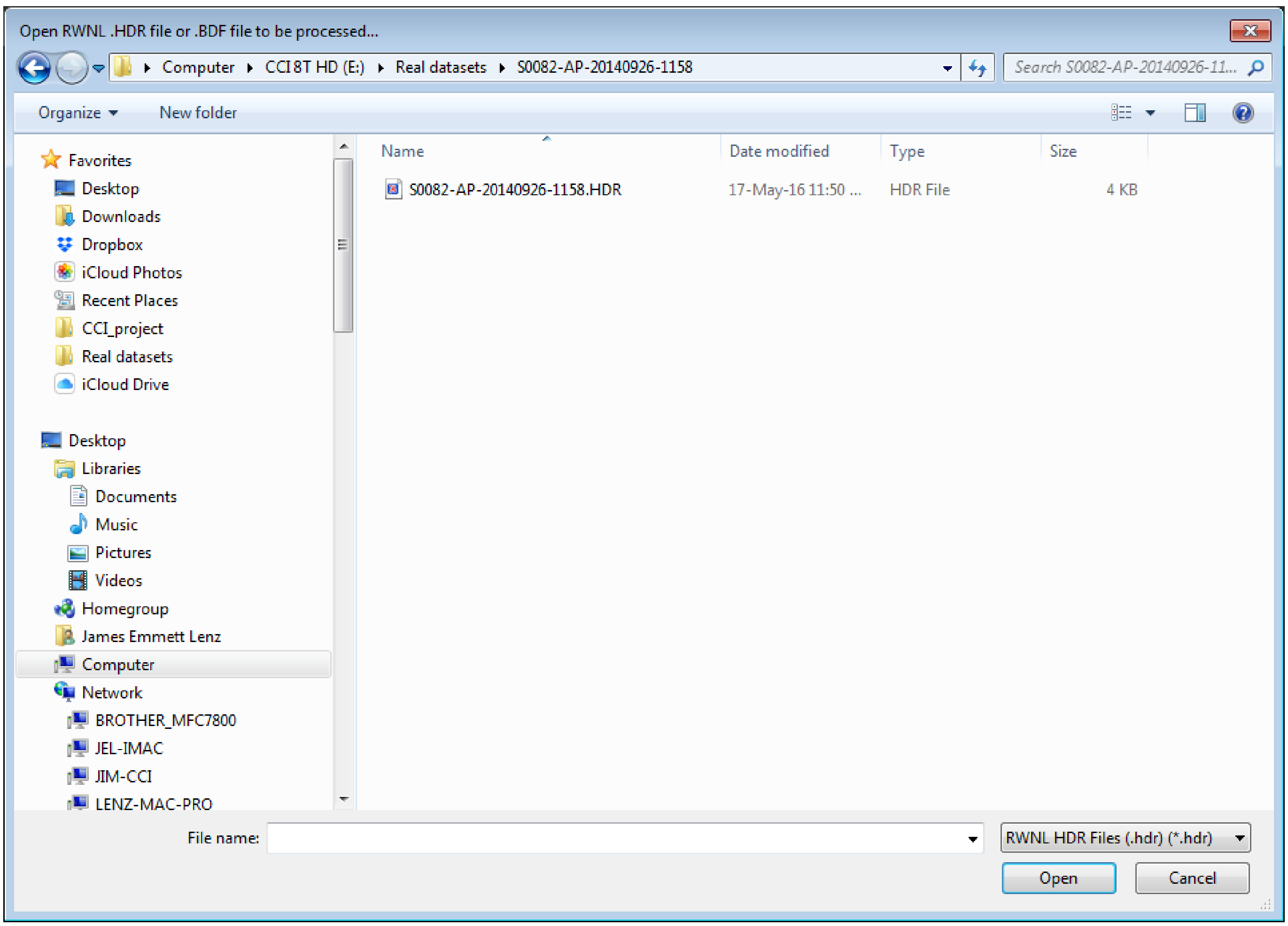
PreprocessDataset Application User Guide

Introduction

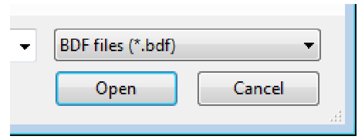
The PreprocessDataset (PD) application is intended to simplify and consolidate the early processing of physiological datasets acquired through the BioSemi system. The processing functions include simple channel selection and renaming, simple data decimation, global detrending of channel data, various referencing options, and IIR filtering of channel data. PD also can compute the surface Laplacian (SL) on interpolated (and smoothed) EEG data when electrode locations are known. PD accepts both RWNL complete datasets and “naked” BDF data. In the latter case, the SL calculation is unavailable since no electrode locations are available. [Future expansion information will be consistently shown in square brackets ([]) in the remainder of this document. PD will be expanded to accept exported EEGLAB datasets in the .SET (.MAT) format.]

