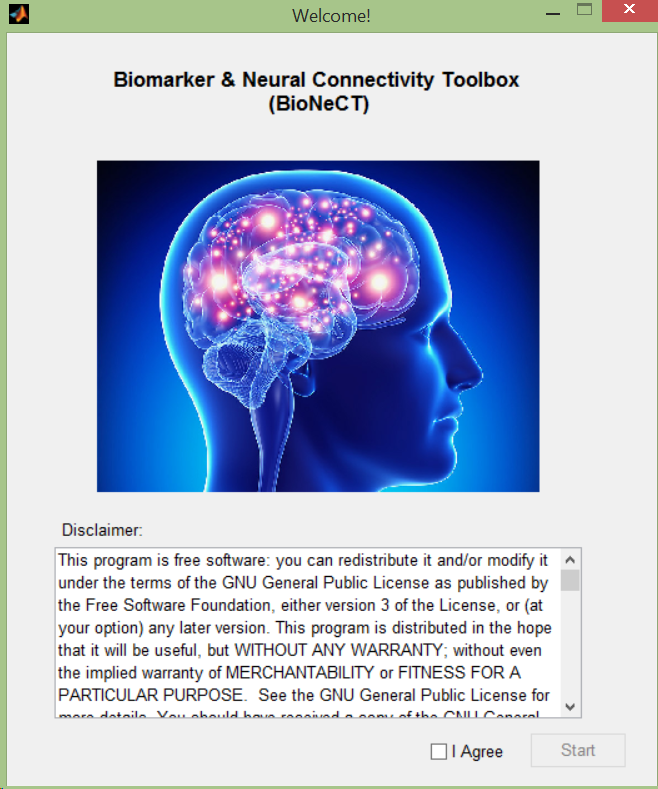
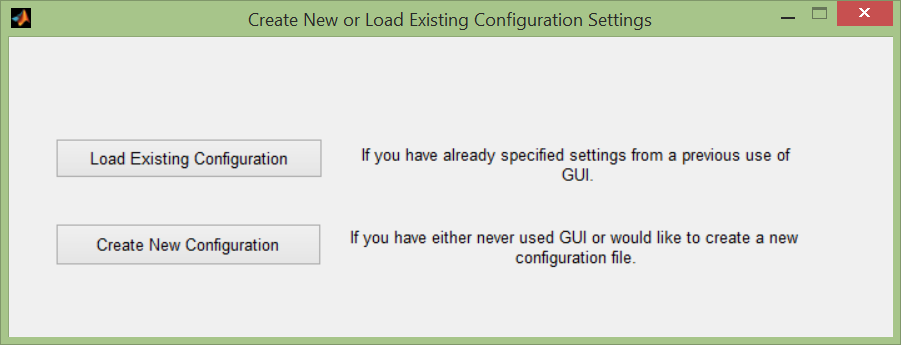
Biomarker and Neural Connectivity Toolbox (BioNeCT)

**2. Getting Started with BioNeCT**

To begin, type *bionect* in the Matlab command window. This will open the welcome page of the toolbox.



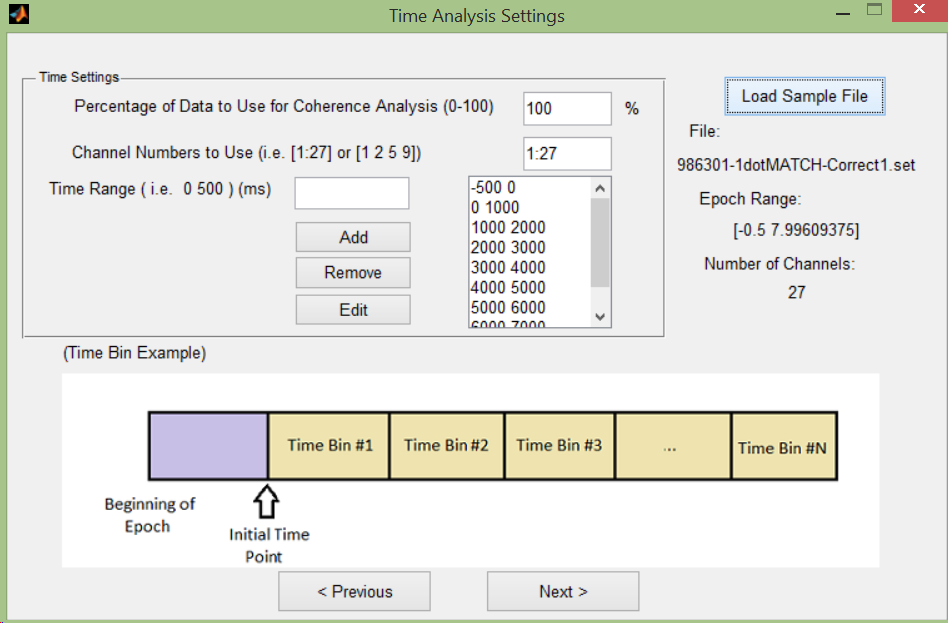
After agreeing to the Terms and Conditions and pressing Start, the following window should appear:



For new users, or those wishing to analyze data from a different study, choose **Create New Configuration**. Alternatively, if you have already used BioNeCT before and have a study/analysis available to load, choose **Load Existing Configuration**.

