Making oscillation detection more robust

Kieran A. Pawluk[1], Tamari Shalamberidze[2], Jeremy B. Caplan[1,2]
[1]Department of Psychology, University of Alberta, Edmonton, AB, T6G 2R3, Canada
[2]Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, T6G 2E1, Canada
*Correspondence: jcaplan@ualberta.ca.

Abstract

Background:
Neural oscillations are important for understanding cognitive functions. To quantify them, certain methods, including *Better OSCillation detection* (BOSC), distinguish oscillatory activity from non-oscillatory 1/f background activity and derive detection thresholds in order to disregard most background signal. When successful, this produces detection criteria that are fairly calibrated across frequencies. However, if the background estimate is misaligned, this can backfire and potentially introduce a frequency bias.

New method:
The optimized BOSC method incorporates several improvements after testing each independently and as combinations before comparing them all together with the standard BOSC method. The improvements in question are: removing high-power values across frequencies, using median rather than mean power values, and robust regression.

Results:
The new BOSC method showed enhanced performance when using shorter time windows and when substantial power existed at one end of the measured spectrum. Synthetic signals were used to demonstrate further versatility and the limitations of the new method.

Comparison with existing methods:
The standard BOSC method fared reasonably well aside from some extreme edge cases. Outcomes suggested that at very short time windows, or when artifacts or lopsided power spectra are a concern, the optimized BOSC method could result in a more selective fit that shows greater alignment with the coloured-noise background signal.

Conclusion:
The standard BOSC method performs well in many typical scenarios, but the optimized version is ideal for less conventional scenarios and addresses many of the shortcomings of the standard method in these cases.

Kieran A. Pawluk https://orcid.org/0009-0008-2076-2439
Tamari Shalamberidze https://orcid.org/0000-0003-3997-9837
Jeremy B. Caplan https://orcid.org/0000-0002-8542-9900

Keywords: electroencephalography, 1/f noise, oscillation, spectral analysis, alpha

1. Introduction

Neural oscillations, rhythmic brain activity produced by synchronized neuronal firing, are associated with numerous brain states and cognitive functions so quantifying them is very important for understanding the neural basis of cognition and behaviour (e.g., Buzsáki and Draguhn, 2004). The quality of neural oscillation research hinges on how accurately rhythmic activity can be quantified. This is not trivial because oscillations are typically not sustained but are sporadic in duration and not constant but variable in both amplitude and frequency. Also, the fact that oscillations are embedded within signal that has non-zero background power values at all measured frequencies makes this even more challenging. Fortunately, the form of this background power is known to be coloured noise, of the form $P(f) \propto 1/f^{\alpha}$, where P is power, f is frequency and and α is an exponent that is typically between 1 and 2 (Shlesinger and West, 1998). Numerous researchers have exploited this to either study the background signal itself (like Samaha and Cohen (2022) and Donoghue et al. (2020)) or to calibrate measures of oscillatory activity by defining an oscillation as a signal that deviates significantly from the standard background signal (Caplan et al., 2001; van Vugt et al., 2007; Whitten et al., 2011; Hughes et al., 2012; Kosciessa et al., 2020; Seymour et al., 2022). One widely adopted such method is Better OSCillation detection (BOSC; Caplan et al. (2001); Whitten et al. (2011)), which uses power thresholds calibrated relative to an estimate of the background along with a duration threshold that scales with the oscillatory period (Figure 1). Although adopted by more than 32 labs in over 52 publications, the BOSC method has known vulnerabilities shared by any method that is based on fitting the coloured-noise spectrum. A particular example is that whenever the background estimate is misaligned the thresholds will be calibrated incorrectly and this is especially problematic with errors in the slope. This is because inaccuracies in the slope estimate reintroduce a bias across frequency even though the goal of estimating the background spectrum is to remove this frequency-bias. Here we estimate and illustrate the severity of the problem in general and in exaggerated problem conditions. We then compare several approaches to help the method fit more of the background signal and less oscillatory or artifactual signal. Before we do so, we briefly elaborate on the challenges in quantifying oscillations and potential applications of a more stable method.

1.1. Quantifying oscillations and challenges

The alpha rhythm (8–12 Hz) is one of the most robust known rhythms in the EEG, easily evoked when participants close their eyes. For these reasons, it is a very good test case for signal-processing methods. When this was first discovered by Berger (1929) the oscillations were identified by visual inspection but later they were identified with Fourier Transforms to compute the power spectrum while looking for a peak at a particular frequency (Klimesch, 1999). However, signals in the brain are non-stationary and while windowed Fourier Transforms can address changes over time, wavelet transforms are better suited to

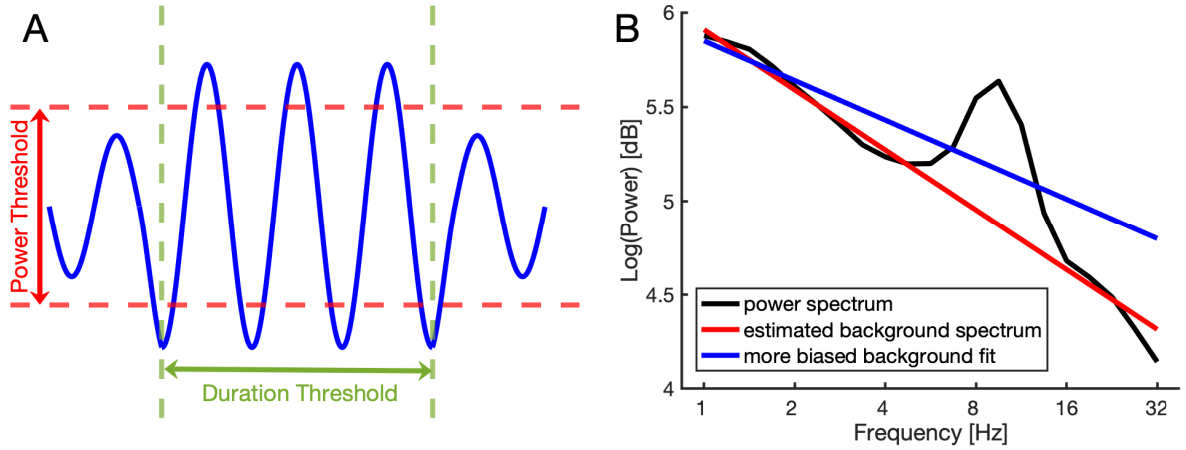

Figure 1: A schematic representation of the thresholds used in the BOSC method along with the background fitting of power spectra. (A) Illustration of power threshold and duration threshold on a section of EEG signal. Both thresholds must be exceeded for an oscillation to be recorded. (B) A power spectrum with a background fit which is used to derive the power thresholds at each analyzed frequency. Also shown is an incorrect fit which shows how a bias across frequency will occur if the slope of the background fit is wrong.

this non-stationarity. This is because the wavelet family is scale-invariant, meaning that wavelets are comparable (i.e., the precision, or resolution, in both time and frequency are a constant proportion at all frequencies) across frequencies (Schiff et al., 1994).

Since background signal power has the form $P(f) \propto 1/f^{\alpha}$, oscillatory power needs to be adjusted accordingly at each frequency. Imagine if one were to select only one single power threshold to label signal as oscillatory— power at lower frequencies would nearly all be above that threshold and power at higher frequencies would nearly all be below that threshold.[1] Several methods of analyzing oscillations, including BOSC, start by estimating the background spectrum such that measures can be rescaled relative to this estimate. Applications of BOSC typically go one step further, and label segments of the recording as oscillatory when the signal deviates substantially from that theoretical background signal. Therefore, for oscillation detection the power thresholds must differ systematically by frequency to account for the coloured-noise background form, and the BOSC method does this by estimating the background spectrum using linear regression of the power spectrum in log-log coordinates where it should be close to a linear function (Whitten et al., 2011). Because power is computed as the square of a complex value, the sum of the square of the real and imaginary parts (each of which is approximately Gaussian-distributed) should produce a $\chi^2(2)$ distribution at each frequency like the example in Figure 2A (Percival and Walden, 1993). Power thresholds are thus set to some high percentile (usually the 95^{th} percentile) of the $\chi^2(2)$ distribution after estimating mean power at each frequency using the background

[1]To offer some intuition: For a given amplitude, higher-frequency components reflect more energy. If power did *not* decrease as frequency increased, that would sum to infinite energy, which is not plausible. This means the observed decrease in power with increasing frequency (as in EEG signals) may be seen as reflecting a physical constraint.

estimate. An oscillation is "detected" at frequency f when the power threshold is exceeded for a minimum consecutive time, defined by the duration threshold, set to some number of cycles at f (usually three). The step-by-step process of BOSC from loading signal data to detecting oscillations is illustrated in Figure 6 (red pathway). After detections are identified throughout the entire signal, we can then calculate $P_{episode}$ values at each analyzed frequency; these represent the proportion of time (of the total signal length) during which detections are present.