Opening a dataset

To open an RWNL dataset, open its .HDR file using the file selection dialog.

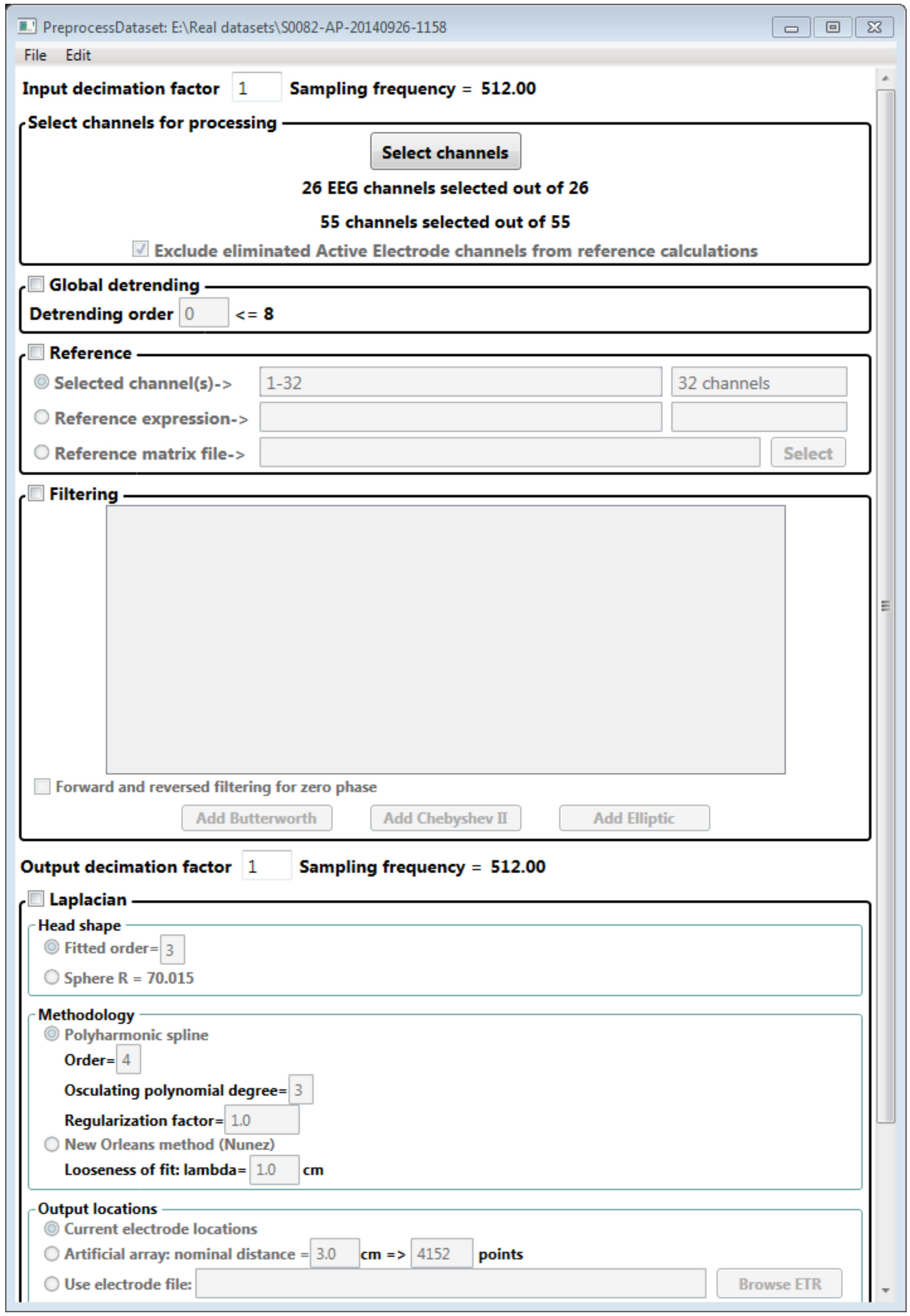


To process a "naked" BDF file change the file type filter via the dropdown, selecting "BDF files".