**2.1 Creating an Analysis**

If you chose to create a new configuration, a Time Analysis Settings window will appear:



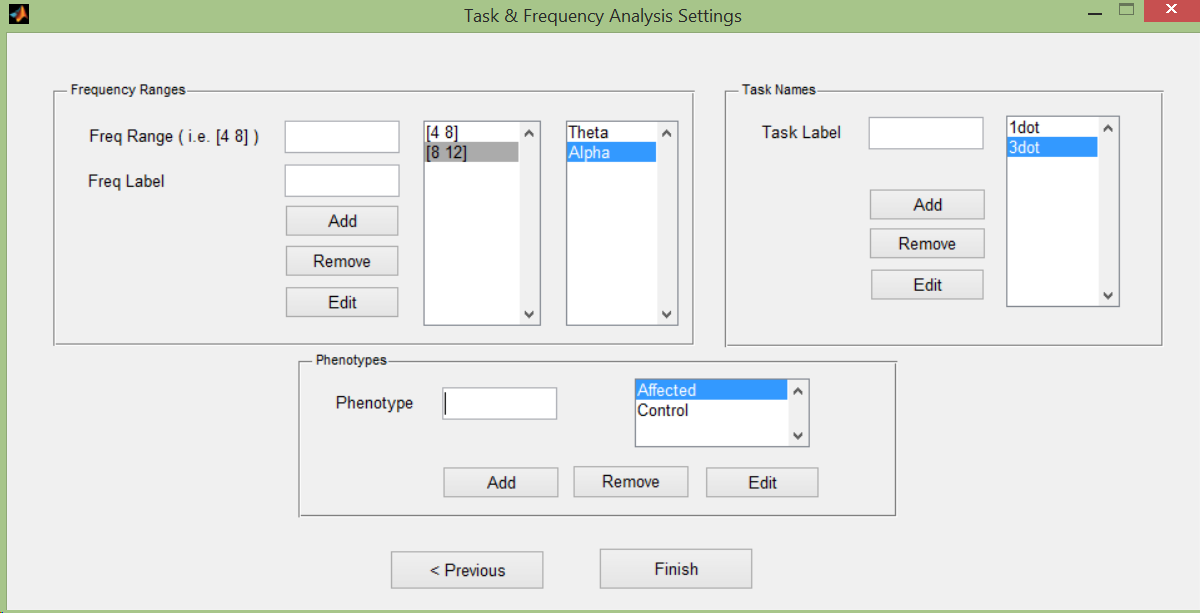
Here you can specify which percent of your data, channels, and time ranges in which to perform coherence calculations.

*Percentage of Data to Use for Coherence Analysis:* This should be a value between 1 and 100, and represents the percentage of the data that will be used for coherence calculations.

*Channel Numbers:* The channels which will be used for coherence analysis. This can be either a range of channels (i.e. 1:27) or a list of channels (i.e. 1 2 3 5 8 10 …etc). Note that using too few channels may affect future graph measure results due to small connectivity matrices.

*Time Range:* Individual time bin ranges (in milliseconds) in which to calculate coherence. The time range limits of your data can be seen by the **Load Sample File** button. Note that smaller time ranges can negatively affect coherence calculations in lower frequency ranges.

Once complete hit **Next**, and a Task/Frequency Settings window should appear:



Here you can specify which frequency ranges to perform coherence analysis on, along with what the task/condition (i.e. eyesopen) and phenotype (i.e. control) names should be.

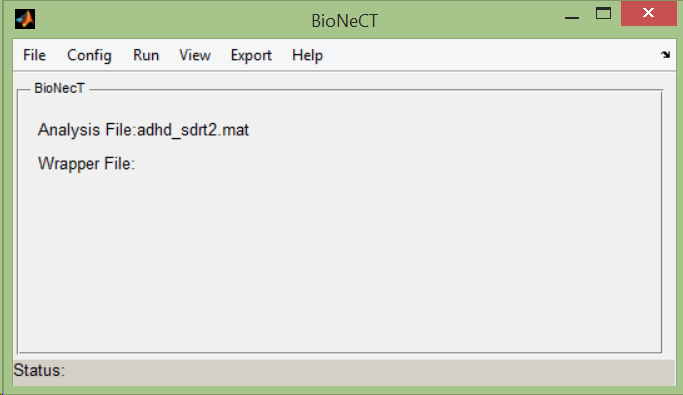
Note: Labels should not be entirely numerical or consist of non-alphabetical characters.

* *Frequency Range:* Specify beginning/end of range (i.e. 4 7).
* *Frequency Label:* Frequency name to be associated with the respective range.
* *Task Label:* Name of condition/task that you want it to be referred to as (EyesOpenRest, EyesClosedRest, etc)
* *Phenotype:* Phenotypes of the respective data groups

Once complete hit **Finish**. You will be asked to save the analysis settings to a .mat file. This will be the file you load in the future when coming back to do any analyses with these same settings.

**3. Main GUI**

After saving your new analysis or loading a previous analysis, the main BioNeCT window will appear.