If the estimate of the background spectrum is well-aligned with the true background, using a fixed-percentile power threshold will be unbiased across frequencies. However, if the background estimate is misaligned, especially if the slope is wrong, that wrong slope will introduce a new bias in oscillation analyses: low frequencies will be detected with a more liberal power threshold than high frequencies, or vice versa depending on whether the slope is under-estimated or over-estimated.

In summary, wavelet transforms are best-suited for non-stationary electrophysiological signals, so BOSC makes use of them before estimating the $1/f^{\alpha}$ background during spectral analysis to derive detection thresholds that should be exceeded by oscillatory activity. However, a frequency bias can be introduced if the slope is wrong so robustness in the estimate is desired. We devised several scenarios where a more robust background fit would be advantageous, as we summarize next.

1.2. Circumstances that could benefit from a more robust background-spectrum fit

Because previous applications of BOSC and similar methods have used several minutes or more of signal to estimate the background spectrum, a scenario of interest is using short time windows within a signal. Standard use of a long segment of data to estimate the background has the obvious advantage of using more data, thus increasing the signal-to-noise ratio of the power spectrum and possibly avoiding many cases where the slope could be misaligned. However, this breaks down if the spectrum is very non-stationary and changes throughout the recording. For example, overall gain shifts due to electrodes slipping or drying out over time or gradual mental state changes can make the background spectrum appear quite different over time.

Another reason for investigating BOSC using short data windows from a given signal was because using a smaller window could adapt to slow drifts in the spectrum. This could also possibly help calibrate the thresholds better for oscillation detection within that particular window since the rest of the signal would not influence the background fit for that particular region. Using a shorter window could also reduce computation time since wavelet transforms and Fourier transforms scale superlinearly in the number of time points. Increased computational speed could be particularly helpful for real-time applications of EEG analysis. The tradeoff, however, is that with a shorter background-estimate window, the estimate becomes more unstable, which could insert more noise into oscillation analyses than using a longer window. A more robust background-estimate method is needed to achieve good-quality performance at short time windows.

The second circumstance of interest – signals with oscillations at or near the edges of the measured power spectrum, was investigated because a peak at the edge of the power

spectrum could have a large influence on the slope, tilting the background fit. This has a worse effect than a peak near the middle of the range of analyzed frequencies which would pull the fit upwards but with a far smaller effect on the slope. The former would produce the bias across frequencies regarding the power thresholds discussed above so correcting the background fit would be useful when applying the method to signals with oscillations near the edges of the power spectrum. Additionally, other artifacts that introduce power values that deviate from the background spectrum could affect the fit.

In sum, the two scenarios of interest that could benefit from a more robust background fitting method are: the use of short time windows within a larger signal, and signals with oscillations near the edges of a sampled frequency range.

1.3. Previous improvements to robustness of the background fit

The BOSC method has been shown to be relatively robust to state changes regarding the background fit and calculation of thresholds (Whitten et al., 2011; Hughes et al., 2012). However, anything that influences the regression will affect the thresholds and this was seen with peaks in power spectra along with various artifacts. An improvement to the background fit that was used in the standard version of BOSC and continued to be used in our optimized version involved converting power values to log form before averaging rather than averaging power values and then log-tranforming (van Vugt et al., 2007). Later improvements focused on the fitting algorithm involve eBOSC (extended BOSC) by Kosciessa et al. (2020) which used robust regression and fBOSC (FOOOF + BOSC) by Seymour et al. (2022) which combined BOSC with the FOOOF (fitting oscillations and one over f) method by Donoghue et al. (2020) which allowed the option of estimating a non-linear background spectrum. While useful, these modifications still assumed that the background estimate is aligned with the actual background after minimizing the effects of spectral peaks. In this paper, the improvements made to BOSC helped to address the potential less-obvious influence of various other artifacts by evaluating the background fit using theoretical properties, notably the $\chi^2(2)$ distribution of power values rather than assuming that the best fit to the points on the power spectrum was the best background estimate. Getting the background fit right, especially the slope, has always been particularly important since, as already mentioned, an error in the slope can introduce a bias across frequencies.

In sum, the background fit has been previously improved by: using mean values after log-transforming, along with using alternative regression forms to ignore outliers or perform a non-linear fit, but an evaluation using the theoretical properties of the background, as we do here, could help address less-obvious effects of artifacts.

1.4. Outline

In this paper, we explore methods of improving the general robustness of the BOSC method using empirical data with varying degrees of alpha oscillations along with synthetic data with oscillations simulated at various frequencies and amplitudes.

Our first goal was to improve the robustness of the background fit at shorter data windows after first testing standard BOSC. To evaluate the robustness of the regression-based

estimate of the background spectrum, we considered several outcome measures. We reasoned that since the true background spectrum should be relatively stable over timescales of a few minutes, the variability of a given method when estimating the background across successive background windows could be used as an outcome measure for robustness. The first way we quantified this variability was by computing the standard deviation of the slope estimated by the regression, over successive time windows. The second outcome measure was the standard deviation of estimated intercepts. We were more concerned with the standard deviation of the slope because a wrong slope introduced a more serious problem: a bias across frequencies that can artifactually produce ceiling or floor oscillation detection at high or at low frequencies (Figure 1B). A wrong intercept changes the degree to which thresholds are conservative or liberal, but comparably at all frequencies. Figure 3A,B plots these measures as functions of window size, and both began to increase sharply at around the 30 s window and below. While this showed greater consistency, it did not necessarily mean the background estimates were more aligned with the true background. Since we know that a signal of pure non-oscillatory background has a theoretical power value distribution of $\chi^2(2)$ at each frequency, we can compare that with the distribution of the actual power values based on the background fit using the Kolmogorov-Smirnov d statistic (KS d) which quantifies the maximum difference between two cumulative distribution functions (CDFs) (Figure 2B). Therefore, lower KS d values indicate higher alignment with the theoretical $\chi^2(2)$ distribution so the KS d values for each window size are an outcome measure that evaluates the alignment of the background fits with this theoretical standard. For our specific outcome measure, we began by taking the median KS d value of the analyzed frequencies in all windows within a signal before averaging those values from all windows of each size (e.g., average the median KS d from all 5 s windows, repeat for all other window sizes). Compared to slope and intercept SD, this averaged median KS d (simply referred to as median KS d) remained much more steady across window size and although there was an increase with shorter windows there was also a lot of overlap of boxes across window sizes (Figure 3C).

Our second goal was to test and improve the robustness of BOSC for signals with oscillations near the edges of the power spectrum along with other extreme cases. We first investigated this using empirical data with a "partial-spectrum" test. This test used signals with an alpha peak that we cut off artificially at the peak frequency so the peak would be right on the edge of the spectrum. The slope and y-intercept of the background fit were compared with that of the full spectrum. In this test, robustness was indicated by small differences in slope and y-intercept between the full spectrum and the partial spectrum. For standard BOSC, a power spectrum with an alpha peak would pull the regression line up but with a partial spectrum the regression line becomes tilted as well (Figure 5). A similar test using simulated signals was done and reported later where simulations with peaks at extreme frequencies, along with various other anomalies were compared to simulations with alpha oscillations.