Some assumptions are made about the form of the underlying BDF files. All EEG electrode channels should have a "Transducer type" of "Active Electrode" in the BDF header (called AE channels hereafter). If location data is available in the form of an .ETR (electrode) file in the RWNL dataset (i.e., not a "naked" BDF file), the electrode names in the .ETR file and the BDF header record must match exactly, including capitalization. Those AE channels having a matching location are termed "EEG channels". Note that some data acquisitions have non-EEG channels that are AE. These should be carefully considered, particularly when performing referencing functions (see below). In "naked" BDF datasets, all AE channels are assumed to be EEG channels (although without location; so SL can't be performed) and thus non-EEG channels should be manually eliminated in the channel selection function to avoid problems.

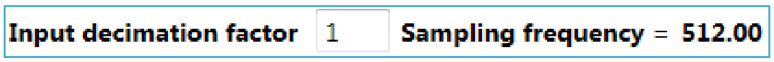
The primary dialog box is then displayed (as shown at right). One sets the various parameters to describe the preprocessing functions to be performed, each of which will be described in greater detail in the following paragraphs.



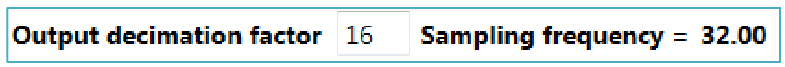
Parameter files

PD uses the concept of “Parameter files” to make repeated use of the same preprocessing protocol easier. One can set up the conditions of a preprocessing protocol in the dialog box and save it using the menu “File -> Save parameter file…” in the upper left hand corner of the dialog. A parameter file may then be reloaded when any other (similar) file is loaded using “File -> Open parameter file…” menu item. This will change all the dialog settings to the condition in which they were saved. This does not apply to Channel selection, since channel rejection is usually on a file-by-file basis.

Decimation

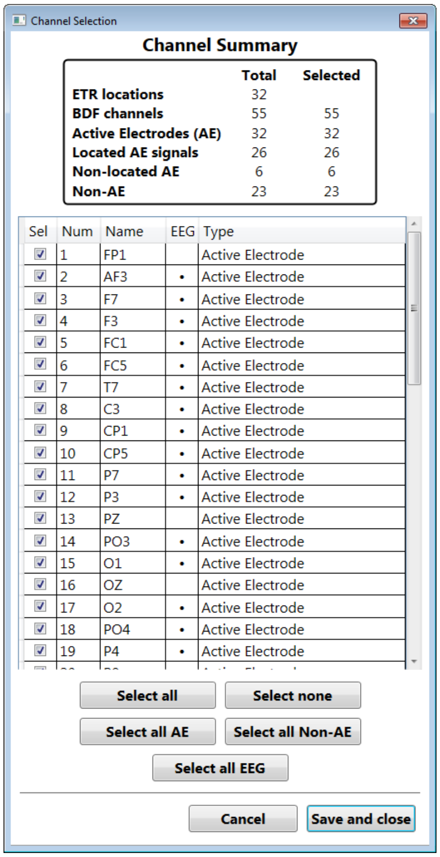


PD performs N-decimation by including every N^{th} data point. No filtering or moving average is used. For this reason there are two types of decimation: decimation before any processing occurs (“Input decimation”) and decimation that is performed after IIR filtering, but before SL (“Output decimation”). If N_1-decimation is used in the first case and N_2-decimation in the second, the final result will be $N_1 x N_2$-decimation in the final output. It is the user’s responsibility to assure that aliasing around the Nyquist frequency is controlled by appropriate filtering. Note that, in the first decimation, aliasing is controlled by the prefiltering of the signal by the BioSemi data acquisition system, and, in the second, by the IIR filtering performed in PD. Only integer decimation is allowed. The final output BDF file will have a record length with the same number of data points, but with an expanded duration of $N_1 x N_2$x(original record duration). This is usually $N_1 x N_2$ seconds. After entering the decimation factors, note that the resultant sampling frequencies are displayed.