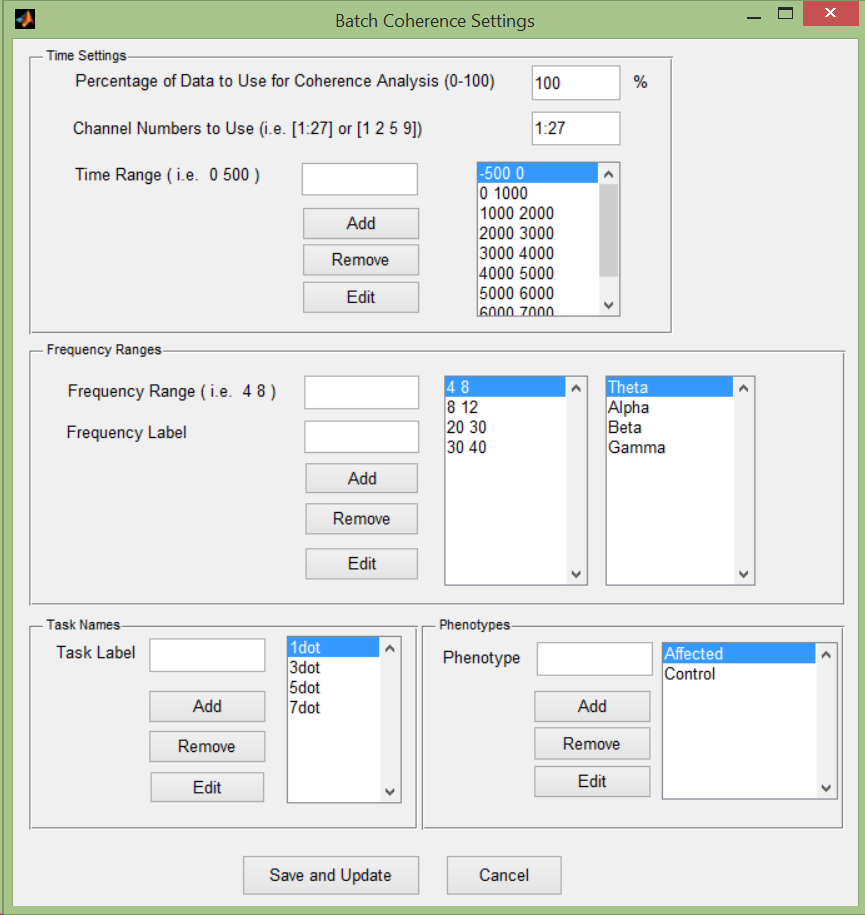
The main BioNeCT window has menu tabs **File, Config, Run, View, Export, and Help**, all of which will be discussed later in this tutorial. The window also displays a few bits of information:

*Analysis File*: Current analysis file being worked on.

*Wrapper File*: Batch file for all of EEG data (discussed shortly in another section)

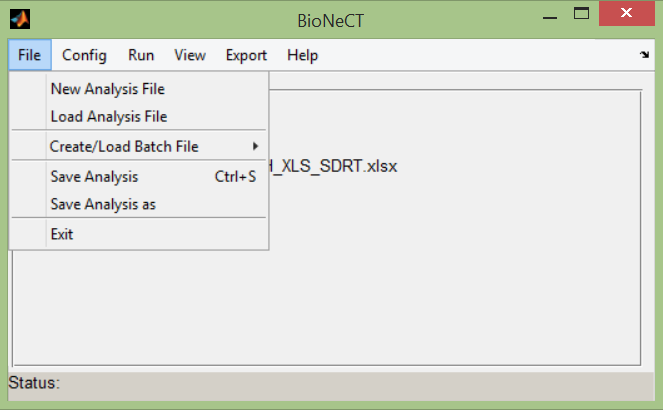
*Status:* Updates according to recent actions performed.

Note: *You can double check or edit any of the analysis settings previously set by selecting Config > Configuration Settings in the menu, which will pop up the following window:*



**4. File Menu**

Let’s take a look at the **File** menu:



There’s a few options here:

*New Analysis File*: Clear the calculations (graph theory, spectral analysis, etc) from the analysis file for a fresh start, saving the configuration settings (time ranges, frequency ranges, etc).

*Load Analysis File*: Load another previously created analysis file.

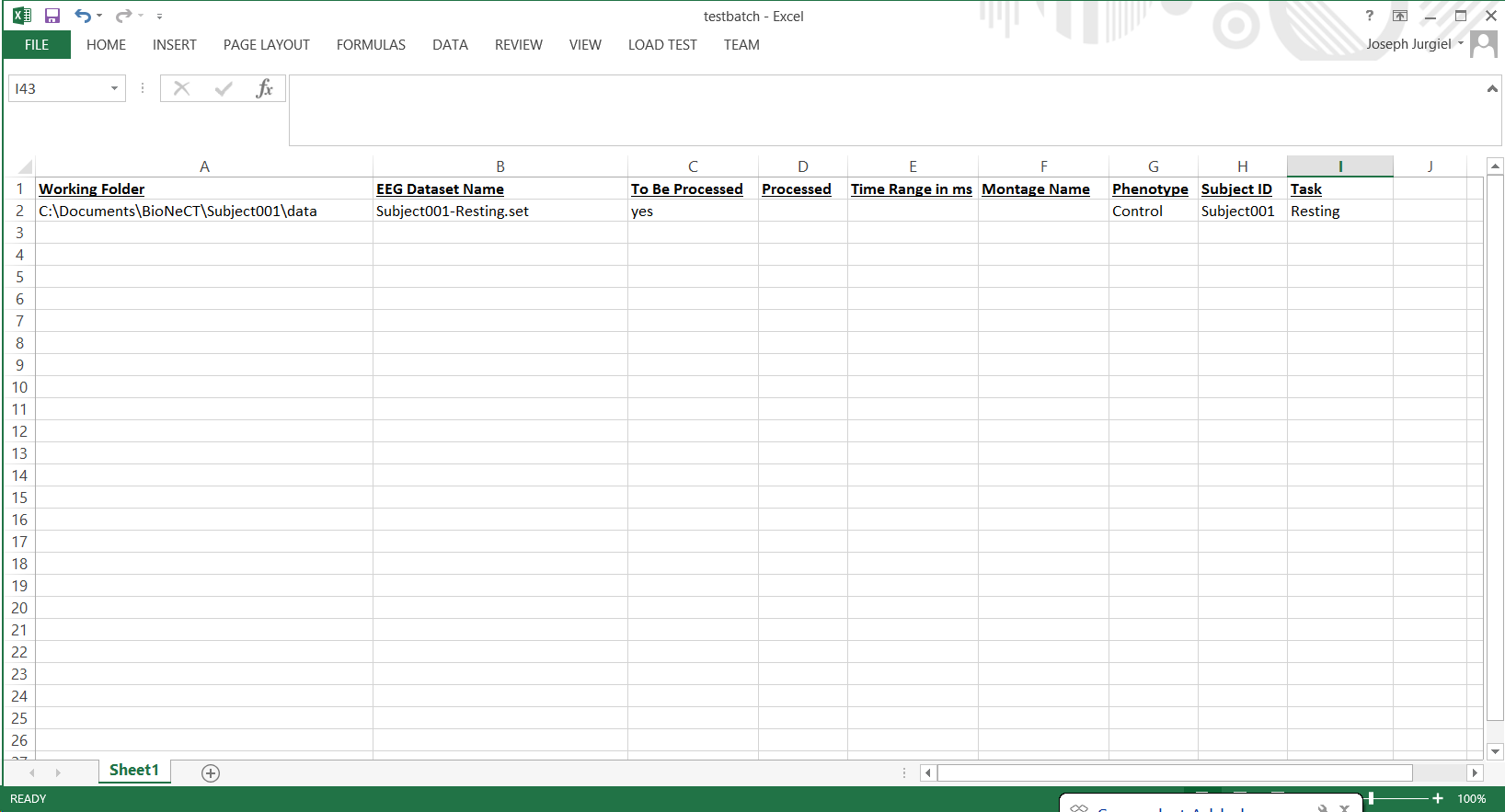
*Create/Load Batch File*: Create or load a previously created .xlsx file which contains file name and folder path information for all data to be used in the analysis.