We had several ideas for the improvement of the background estimate:

1) Robust regression. Frequencies that deviate from the background signal we are trying to estimate may tend to be more variable so a robust regression, used in eBOSC (Kosciessa et al., 2020), could be beneficial by reducing the influence of outliers. We used MATLAB's

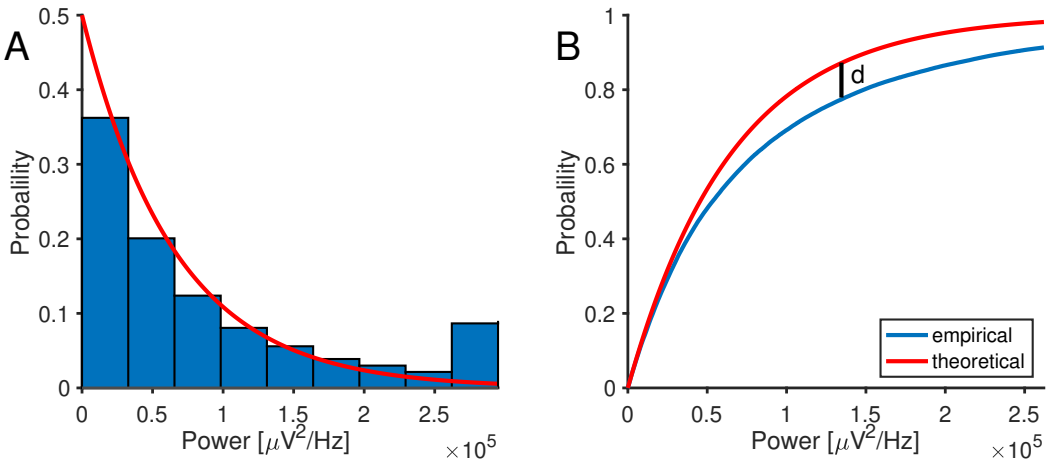

Figure 2: Empirical vs theoretical probability distribution function (PDF) and cumulative distribution function (CDFs) of power values at one frequency of interest. (A) Comparison of the empirical PDF (bar graph) from the Pz electrode of a single subject (sampling rate of 500 Hz) to the theoretical $\chi^2(2)$ PDF (red line). The last bar indicates overflow collecting power values above the range plotted. (B) Comparison of the empirical CDF using the same data to the theoretical $\chi^2(2)$ CDF. The KS d statistic indicating the maximum vertical distance between the curves is also shown.

robustfit.m function, further discussed in section 3.1.

2) Median power. Because deviations from the background signal will typically consist of very large power values, exchanging mean for median may reduce the influence of outliers. The means can then be calculated from the medians through the ratio between the two in a $\ln(\chi^2(2))$ distribution since the regression is done in log-log coordinates after accounting for the scale between the ln(Power) and $\ln(\chi^2(2))$ distributions (Eq. 2).

3) High-power removal method. Consider that power values that are not part of the background spectrum we are trying to estimate usually take on very high values (e.g., Figure 2A). If we screen out those high-power values, we may be able to more accurately estimate only the background signal. After a first pass to estimate the face-value power spectrum, we then remove any power values exceeding a high threshold (based on the $\chi^2(2)$ distribution) at any frequency in order to reduce the effects of high-power values, characteristic of oscillatory activity or possibly artifacts, on the background estimate.

4) Frequency-subset method. Consider that frequencies that have little or no oscillatory signal present come closer to our theoretical model of the background signal. We might improve the robustness of the regression by simply focusing the regression on a small set of "good" frequencies while ignoring frequencies with suspected oscillatory (or artifactual) activity.[2] After a first pass to estimate the face-value power spectrum, we performed the regression using a subset of frequencies that best represent the non-oscillatory background rather than using all frequencies. To determine these best-fit frequencies we initially investigated the behaviour of the standard deviation and coefficient of variation of power as a function of frequency (SD(f) and CV(f)). When these did not show any promise we made

[2]Thanks to Michael J. Kahana for this suggestion.

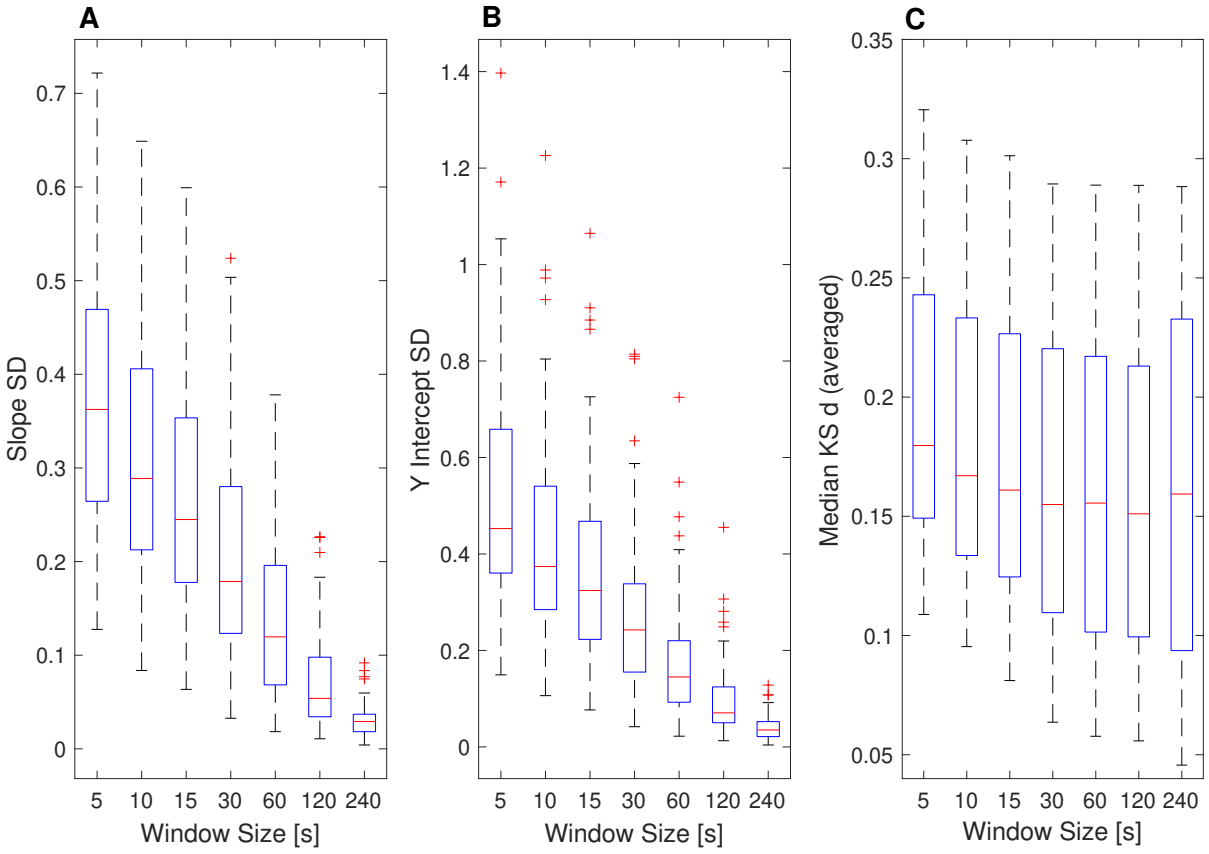

Figure 3: Measures of variability across background-spectrum estimates (time windows) as functions of window size across the 48 participants, electrode Pz. Standard deviation of the slope (A) and y-intercept (B) and median KS d averaged over windows of the same size (C). Medians are shown by orange lines inside the boxes, the tops and bottoms show the 75th and 25th percentiles, respectively (difference is interquartile range or IQR) and the ends of the whiskers show the maximum and minimum values excluding outliers. Outliers are shown with small orange crosses and are at least 1.5*IQR above the 75th percentile or 1.5*IQR below the 25th percentile. (A) At the 30 s mark and below the median slope SD sharply increases along with the spread of the values. (B) The median y-intercept SD also increases with windows under 30 s but to a lesser degree and the spread is relatively unchanged. (C) Median KS d also increases at the 30 s window but at a much slower rate and the difference is only pronounced by the 5 s window. The area of the boxes also overlap significantly so there is little change across window size for median KS d compared to the other outcome measures.

use of Kolmogorov-Smirnov tests and the respective d statistic (KS d) that shows how well the distribution of power at each frequency resemble the theoretical $\chi^2(2)$ distribution.

Combinations of these four individual modifications were also explored. They too were then compared with standard BOSC and individual modifications with the ultimate goal of optimizing the BOSC method to have increased robustness, particularly for short data windows and signals with oscillations at extreme frequencies. The potential step-by-step processes of these various combinations are shown in Figure 6 along with that of the standard BOSC method and the method we found to have optimized robustness the most.

2. General Methods

Here we describe how the empirical signals and synthetic signals were gathered before use in this paper along with the specifics of the analytical approaches used to test and improve the robustness of the BOSC method under varying conditions.

2.1. EEG data

The data we used in this paper consisted of resting state EEG recordings along with various simulated signals generated using slightly modified code from eBOSC by Kosciessa et al. (2020). The core scripts for the optimized BOSC method including generation of simulated signals can be found under Data and Code Availability.