Channel selection

Channel selection (and renaming) is performed in a secondary dialog by clicking the “Select channels” button in the main dialog.



Here is an example of the selection dialog for an RWNL dataset. Note the summary of the number of channels available and the number selected. This is updated whenever changes are made. BDF channels are classified as AE, EEG, non-located AE, and non-AE. The number of available electrode locations is shown form reference.

Channels are selected for inclusion in further processing by using the checkbox in the first column. The BDF channel number is in the second column. The third column is initially the name of the channel in the underlying BDF file, but may be edited to change the name in the output record. Name change can also be used to match to a name in the .ETR file.

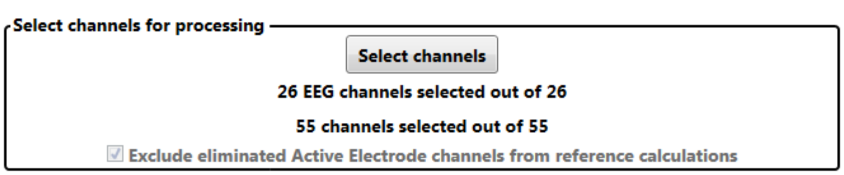
The fourth column indicates if the channel is currently considered an EEG (located AE) channel and thus can be used in any SL calculation.

The fifth column lists the channel transducer type as listed in the BDF header record.

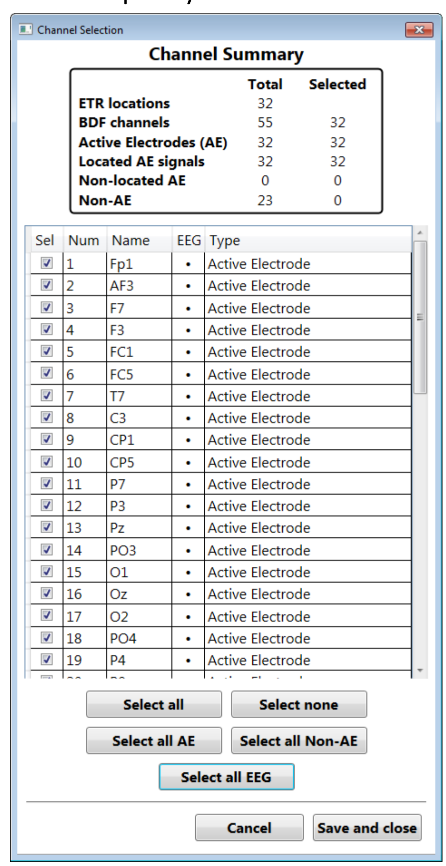
The buttons under the channel list can be used to quickly select various subsets of channels.

Close the dialog by clicking “Save and close” to save any changes or by clicking “Cancel” to ignore any changes.

If one opens a “naked” BDF file, the channel selection is similar, but there are no “located AE” channels, so the EEG column is not displayed.



In the main dialog, there is a checkbox labeled “Exclude eliminated Active Electrode channels from reference calculations” (see above). This is checked by default. If referencing is performed, AE channels not selected are automatically eliminated from the referencing calculation. This is what is usually required, since they are often omitted because of noise or artifact.



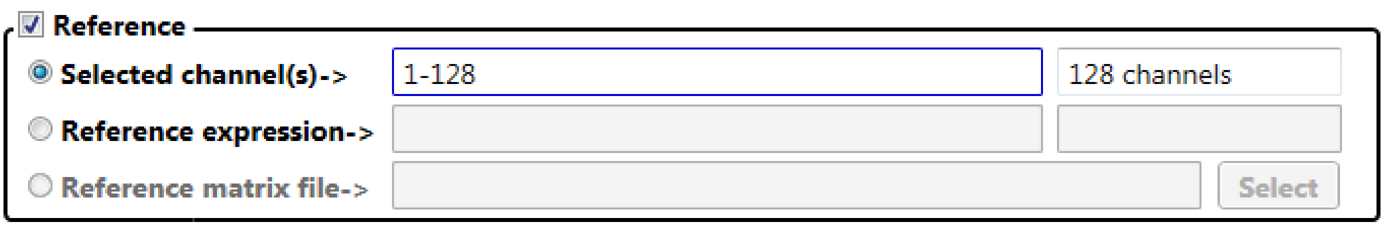
Here is an example of changes that might be made in the previous channel selection dialog (see right). Selection of all AE channels has been performed and all the “misnamed” BDF channels have been edited to match the names in the .ETR file (note again that capitalization counts!). We thus now have 32 selected channels, all of which are EEG (located AE) channels.