*Save Analysis*: Save any calculations (graph theory, spectral analysis, results, etc) to the analysis file.

*Save Analysis As:* Save analysis under a different name.

**4.1 Creating & Loading a Batch File**

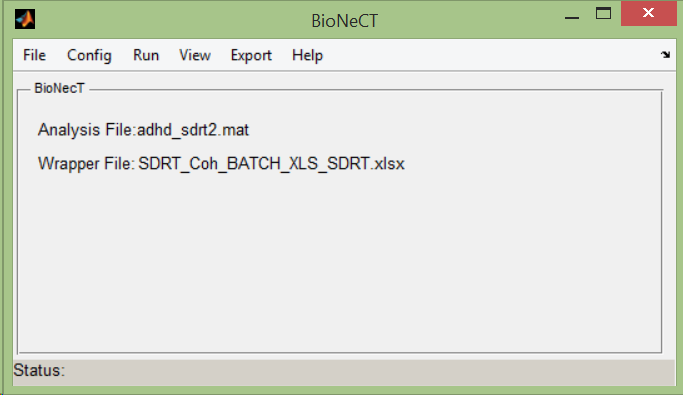
Let’s create a batch file from *File > Create/Load Batch File > Create Batch File*. This will open up an excel file that looks like this:



Each row in the batch file corresponds to a given dataset file. The excel file has several columns that the user should fill out, including the Working Folder, EEG Dataset Name, To Be Processed, Phenotype, Subject ID, and Task:

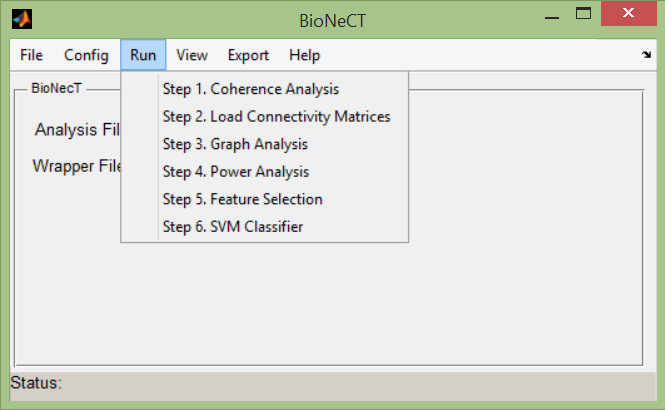
* *Working Folder*: Folder where the EEG dataset is located. This is also where the resulting coherence calculations will be stored.
* *EEG Dataset Name*: Name of the .set file.
* *To Be Processed*: Whether or not you want to run coherence analysis on the respective file, denoted by *yes* or *no*. For example, if you have already run coherence analysis on all subjects and decide to add a few more, you can write *no* in this column for the subjects already processed in order to save time.
* *Processed*: Automatically filled by script with a 1 if coherence analysis was successful on the respective dataset. Cell will be empty if analysis was unsuccessful.
* *Time Range in ms*: N/A
* *Montage Name*: N/A
* *Phenotype/Subject ID/Task*: Phenotype/Subject ID/Task of the respective dataset. Phenotype and Task should match those entered into the Batch Coherence Settings, and may not be entirely numerical or consist of non-alphabetical characters. Subject ID may be numerical or alphabetical.

Once the spreadsheet has been filled out and saved, go to **File > Create/Load Batch File > Load Existing Batch File**. This will update the batch file name on the front of the GUI:



**5 Run Menu**

Shown below is the *Run* menu. This menu contains all of the analysis functions that can be performed on the datasets.



**5.1 Step 1. Coherence Analysis**

Running coherence analysis requires the previous steps:

* Creating and loading a batch file as described previously
* Setting appropriate configuration settings in *Config > Configuration Settings* (% data to use, channels, time ranges, and frequencies)

Coherence analysis can be performed by **Run > Coherence Analysis**. Once finished, there will be a unique .mat file created with the format of (EEG Dataset name from batch file)\_CohValue.mat, placed into the Working Folder for the respective file, for each listed in the batch file spreadsheet. *Connectivity matrices* (which are NxN channel matrices of coherence values between all pairs channels) are placed in the raster\_mat variable of the .mat file in a 4-D matrix with form *#Channels x #Channels x Time Range x Frequency*.

For example, if we have 40 channels, 5 time ranges, and 3 frequencies, the size will be 40x40x5x3, meaning we now have 15 40x40 connectivity matrices in total. Each cell in the 40x40 matrix represents the coherence between the respective pair of channels.