The resting-state EEG recordings were taken from 48 participants in a larger experiment investigating the role of neural oscillations in associative memory and anxiety. The participants consisted of 69 undergraduate students enrolled in an introductory psychology course at the University of Alberta in exchange for partial course credit (mean age = 20.51 years). Written informed consent was obtained from all participants for being included in the study, in accordance with a University of Alberta ethical review board. Recordings were done after participants completed an associative memory task and questionnaires regarding trait anxiety and demographics (none of which are reported here), along with a brief break. The recording location featured an electrically shielded, sound-attenuated chamber and a high-density 256-channel Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, OR), amplified at a gain of 1000 and sampled at 500 Hz. During the recording sessions, participants were instructed to keep their eyes open after a beep until they heard the next beep, and then to close their eyes and open them only after they heard the next beep for a total of four beeps spaced one minute apart. In all, there were two 1-minute eyes-open and two 1-minute eyes-closed cycles alternating after each other with a $\sim$ 30 s buffer in the recordings before the eyes open/closed task, for $\sim$ 4.5 minutes total. Artifact caused by line noise at 60 Hz was notch-filtered and 21 subjects were removed due to excessive artifacts, leaving 48 EEG recordings for analysis in this study, all from the Pz electrode (channel 101). Spectral analysis of the Pz recordings was done using a Morlet wavelet transform with a width of 6 cycles (Grossmann and Morlet, 1985). Frequencies were analyzed for analysis in 21 logarithmic steps ranging from 1 to 32 Hz as shown by the following equation:

$$\mathsf{F} = (2^{1/4})^n, n \in \{0, 1, 2, \ldots, 20\} \tag{1}$$

We generated simulated signals using modified code from eBOSC (Kosciessa et al., 2020), made to resemble the empirical data so they had a total length of 4.5 minutes with a 20-s buffer followed by alternating nonoscillatory/oscillatory 60-s periods (500 Hz sampling rate) and a 10-s buffer at the end (see Data and Code Availability for relevant eBOSC scripts). In the background noise simulation $\alpha = 1$ therefore $1/f^{\alpha} = 1/f$ so a signal of pure noise would have a slope very close to –1 as opposed to the empirical data from the previous experiment where the true background slope was unknown due to variability in α. The seed used in the random number generator for the noise was also kept constant at a value such that the slope of pure noise would remain very close to –1. We inserted oscillations at 10 Hz to mimic alpha rhythms, 20 Hz to mimic beta rhythms and to present a peak near the edge of the power spectrum, and would also add them together to create a simulation with two peaks after simulating them separately. We also used smaller 16 Hz oscillations along with the 10 Hz and 20 Hz oscillations to create a simulation where the peaks appeared wider and more connected in the power spectrum. Additionally, we simulated high-power values in certain cases by including nine large, yet short oscillations that were randomly placed throughout the signal length at frequencies of 1–4 Hz since this is how high-power values seemed to appear most often in empirical data. The amplitude of oscillations was variable from zero to the point where a peak would be very high relative to the empirical data, except for the high-power simulations where the random 1–4 Hz oscillations were always present even if there were zero peak frequency oscillations. The position of the oscillations within the signals remained unchanged, including the seed used for the random generation of high-power oscillations.

2.2. Testing robustness

The first way we tested robustness was for the short time window scenario and we used a "sliding-window" method. This would work such that a window of a given size (either 5 s, 10 s, 15 s, 30 s, 60 s, 120 s or 240 s) would start at the earliest point (just after 6 s) and be "slid" along the length of the signal until the end of the window met the end of the signal minus 6 s. The 6 s excluded from either side is a "shoulder" value used to exclude edge artifacts, and it works out to be 3000 samples. The "sliding" was done by translating the window by one sample after performing the background estimate using BOSC and then repeating until the endpoint was reached (Figure 4). For larger windows the endpoint would come earlier since there is less space to slide the window within a given signal and this ultimately meant that more windows were calculated for smaller window sizes than for larger ones. The various modified methods were tested for robustness across all 48 datasets with the same window sizes. The standard deviation for the slope and y-intercept (slope SD and y-intercept SD) of the regression lines were compared across window sizes as well as across the different methods. Additionally, the median of the KS d values for all frequencies was calculated for each window and these values were averaged across all windows of the same size for a particular signal. This was used as another outcome measure since KS d shows how close power value distributions given the background estimate are to the theoretical $\chi^2(2)$ distribution. Since the linear regression used for the estimate is being evaluated, the

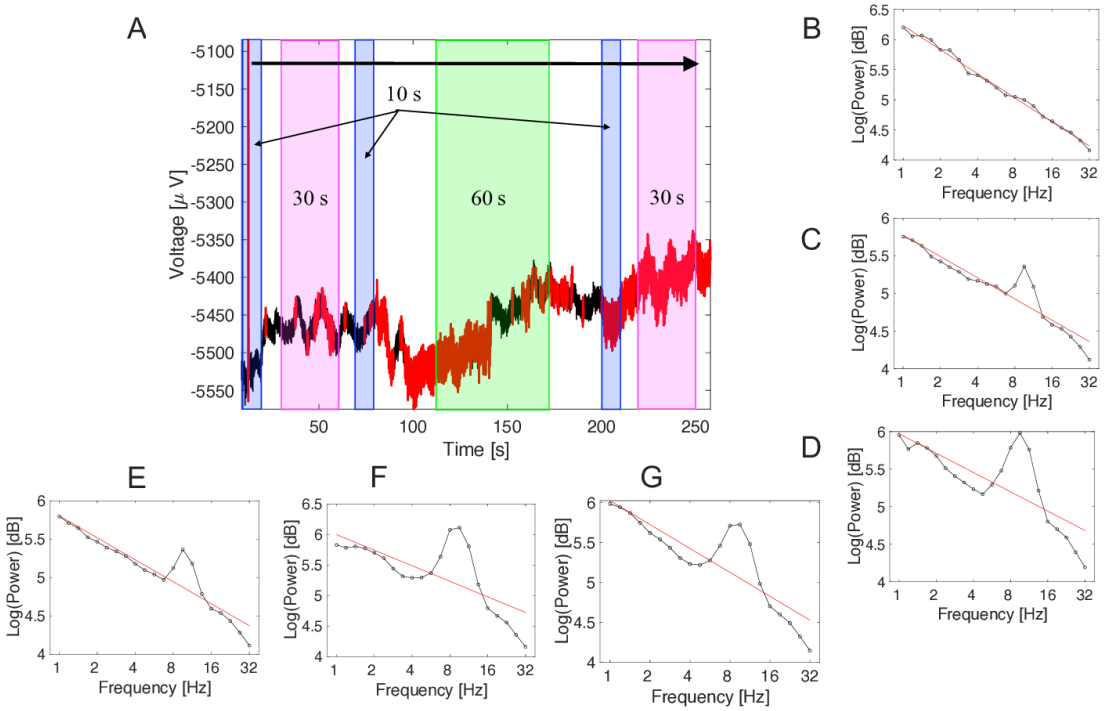

Figure 4: Illustration of the sliding-window method using a sample signal along with power spectra generated from various windows. (A) The raw signal with detected 9.5 Hz (alpha peak) oscillations highlighted in red. Also shown are example windows with three 10-s windows highlighted in blue, two 30-s windows highlighted in purple and a 60-s window highlighted in green. The black arrow at the top indicates that each window starts at the beginning and moves along the signal. (B), (C) and (D) show the power spectra generated by the three 10-s windows as they appear from left to right respectively. (E) and (F) show the power spectra for the two 30-s windows, again as they appear from left to right and (G) shows the spectrum for the 60 s window. The various spectra also show the regression lines used to estimate the background. As a window slides along a signal it generates many spectra and regression lines, the slope and intercept of which are all compared to assess variation.

power values were scaled down to the $\chi^2(2)$ distribution (μ=2) using the regression values before performing the KS test.

All methods were also tested for robustness using a "partial-spectrum" test where each method would compute the background estimate using the power spectrum from the entire frequency range or from a restricted frequency range such that if the spectrum had an alpha peak it would appear on the end rather than in the middle (Figure 5). Essentially, this would be like having a peak at the highest analyzed frequency rather than near the middle of the spectrum in the case of alpha. This in turn would affect the slope of the regression line since a peak in the middle may simply raise the line while a peak on the end can tilt it. The robustness of each method for this test was measured by the difference between the slope for the full spectrum and that of the partial spectrum with a smaller difference indicating the particular method was able to focus on the background and ignore the peaks.