Global detrending

This function eliminates long-term trends in each channel over the entire dataset. The trends are

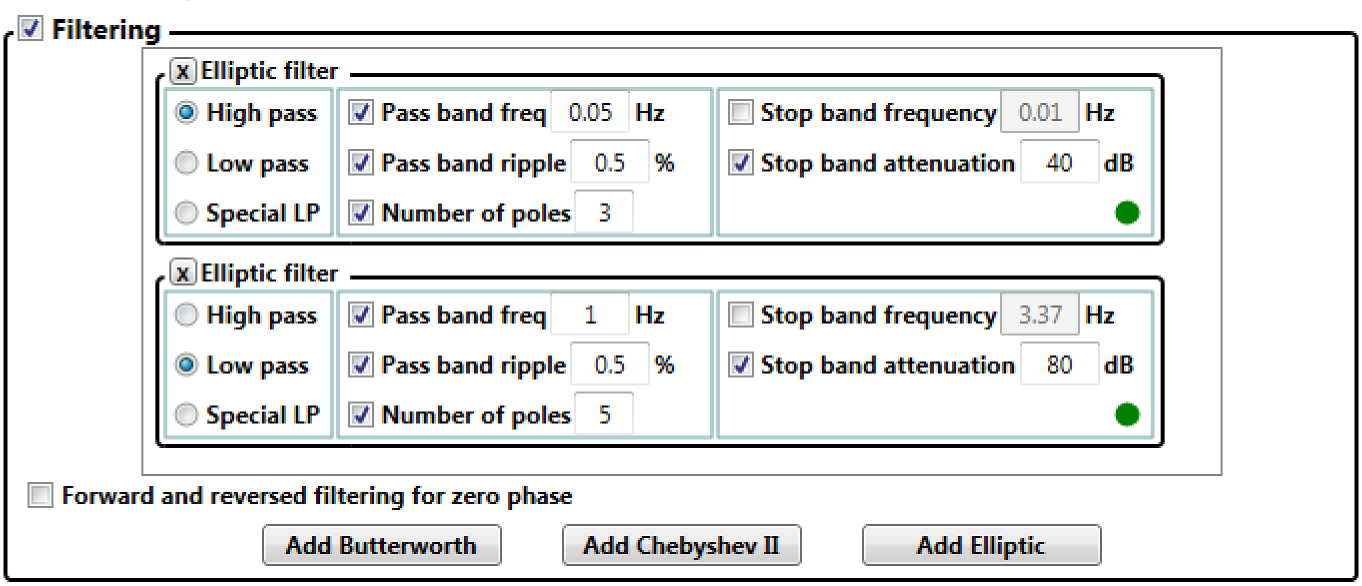
polynomials of the selected order. Thus, 0 implies removal of offsets only, 1 removes linear trends (and offsets), 2 is quadratic, etc., based on least-square curve fit of the entire dataset (after the first decimation).

Referencing



General referencing of the selected AE channels is performed using this section of the main dialog box. If all the selected AE channels are to have the same reference, use the first selection. Common average reference would use (an average of) all the channels for the reference signal. The second option permits use of multiple reference signals. See other documents for further explanation. The third option is currently not implemented. [This option will permit the use of a general matrix to form the reference signal for each channel individually based on a weighting of the signals.] Note that if the “Exclude eliminated …” option is selected in the channel selection box, only selected channels will be used in the calculation of the reference.