**Notes:**

* Running Coherence Analysis can be very time consuming and may take several hours to run and scales based on number of channels, time ranges, frequencies, and data trials.
* Once you have ran coherence analysis once, you will not have to run it again unless you want to analyze different channels, time ranges, or frequencies.

**5.2 Step 2. Load Connectivity Matrices**

Required previous steps:

* Running coherence analysis

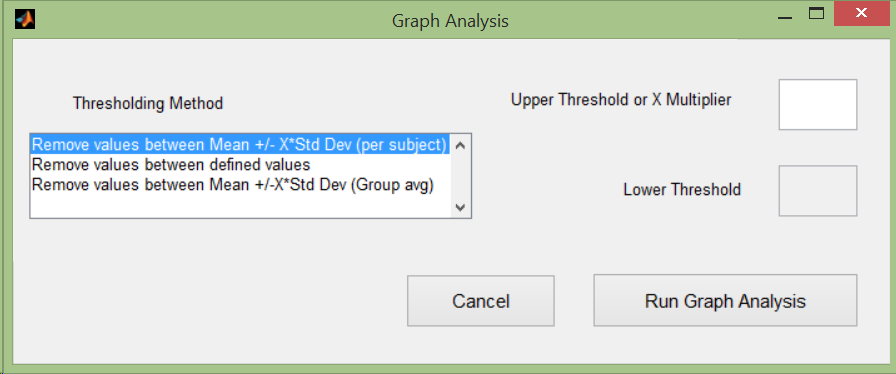
After coherence analysis has been performed and *(EEG Dataset)\_CohValue.mat* files have been created, **Run > Load Connectivity Matrices** will load the connectivity matrices for all files that successfully ran through coherence analysis into Matlab.

**5.3 Step 3. Graph Theory Analysis**

Required previous steps:

* Loading connectivity matrices

Upon selecting **Run > Graph Theory Analysis**, the following window should appear:



Here you can choose which thresholding method (i.e. the way we decide to either keep or get rid of certain coherence values) you would like to use on the data, as well as the magnitude of thresholding for the given method.

Let’s say we have a file with coherence analysis run for 5 time ranges and 3 frequencies, meaning 15 connectivity matrices for this file (one for each time range, and 5 time ranges for each frequency).

* *Remove values between Mean +/- X\*Std Dev (per subject):* This method will average connectivity matrices of all time ranges of each frequency band for a subject/file. So in our example, it will find an average connectivity matrix for the 5 time ranges in a given frequency range. From this average connectivity matrix, it will find the mean and standard deviation, and then apply these threshold values to the original matrices in the respective frequency band.
* *Remove values between defined values:* Apply specified threshold values to each individual connectivity matrix. For example if you choose .3 as the lower bound and .7 as the upper bound, values in the matrix between .3 and .7 will be removed.
* *Remove values between Mean +-/ X\*Std Dev (Group avg):* Going along with the example above, this method will find an average connectivity matrix for the 5 time ranges in a given frequency, but now across all subjects in a respective phenotype. Thus the mean & standard deviation threshold values are calculated from 5 time ranges x # of subjects in phenotype, then applied back to these matrices.

Note: *These methods can result in vastly different connectivity matrices, which can alter graph theory measures. Thus it is best to experiment with different values to observe how many connections are left over after thresholding. An easy way to visualize these matrices are using the View > Connectivity Maps option in the menu, which will be discussed later.*

**If thresholding with standard deviation, a multiplier factor less than 1 is typically suitable. This number should be experimented with to obtain the connection density desired.**

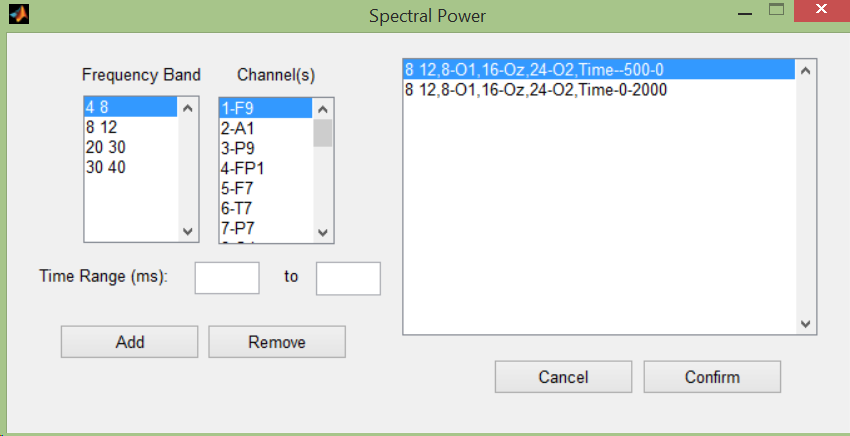
After hitting **Run Graph Analysis**, graph measures will be calculated for each connectivity matrix loaded in Step 2.

**5.4 Step 4. Spectral/Power Analysis**

Required previous steps:

* Loading connectivity matrices

After selecting Run > Power Analysis, the following window should appear:



