of the signal (**Figure 13.b**) of the electromyogram (*EMG)*) referring to the n-second epoch. There is also a PSD graph (**Figure 13.c**), with the two axes (Frequency - *Frequency;* PSD) on a logarithmic scale.

When an n-second epoch is presented, the user can classify it, indicating which stage of the sleep-wake cycle it belongs to. Pressing the button:
1 - AWAKE, the epoch is classified as part of the agreed state set;
2 - NREM, like NREM sleep;
3 - REM, like REM sleep. If you think it is a transition period, the user can press
4 - *AWAKE ← → NREM*, indicating that the epoch is a transition between the awake state and NREM sleep;
5 - NREM ← → REM, between NREM sleep and REM sleep;
6 - REM ← → AWAKE, between REM sleep and waking state.

If the user judges that the period in question does not belong to any of these classifications, press the *0 - None of those* button. After pressing any of these buttons, the next epoch is loaded and all graphs are updated. Furthermore, the user can go back to the previous epoch by pressing the *Previous button* or advance to the next one by pressing the *Next button*.

The *Stop and save inspection button* allows the visual inspection to be ended and saved in the previously defined directory.

When the user completes 40 (approximately) classified epochs for each of the 3 stages of the sleep-wake cycle, a warning indicating the end of the inspection is presented, the data referring to the visual inspection is saved, the visual inspection interface is closed and the Execution of the classification algorithm resumes.

The information input interface for the classification algorithm (**Figure 11**) is kept after the visual inspection is finished, as it still maintains the function of indicating the processes that are running through the *Status text box*. However, at this stage, it is not necessary for the user to interact with the interface, since the algorithm is running in the background. Finally, when sorting is finished, the *Status text box* will display the message *Done*, and after 5 seconds the window will close, keeping only the main interface open.