IIR filtering



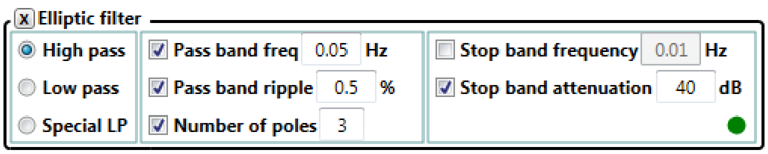
DP allows any of three classes of infinite impulse response (IIR) filters. These are Butterworth, Chebyshev II, and elliptic. Any number of high or low pass filters may be included and these are performed in series on the data channels. [Inverse filters are not currently available, so band stop filtering cannot be performed.] Thus, by combining two filters one may achieve a band-pass filter. Filters are added by clicking on the buttons at the bottom and individual filters can be removed by clicking the "X" button in the upper left corner of each filter panel.

The "Forward and …" checkbox is selected when one wishes to perform both forward and reverse filtering of the data to achieve zero phase shift, but this results in an acausal filter.

The individual filters added each have a "design panel" in the list of filters. Once a correct set of parameters have been entered for a given filter, a green dot will appear in the panel. Each filter type also has a radio button to select high or low pass design.

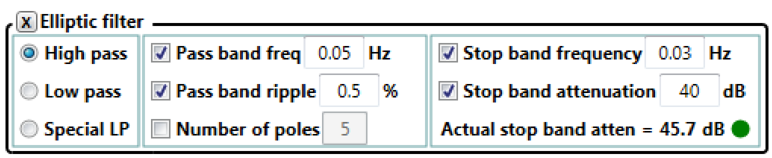
Elliptic filter

Here is the design panel for the elliptic filter. There are a set of 5 different parameters, but the filter design requires that exactly 4 be set and the fifth calculated from those values. The 4 parameters are chosen by the checkboxes.

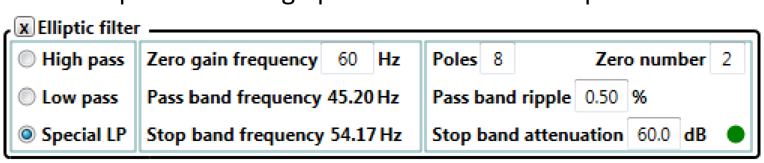


In this case, we've selected "Pass band frequency", "Pass band ripple", "Number of poles", and "Stop band attenuation" as the set parameters and with "Stop band frequency" being calculated at 0.01Hz. One has chosen 0.05Hz for the pass band with a ripple (difference between highest and lowest amplitudes) of 0.5%. The stop band attenuation is a minimum of 40dB and the filter is designed with 3 poles.

Alternatively, one might wish to control the width of the transition band (between stop and pass bands). By selecting the "Stop band frequency" as a parameter, setting it to 0.03Hz and letting "Number of poles" be selected by the application, one sees that 5 poles achieves the design and results in an actual stop band attenuation of 45.7dB (calculated by the application), greater than the design (minimal) attenuation of 40dB.

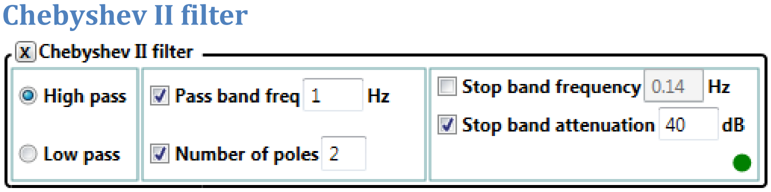


A third elliptic filter design pattern is available: "Special LP". This is a specially designed low pass filter that permits placement of one of the stop band zeros on a particular frequency, completely eliminating that frequency. Here we have placed our nulled frequency at exactly 60Hz and selected the second zero in the stop band for this (slightly more forgiving for inexact frequency). We find that if we use a pass band ripple of 0.5% and a stop band minimum attenuation of 60dB, that our pass band frequency is 45.2Hz. Note that in this case the only choice is to select the null frequency, the number of poles, ripple and stop band attenuation, and the pass band and stop band frequencies are chosen for you.



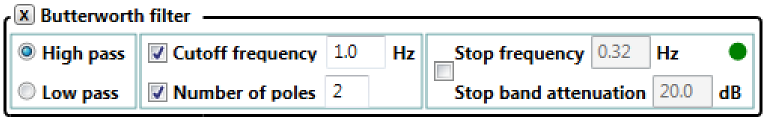
Chebyshev II filter

Three of four parameters are need to be specified for the Chebyshev filter. The Chebyshev



II filter is similar to the elliptic filter, but with the pass band ripple forced to be equal to zero.