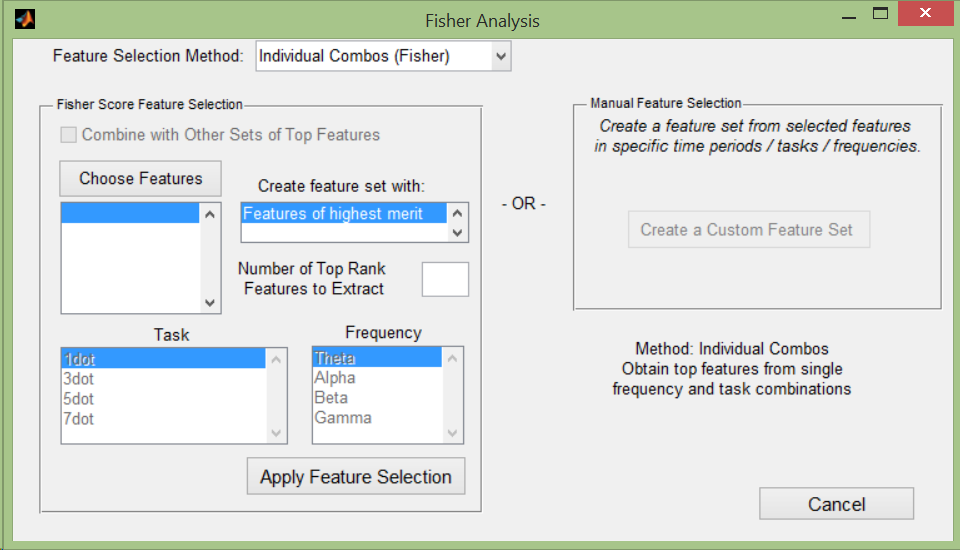
Here you can select the channels (CTRL/SHIFT + Left Click to select multiple channels), frequency range, and time range over which to calculate both average and relative spectral power for the selected ranges.

**5.5 Step 5. Feature Selection**

Required previous steps:

* Calculating graph theory measures *and/or* calculating spectral powers

Once you have calculating either graph theory measures or spectral power measurements (or both), you can perform feature selection via *Run > Feature Selection*, where the following window should appear:



First begin by choosing a Feature Selection Method in the dropdown list at the top. After selecting a method, the appropriate lists/buttons which correspond to the selected method should be enabled.

There are a few options to choose from:

* **Individual Combos (Fisher):**

Create an initial feature set of graph/spectral measures from single combinations of frequency and task. This initial feature set will have number of columns corresponding to *(# Time Ranges) x (# features from Choose Features (below)),* and number of rows corresponding to the number of data files associated with a given task.

From these feature sets, a selected number of features will be chosen via Fisher Score feature extraction for each set, and placed into their new respective feature sets. The Fisher Score method will rank independent features based on within-group and between-group variability of the data. Thus a set of data which is highly separable between groups, and compact within a group, will have a higher fisher score and will be selected for classification use.

After selecting this method, click **Choose Features** to select which graph theory/spectral power measures to use to create their initial feature sets.

You should also specify the number of top ranking features to extract via Fisher Score in the **Number of Top Rank Features to Extract** box. It is recommended to try out different numbers of Top Ranked features, as they may give different classifier accuracies depending on the data. Too few feature may not train the classifier well enough, whereas too many features may overtrain the classifier with less differentiable features.

Click **Apply Feature Selection** to perform the Fisher Score Analysis.

* **Combine freqs/tasks (Fisher):**

Uses the same Fisher Score method as Individual Combos, but instead of single combinations of frequency and task for the starting feature set to extract features from, the user can choose which combination of tasks and frequencies to use as the starting feature set. This initial feature set will have number of columns corresponding to (*# features from Choose Features (below)) x (# Time Ranges) x (# of frequency ranges + tasks selected),* and number of rows corresponding to the number of data files that are available for the selected tasks. For example, if you select 10 graph theory measures to use, 5 time ranges, 3 frequencies, and 2 tasks, the initial feature set will have 10x5x3x2 = 300 columns of features. From this feature set, the top ranking features can be extracted via Fisher Score.

After selecting this method, click **Choose Features** to select which graph theory/spectral power measures to use to create their initial feature sets. You should also specify the number of top ranking features to extract via Fisher Score in the **Number of Top Rank Features to Extract** box.

Lastly, choose which combination of tasks/frequencies you would like to use for your initial feature set via the labeled list boxes at the bottom. You’re free to select as many tasks/frequencies as are available by left clicking on one and *CTRL + Left Click* to select individually or *SHIFT + Left Click* to select multiple at once.

*Note that the initial feature set will consist only of data which is present for all tasks selected, as to keep track of which feature values correspond to which data file. For example, if two tasks are selected in the list box, but there is no data for Subject1 in one of the tasks, Subject1 will be not be included in the feature set.*

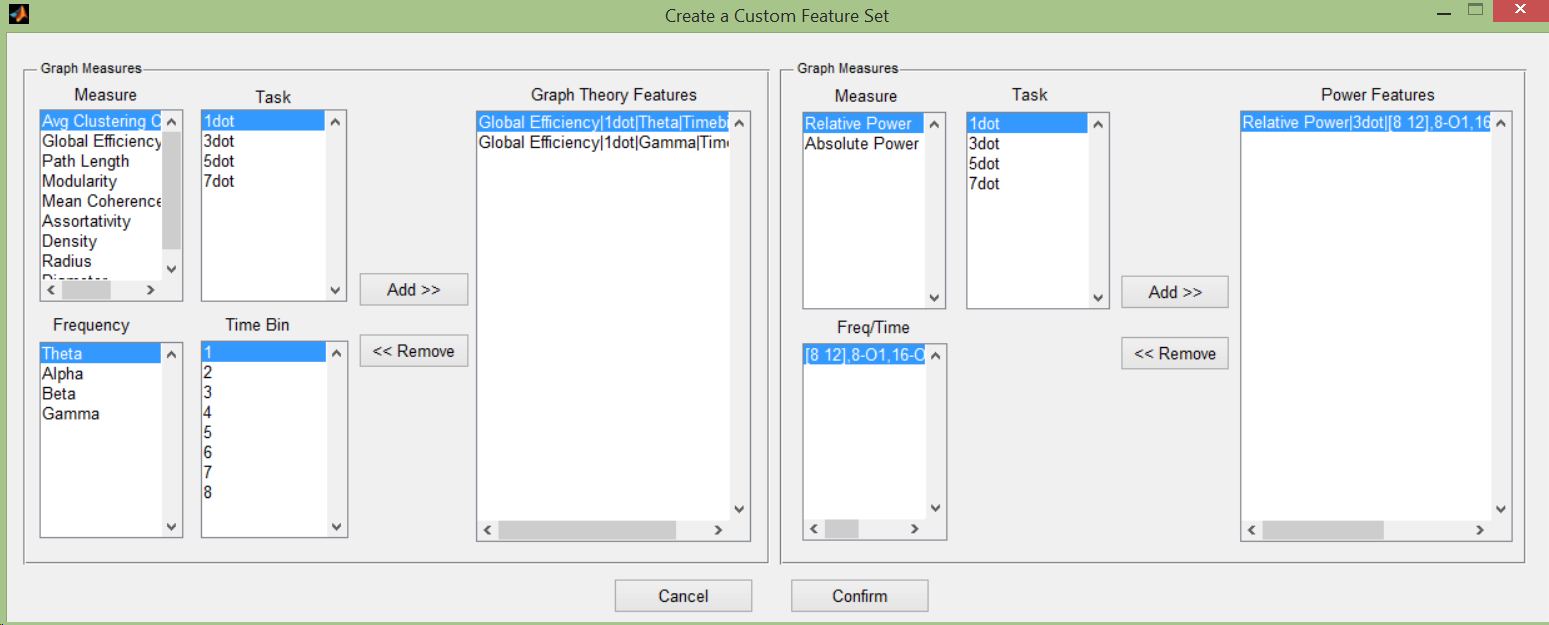
Once finished, click **Apply Feature Selection** to perform the Fisher Score Analysis.