Once the best-performing method from these tests was selected we then shifted focus

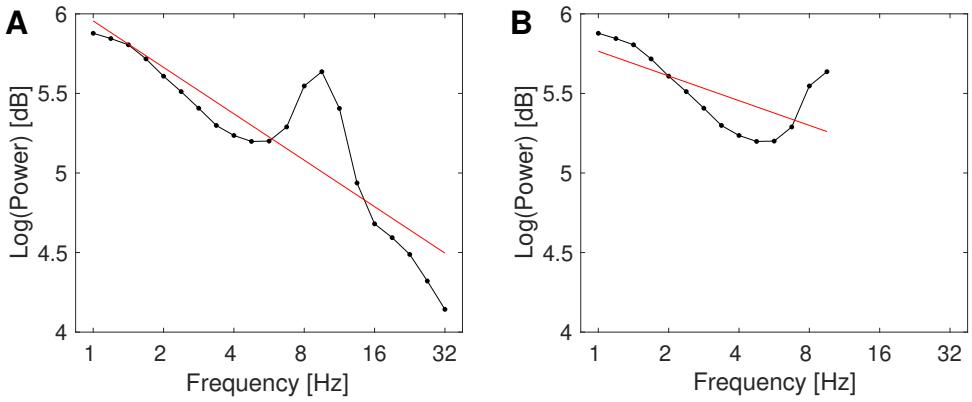

Figure 5: (A) Power spectrum of full frequency range with background estimate from standard BOSC using least-squares regression. The peak causes the regression to be skewed upwards but the slope is relatively unaffected. (B) Power spectrum with background fit using partial frequency range that ends in the middle of the alpha peak. The peak being at the end of the spectrum rather than near the middle line causes the regression to tilt rather than just being pulled upwards so the effect on the slope is very pronounced.

from the background estimate to oscillation-detection consistency. The optimal method was compared to the standard BOSC method using the partial-spectrum paradigm at first before expanding the idea using simulated signals such that the peak could be moved anywhere on the spectrum including high frequencies (which the partial spectrum emulated). Additionally, multiple peaks were simulated along with simulated high-power values. The sliding-window test was also repeated using simulated signals to test which scenarios would result in the greatest difference in performance between the optimized method and the original. For this test the amplitude of the oscillations in each signal was varied six times so there were six data points for the standard and optimized BOSC methods. Since the slope from the power spectrum of the simulated signals was preset to be very close to –1 with no oscillations added, this meant that the mean slope of the regression lines from all the windows could also be used as an outcome measure. This was done by comparing the difference in mean slope and –1, with smaller differences indicating a closer match between the slope of the background estimate and the actual signal.

Computation time was also tested in two ways for standard BOSC, the individual modifications and the optimal method using a 2019 iMac running MacOS 10.15.7 with a 3.6 GHz 8-core Intel Core i9 processor. The first test involved 100 trials of a script with the standard BOSC functions from the wavelet transform to determining $P_{episode}$ values, with the time of each trial along with the background fit by itself being recorded. A simulated signal with 10 Hz oscillations, and generated in the manner described above was used as data. The second test focused on the background fit time only for windows of the different sizes in the sliding-window test, and obtained these data windows from the same signal as the first test. The number of windows for each window size per trial was adjusted so the total amount of data to analyze remained constant, and the results were once again averaged over 100 trials.

In summary, the robustness of the BOSC method and the various proposed modified

forms were tested for the scenarios of short time windows and oscillations along the edges of a sampled frequency range using a "sliding-window" test and a "partial-spectrum" test for each, respectively. This was done using both experimental EEG recordings and simulated data.

3. Testing various robustness modifications

Next we describe each of the modifications listed earlier in detail, report the outcome measures of variability and alignment with theory for each using the tests described in the previous section, and discuss any specific contributions to improving robustness that are found.

3.1. Robust regression

MATLAB's robustfit.m function reduces the influence of outliers during a robust regression of all analyzed frequencies with an algorithm called iteratively reweighted least squares (IRLS) with the bisquare weighting function (The MathWorks Inc., 2024). This was first used in the eBOSC method by Kosciessa et al. (2020) who showed that this is useful for removing the influence of spectral peaks and providing a regression line with a tighter fit to the remaining points. For the sliding-window test there was no advantage to using robust regression compared to ordinary least-squares (Figure 7) but it did slightly increase computation time (Figure 9), likely because robustfit.m involves an iterative algorithm as opposed to polyfit.m. However, robust regression performed very well in the partial-spectrum test due to the algorithm's ability to exclude outliers (Figure 8). During the tests with simulated signals it was still able to exclude peaks at various frequencies, sometimes even multiple peaks but the simulated signals with high-power values (generated with many small peaks) were able to throw it off, along with signals with multiple wider peaks that stretch across more frequencies (Figure 11). This is presumably because once enough points are affected the algorithm no longer considers them to be outliers.

3.2. Median power instead of mean power

Since the median is less affected by outliers than the mean, using the median of the power values (in log form) instead of the mean at each frequency resulted in noticeably better performance on both tests (Figure 7 and Figure 8). Note that we used the *mean* log-power values to line up the background estimate; those means were computed from the medians:

$$\mathsf{mean}(\mathsf{log}_{10}(\mathsf{Power})) = \frac{\psi(1) + \mathsf{ln}(2) + \mathsf{median}(\mathsf{ln}(\mathsf{Power})) - \mathsf{ln}(\mathsf{median}(\chi^2(2))}{\mathsf{ln}(10)} \tag{2}$$

This equation makes use of the ratio between $\mathsf{mean}(\mathsf{ln}(\chi^2(2)))$ and $\mathsf{median}(\mathsf{ln}(\chi^2(2)))$ with the latter being equal to $\mathsf{ln}(\mathsf{median}(\chi^2(2)))$ ($\mathsf{median}(\chi^2(2))$ computed with MATLAB's chi2inv.m). The mean of $\mathsf{ln}(\chi^2(2)))$=$\psi(1)$+$\mathsf{ln}(2)$ which is equal to $\mathsf{ln}(\mathsf{geomean}(\chi^2(2)))$ where geomean refers to the geometric mean, as opposed to the arithmetic mean, or common mean (Lee, 2012; verified in MATLAB). This distinction is important because it allowed us to correct a

glitch in the original BOSC code where the power thresholds for detection are determined using the regression values converted from log form and the $\chi^2(2)$ mean of 2 when the $\chi^2(2)$ geometric mean of $e^{\psi(1)+ln(2)}$ should be used instead.[3] This is also relevant for deriving the thresholds used in the high-power removal method along with scaling down the power values for KS tests when regression values are used. Using the medians instead of the means did increase computation time (Figure 9), presumably because medians require sorting.

3.3. Removing high-power values

One of the main reasons power distributions deviate from the theoretical $\chi^2(2)$ distribution is due to an abundance of extreme, high-power values (Figure 2A, overflow bar on right). If we could remove those, we would have a more pure distribution, which would allow us to selectively fit the background signal more accurately since the influence of those non-background values would no longer be present. After an initial regression, the fit values were used to scale the $\chi^2(2)$ distribution with the power value distributions. Power values exceeding a high threshold based on the $\chi^2(2)$ distribution (specifically, the 99.9th percentile) were removed at each frequency and then another regression was performed using the mean of the remaining power values (again converted to log form first). Overall this proved to be an effective improvement on the sliding-window and partial-spectrum tests (Figure 7 and Figure 8 respectively), but it did increase computation time due to the additional regression and high-power removal operations (Figure 9).

3.4. Using a subset of frequencies

We reasoned that the background fit should be more accurate and robust if we only use frequencies that have little oscillatory activity (or artifact) present before performing the regression. These "best" frequencies were determined in an initial regression using KS d scaled with regression values to the $\chi^2(2)$ distribution just as it was used as an outcome measure. However, other ways of obtaining the KS d values were also tested, namely using the actual mean of power values at each frequency, using the median power values, and using the regression values before removing values above a threshold (the 99.9th percentile of the $\chi^2(2)$ distribution) and then using the mean of the remaining subsample. Scaling using mean power values produced inconsistent patterns with the peak frequencies sometimes having the lowest d-values, but scaling using the regression values was much more consistent and the peak frequencies usually had much larger d-values compared to other frequencies (Figure A2A,C). The method of using the regression values, removing high values with a 99.9% threshold and then using the subsample generally showed this pattern too and the KS d values were generally much lower, but the peak frequencies were not as distinguishable, but scaling using median power values produced results with low KS d values as well but also tended to have distinguishable peak frequencies (Figure A2B,D).