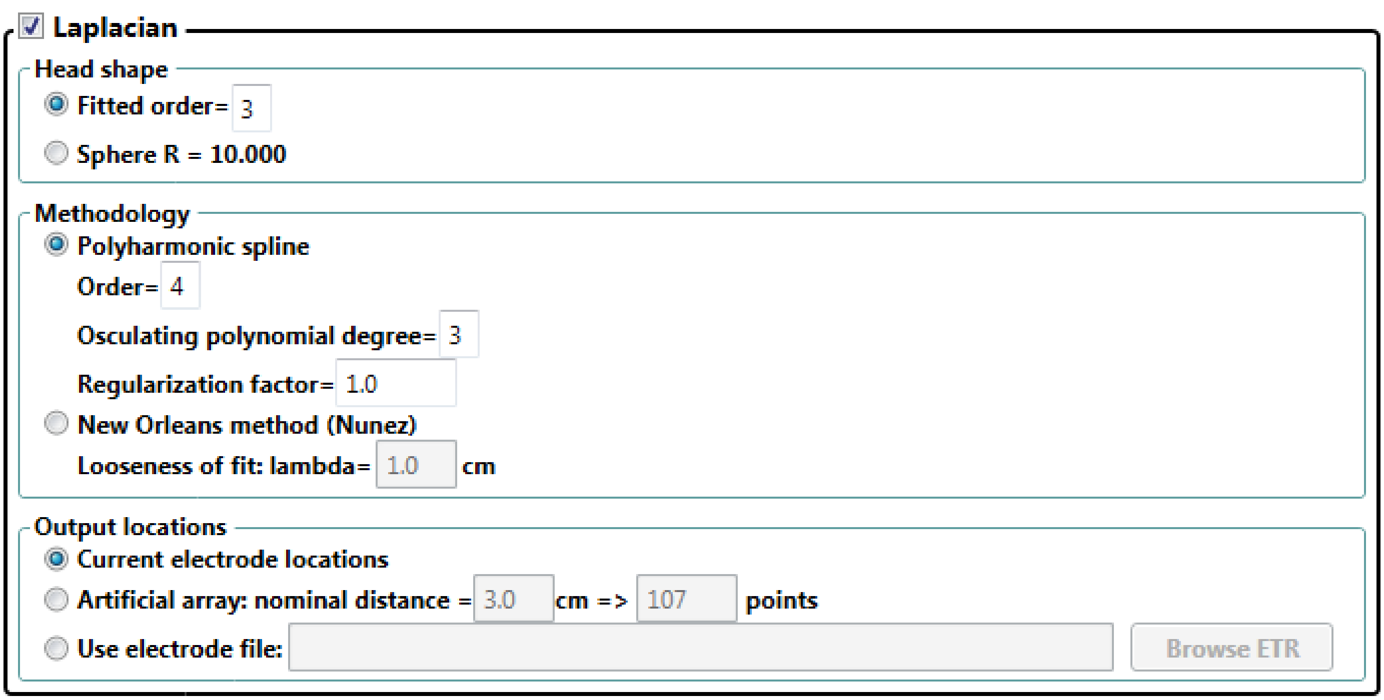
Butterworth filter



The Butterworth filter design requires selection of two parameters from a choice of 3. The two usually selected are “Cutoff frequency” and “Number of poles”. An arbitrary point of 20dB down is then used to calculate the “Stop frequency”. In a Butterworth filter there is no pass band ripple and attenuation in the stop band increases the farther one goes from the cutoff frequency.

Surface Laplacian

DP implements calculation of a surface Laplacian (SL) of the EEG portion of a dataset. SL is calculated from a smoothed interpolation of the instantaneous EEG values, generating an estimate of the scalar electric field near the surface of the scalp. The SL is then based on this estimate using either of two head shapes: spherical or a smoothed estimate of the head from the electrode locations. Thus there are three classes of choices that need to be made: 1. Head shape; 2. Interpolation/extraction/smoothing methodology; and 3. Output locations.