Also available is a **Combine with Other Sets of Top Features** checkbox, which allows the user combine the Fisher scored feature set obtained above with another Fisher scored feature set that was obtained using a different combination of tasks and frequencies. To do this, simply run the feature selection once as described above, then check **Combine with Other Sets of Top Features**, and repeat the feature selection method with preferred features, number of top features, and tasks/frequencies.

* **Create a Feature Set (Manual):**

This method allows the user to manually select features to use to create a custom feature set. For example, the user may be interested in using certain graph theory measures in a certain time and frequency range as their feature set, regardless of how the Fisher Score may rank them.

After selecting this option, the following window will appear:



Here you are free to select specific, individual graph theory and/or spectral power features that have been calculated in previous steps.

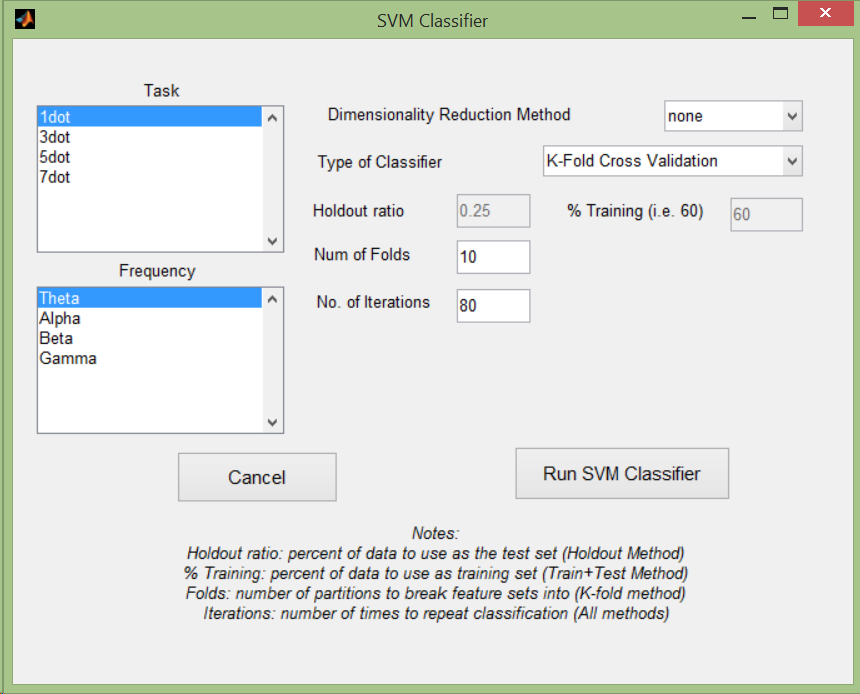
*Note there is no Fisher Score feature extraction here; the feature set simply consists of the features selected in the panel above.*

**5.6 Step 6. SVM Classifier**

Required previous steps:

* Feature Selection

SVM Classifier will take in the feature set(s) obtained through the previously selected feature selection method, and output a classification accuracy (i.e. how well the classifier can discriminate between the phenotype groups using the feature sets obtained). After selecting **Run > SVM Classifier**, the following window should pop up:



First begin by selecting the **Type of Classifier**. Different options will either be enabled or disabled based on which type of classifier you choose to use.

There are (3) types of classifiers currently:

* *Holdout Method:* Randomly partitions the subjects into a training set and test set, where the Holdout Ratio is the percentage of subjects used for the test set (10-20% recommended).
* *K-fold Cross Validation:* Randomly partitions the subjects into K groups (specified by **Num of Folds** user input), where K-1 groups are used for training the classifier, and the remaining group is used to test the classifier. This process is repeated K times such that all groups are used as the test set once.
* *Train + Test Set:* Randomly partitions subjects into a training and test set, where the % Training is the percentage of subjects to be used to train the classifier.

As these classifiers randomly select which data to use as the training and test sets, accuracy results can vary depending on which subjects are placed into which sets. **No. of Iterations** defines the number of times to repeat the classification, where after the classifier accuracy results are averaged for a final result.