Initially we also investigated SD(f) and CV(f) since any frequency with a peak should show high general deviation a lot from our model of the background. We mainly looked

[3]This was first flagged by Seymour et al. (2022) in the fBOSC code.

for whether the peak frequencies in the power spectrum also produced peaks (or potentially valleys) in the pattern of each outcome measure which would single them out. We first tested SD(f) with the idea that less variance means that frequency may better represent the background and would bring with it a benefit of computational simplicity. However, this measure proved to be inconsistent; sometimes peak frequencies had high SD values but not other times. Next we tried CV(f), hoping it would compensate for the scale differences in the 1/f form, but CV was even less effective (Figure A1). With both of these measures proving ineffective it was clear that KS d had more potential, and after testing the four versions of obtaining KS d on the sliding-window test the method using regression values proved to be most effective. However, the benefits were marginal (Figure 7), and using KS d substantially increased run-time due to the complexity of kstest.m, especially the sort functions within. Using custom code for the KS tests did largely improve the computation time but it was still relatively costly (Figure 9). This frequency subsetting approach was not included in the optimized method, although KS d proved to be a useful outcome measure for assessing the quality of a background estimate.

3.5. The best-performing method used a combination of modifications

In addition to the standalone modifications, we tried many combinations of the pipeline using the different variants of each modification (Figure 6). The method that showed the most overall improvement over standard BOSC involved a combination of removing high-power values, using median-based regression instead of mean-based and using robust regression (Figure 7 and Figure 8). Specifically, this method started with an initial median-based regression followed by removing high-power values above thresholds based on the initial regression values and the 99.9th percentile of a $\chi^2(2)$ distribution before performing a second regression that is also median-based but also used robustfit.m. The combination of all these modifications did substantially increase computation time, particularly the combination of median-based regressions and high-power removal (Figure 9).

4. Comparing the new optimized BOSC method to the standard version

4.1. Oscillation-detection performance on experimental data

We first compared the detection performance of the two methods using the same experimental data from earlier when testing the various modifications. Specifically, comparison between methods was done using the $P_{episode}$ values (probability of detection within a time window) along with a visual inspection of detections from each method (Figure 10). In general, the optimized BOSC method resulted in slightly more detections across frequency and had lower power thresholds for detection since it was less affected by high-power values. However, the detection rate for signals with power spectra containing no peak or an alpha peak did not differ by much. Differences became more apparent when using partial spectra for the same alpha peak signals just like the more apparent difference in background fit quality, with optimized BOSC showing much more consistency between partial and full spectra.

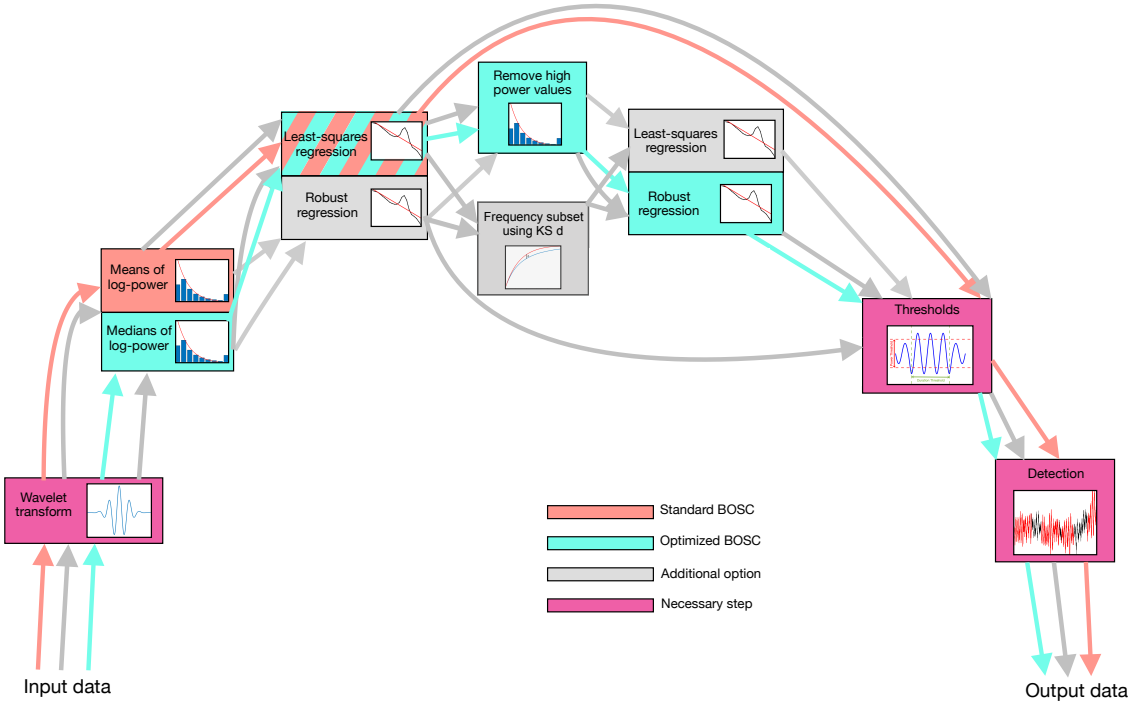

Figure 6: Flowchart illustrating the possible steps that can be taken by different versions of the BOSC method, with the five-step standard method and seven-step optimized method shown with light red and cyan arrows respectively. Also shown are options that were used by neither but could be used in different variations of BOSC, along with the necessary steps that are essential to BOSC. While the regression for the background fit is not colour-coded as a necessary step, at least one regression must be performed. The icons within each box are: a Morlet wavelet for the wavelet transform step, a probability distribution function of power for steps involving it (means and medians of log-power, remove high power), a power spectrum with a background fit for the regression steps, a cumulative distribution function and d statistic for the frequency subset step, and a threshold diagram and sample detection output for the last two corresponding steps.