For head shape there are two choices. On can assume a spherical head where the radial coordinate of the electrode location is replaced with a constant consisting of the radius of a sphere fitted to the electrode actual locations. This is equivalent to projecting each electrode radially onto the spherical surface. Otherwise one can fit a “smoothed” head shape by fitting the electrode locations to spherical harmonics, up to the stated order. The higher the order, the more exact the fit, but resulting in less smoothing of the surface. Zeroth order is equivalent to a spherical fit.

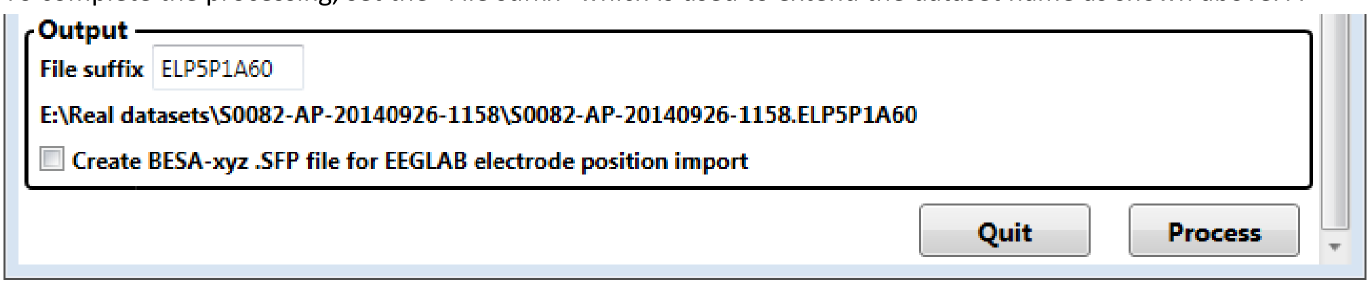
There two major classes of methodologies, both concerned with making a smoothed estimate of the scalar electrical field in the vicinity of the surface fitted to the head, based on the instantaneous EEG signal at the electrodes. The first is a spline technique based on Greene's functions of a given order. This is equivalent to the so-called "thin plate" spline in two dimensions, though this is now a three dimensional field. Increasing the order results in a more "rigid" fit (and quickly becomes unstable), while lower orders provide more "flexibility" in the fit. The regularization factor controls how close the fit is required to be to the actual measured field values, higher values giving a "smoother" fit, but with a higher mean square error to the actual EEG values. The degree of the osculating polynomial is usually one less than the order of fit, but can be made less.

The second methodology is fitting the voltage field to an ad hoc set of functions, similar to Greene's, but not guaranteeing the energy minimization that the latter provide. This method has been studied by Paul Nunez and is sometimes termed the "New Orleans" method. There is a single parameter which controls the smoothness vs. exactness of fit to the data points.

There are three choices for selecting output locations. The surface Laplacian of the interpolated field is calculated at each site and output as a new channel. Although these can be the original electrode sites, often it is better to have another set of locations, perhaps more densely located for data presentations. Thus an "artificial array" of sites may be chosen by selecting a "nominal distance" between sites. The number of sites is calculated from the distance selected. Finally, one can chose a particular .ETR file to locate the output sites, which might be useful when comparing various preprocessing techniques.

Processing and output

To complete the processing, set the "File suffix" which is used to extend the dataset name as shown above. A



warning is presented if this results in a clash with any other dataset file name, which will be overwritten if one proceeds.

If one checks the "Create BESA-xyz …" box, a file with .SFP extension is created containing the EEG electrode positions (or output sites if SL performed) in BESA Cartesian coordinates which can be directly imported into EEGLAB. In particular, BESA coordinates are the same as RWNL coordinates: the X-axis is toward the right ear, the Y-axis to the nasion and the Z-axis is toward the vertex. EEGLAB automatically rotates these coordinates into its system. This option is only available when processing RWNL datasets.

Clicking the "Process" button in the lower right corner begins the computation. Note that this button will only be enabled if all the various parameters are correctly and consistently set. A progress dialog is displayed while processing is underway. If a RWNL dataset is being processed, a new, valid, RWNL dataset is created; if a "naked" BDF file is processed, only a BDF file is created.

In addition, a log file is created (extension .log.xml) in XML format which describes the processing that was performed, including some of the results of that processing (e.g. detrend coefficients).