Depending on which feature selection method was used, the **Task and Frequency list boxes** may be enabled or disabled. If the Individual Combinations method was used, here you can specify which single combinations you would like to input to the classifier one at a time. If *Combined Freq/Task* or *Manual* feature selection were used, this option is disabled, since there is only one available feature set to use.

The **Dimensionality Reduction Method** dropdown box has an option for further feature reduction (i.e. PCA). However, note that if PCA is used, the features before PCA is used can’t be described by the original features selected.

**Features**

**Subjects**

**Frequency/Task Combinations**

**The suggested setup is using K-fold Cross Validation with # of folds = 10, # of iterations = 100, and no dimensionality reduction.**

Note: After running the classifier, the results, along with features, and configuration parameters selected will be output to a log text file with name (analysis file name)\_Logfile\_(data & time).txt.

**6. Exporting**

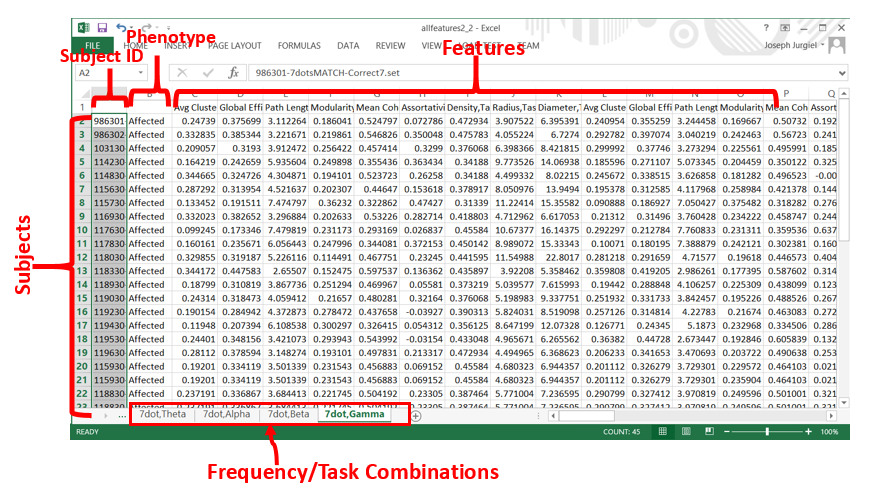
Previous required steps:

* Feature Selection

**6.1 Exporting Features**

If you would like to save your feature sets to an excel file for something such as statistical analysis, you can export them via Export > Features. There are (2) options here:

* Export All Features: This will export the entire initial (non-fisher scored) feature set(s) from the feature selection method performed, in the format below. The image below is from an Individual Combos All Features.



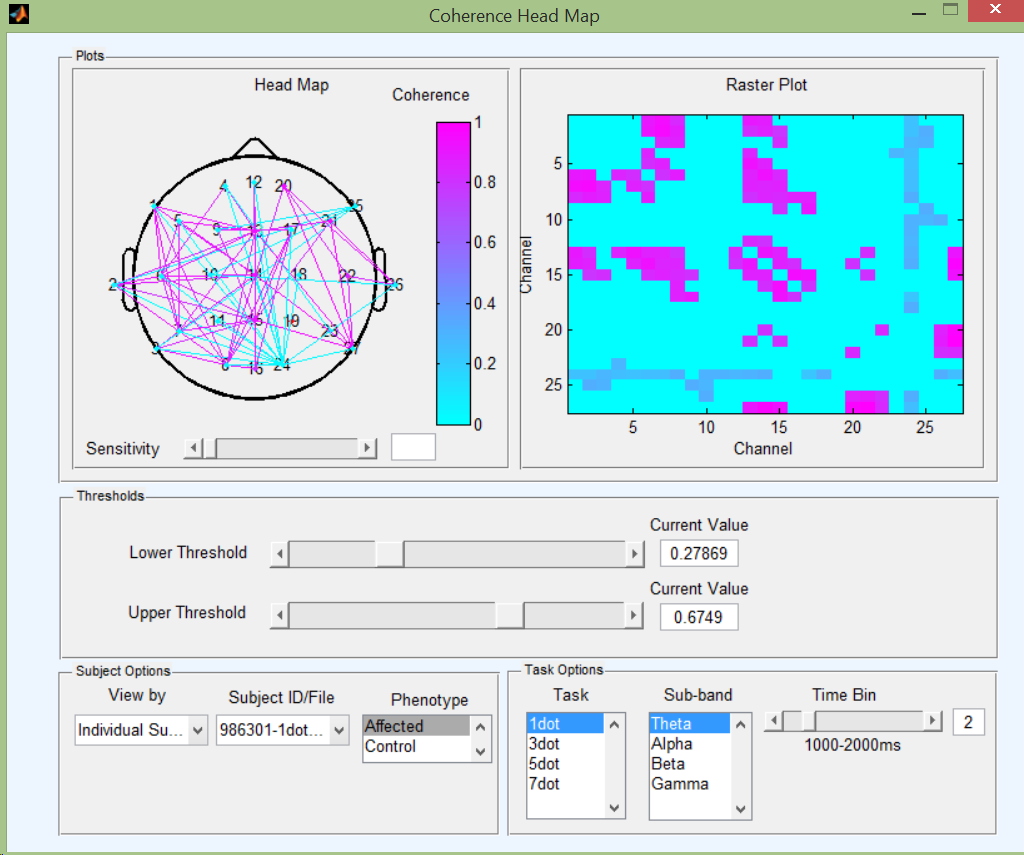
* Export Top Features: Export the top features from the previously chosen feature selection method.

**7 Connectivity Maps**

Previous required steps:

* Running Coherence Analysis
* Loading Connectivity Matrices

If you are interested in visualizing the connectivity via coherence, you can do that in the **View > Connectivity Maps** menu. A window similar to the following will appear:



Here you look at connectivity through different methods, such as raster plots, 2-D head maps, and schemaballs.