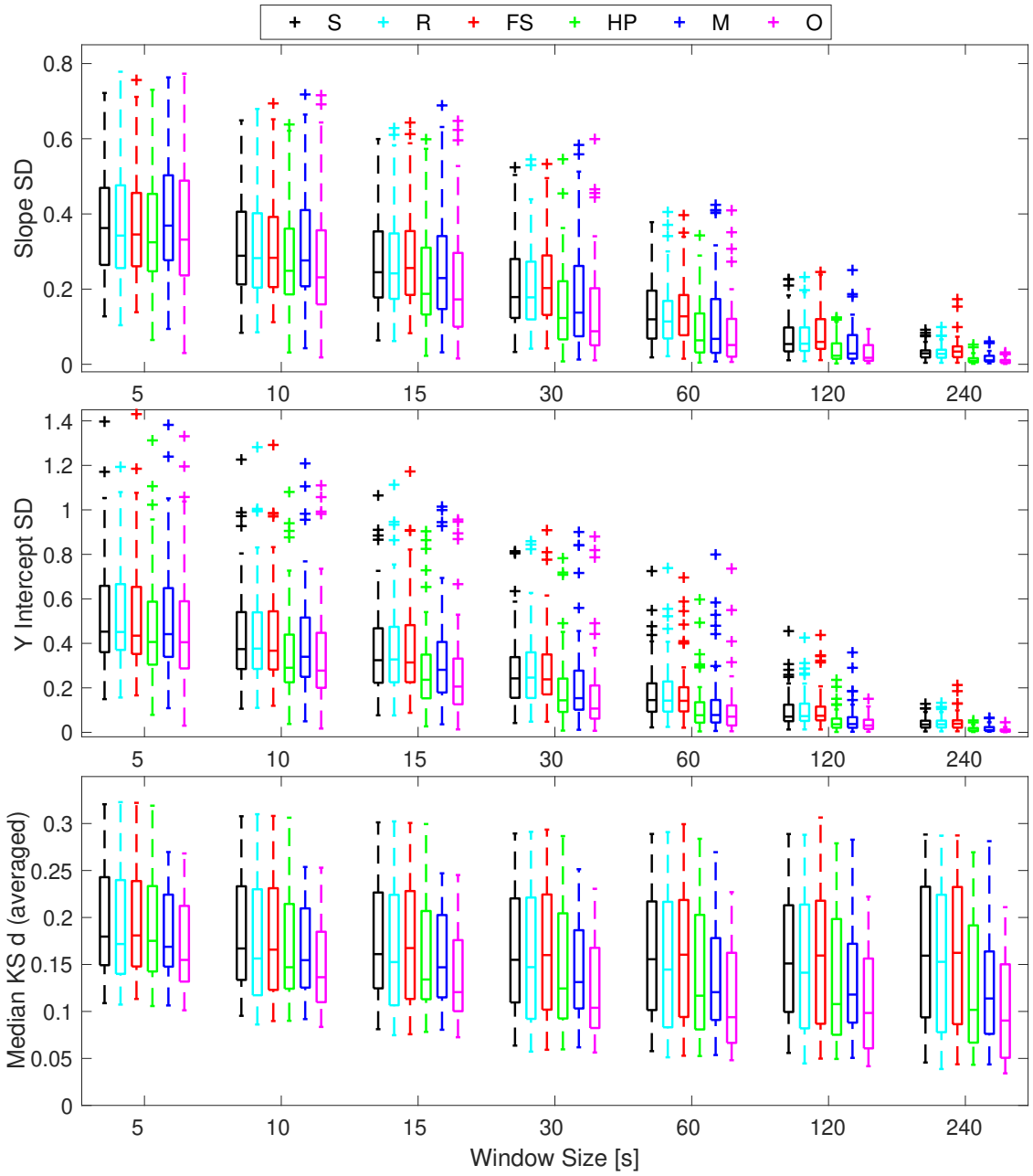

Figure 7: Slope standard deviation (top), y-intercept standard deviation (middle) and median KS d (bottom) as functions of window size using the 48 datasets for each method. Methods are depicted in different colours; referring to the letters in the legend at the top: S - the standard BOSC method, R - robust regression, FS - frequency subset method, HP - remove high-power, M - median-based regression and O - optimized BOSC method. For slope and intercept SD, optimized BOSC performs better than standard BOSC even at shorter windows (e.g., 15 s optimized vs 30 s standard) but the pattern of worsening relative performance below 30 s remains. For median KS d this is also the case for optimized BOSC such that performance improves with longer windows as opposed to standard BOSC where it stays relatively constant, so the performance difference increases with longer windows.

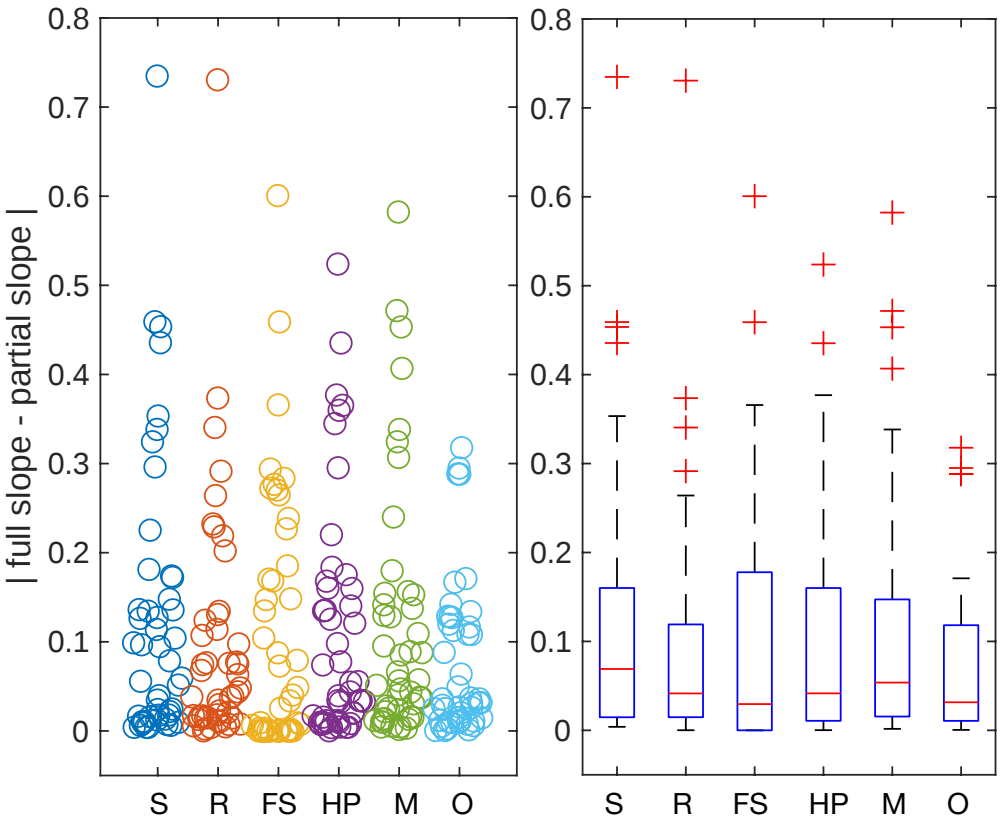

Figure 8: Swarm (left) and box (right) plots for standard BOSC and tested modifications showing the slope difference between the full spectrum and partial spectrum for the 48 datasets. The high-power removal method (HP) and robust regression (R) show the smallest differences, along with the frequency subset method (FS) but with more spread. The former two were combined with the median (M) to form our fully optimized BOSC method (O). The standard BOSC method is labelled S.

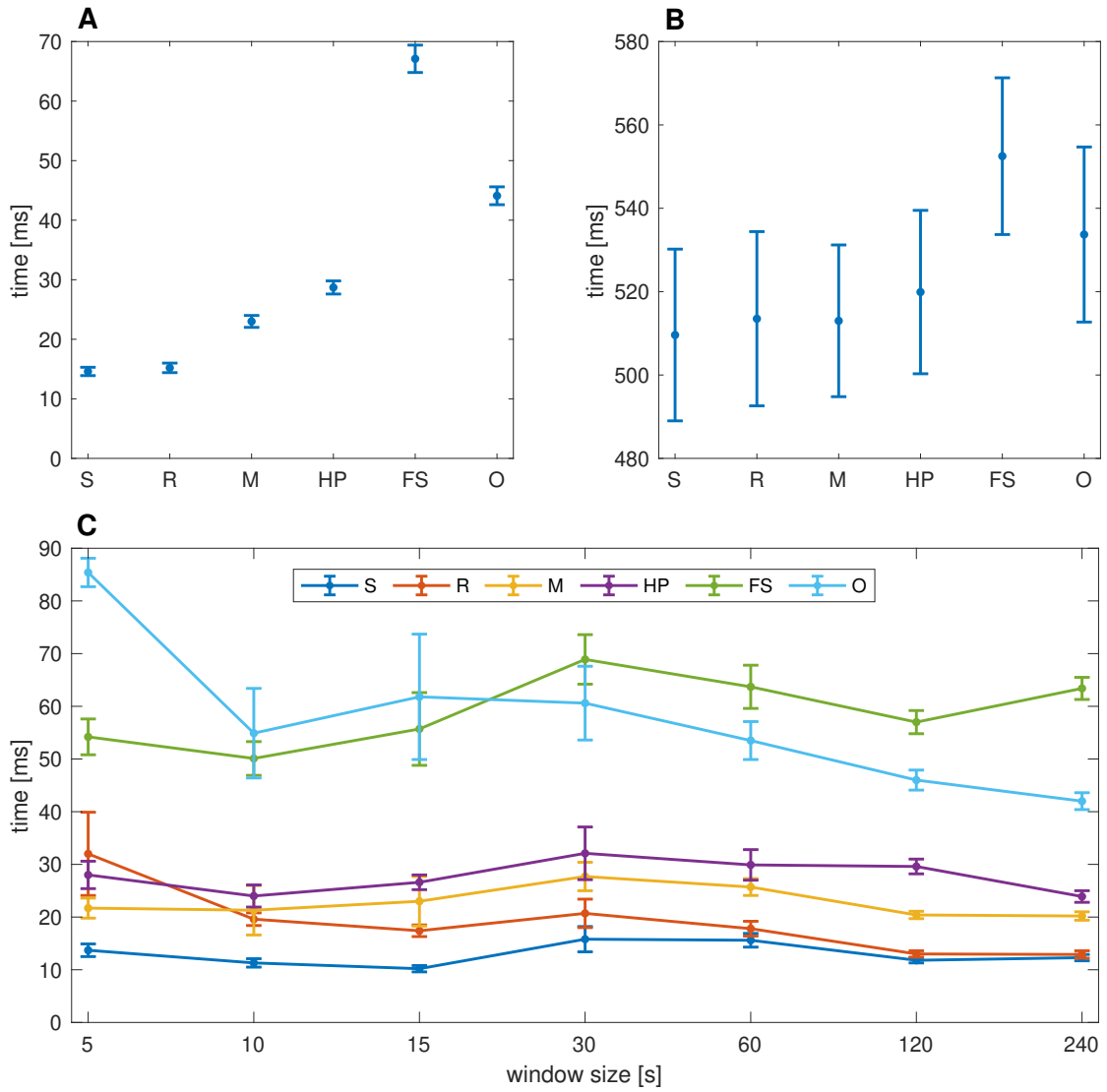

Figure 9: Computation times for standard BOSC (S), robust regression (R), median-based regression (M), removal of high-power (HP), the frequency-subset method (FS) and the fully optimized BOSC (O). the individual modifications and the combination used in our optimized version of BOSC. (A) computation time for the background fit alone for each method. (B) computation time for the entire BOSC method from the wavelet transform to retrieving $P_{episode}$ values after detection. Computation time was averaged over $N = 100$ trials. (C) computation time for the background fit alone for each window size (repeated for a respective number of windows) for each method. The number of windows decreased at the rate the window size increased so the amount of data being analyzed remained constant. For all three tests computation time values were averaged over N=100 trials, in (C) one trial considered all the windows for each window size (e.g., 48 × 5 s windows or 1 × 240 s window). A simulated signal with 10 Hz oscillations, identical to that seen in Figures 11A and 12 was used for all three tests, which were performed using a 2019 iMac running MacOS 10.15.7 with a 3.6 GHz 8-core Intel Core i9 processor.

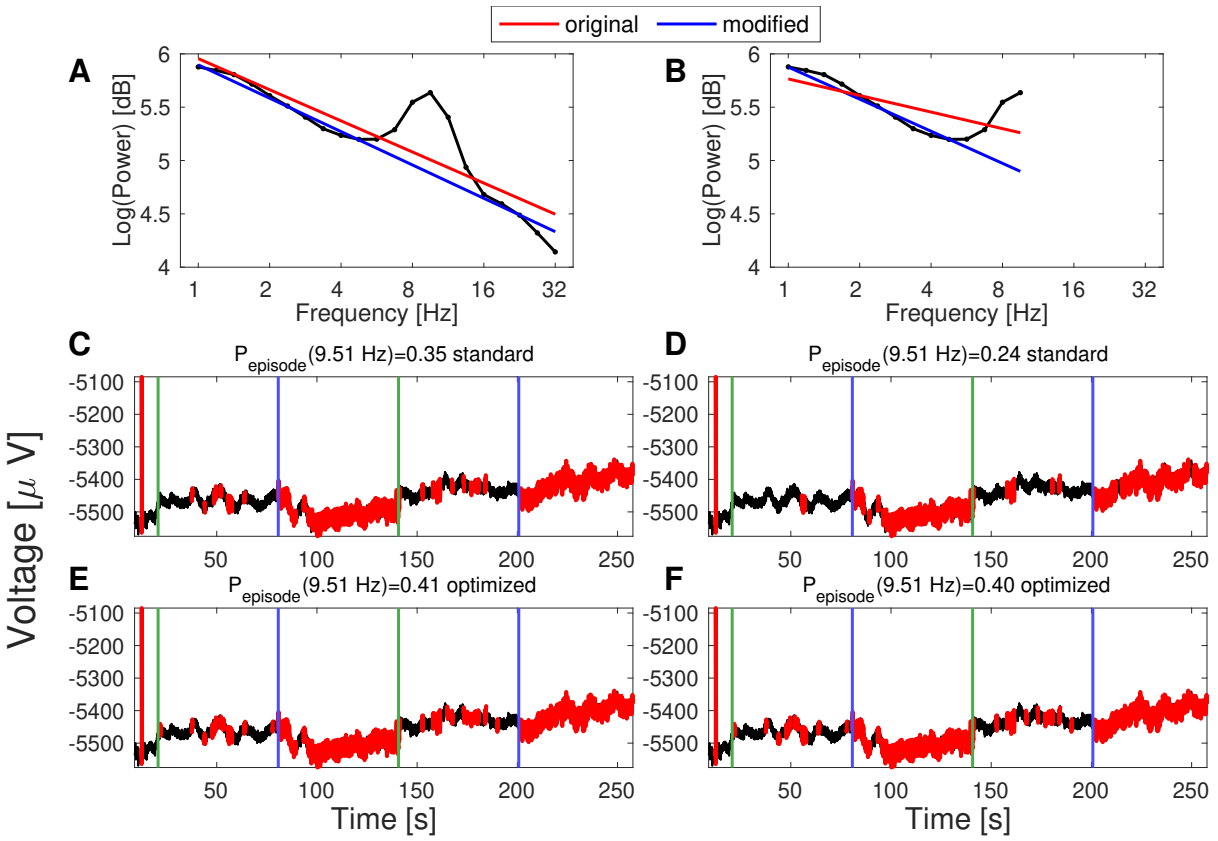

Figure 10: Background fit and detection consistency for the standard BOSC method and the new optimized version using full and partial spectra. Background fits are shown by the power spectra with (A) showing the spectrum for the full frequency range and (B) showing the same for the partial spectrum that cuts off at the alpha peak. (A) also shows fit of standard BOSC being pulled up by the peak compared to optimized BOSC, but (B) shows the fit of standard BOSC being tilted as well while optimized BOSC is barely affected. C–F show the raw signal (black) with detected episodes (red) superimposed, also points where participants were asked to open their eyes and close their eyes are indicated with green and blue vertical lines respectively. (C) shows the detections and $P_{episode}$ from standard BOSC when using the full spectrum while (D) shows the same for standard BOSC with the partial spectrum. The same information is shown for optimized BOSC with (E) showing the results from using the full spectrum and (F) showing the results of using the partial spectrum. Immediately notable is the difference in detection consistency as shown by the $P_{episode}$ values from full and partial spectra for optimized BOSC (E and F) being very close (0.41 vs 0.40) while the $P_{episode}$ values from (C) and (D) for standard BOSC are more distant (0.35 vs 0.24). While (C) and (E) show slightly less detection with the full spectrum for standard BOSC vs optimized BOSC, this is due to the higher power thresholds in again shown in (A). The greater difference using the partial spectrum, shown by (D) and (F) is due to the threshold bias across frequency that was shown by the tilted slope of standard BOSC in (B). While the $P_{episode}$ values differ, visual distinction is better highlighted in a zoomed-in portion during the eyes-closed phase in Figure A3.

4.2. Assessing background fit and detection with simulated signals

We evaluated background fit by checking how much each method deviated as a result of the various peaks and detection performance again used consistency again but between the extreme frequency peaks and more centered peaks in alpha (10 Hz specifically). We also tested the performance of robust regression by itself as this was one of the main changes in methods like eBOSC when estimating the background spectrum. In doing so we found that background fit quality was on par with optimized BOSC for these simulations and could even slightly exceed it in certain cases (Figure 11A,B,E). This is likely because the robust regression can still handle extremely high power values and there are no other artifacts in the simulations. This carries over to detection performance (Table 1 and 2).

To distinguish the capabilities of optimized BOSC from robust regression we focused on how the latter cannot address skewed values that are not outliers. A good example was the skew from high-power values in the empirical data, so we used simulated signals designed to mimic the effects of high-power values from 1–4 Hz. During tests with these simulated signals, optimized BOSC showed better background fit quality compared to robust regression and standard BOSC (Figure 11C,D). Another more complex simulation produced two peaks at 10 Hz and 20 Hz, along with a 16 Hz peak in between but with half the amplitude in order to make the peaks appear wider and connected. This would mean more points would be affected instead of a few that could be recognized as outliers, however, this example not only decreased the performance of robust regression but of optimized BOSC as well especially as the oscillations grew larger (Figure 11F). Detection performance showed very similar patterns as background fit quality (Table 1 and 2) Therefore, this example simulation helps establish the limits of this optimized BOSC method.

Overall, simulated signals with peaks at extremely high or low frequencies showed differences in both background fit quality and detection performance similar to that seen with the partial spectra.

4.3. Sliding-window tests with simulated signals

Next we performed another sliding-window test between the original BOSC method and the new optimized form using the simulated signals. These compared six different oscillation amplitudes so as a result there were six data points per window size as opposed to the 48 (participants) from the empirical data. Performance was once again evaluated with the standard deviation of the slope and y-intercept of the background fit (slope SD and y-intercept SD) and with resulting median KS d values by window size. For slope SD and y-intercept SD, specifically, the performance difference of each method often remained similar at each window size, with the exception of the windows spanning 120 s and 240 s where performance was nearly identical (Figure 13 and 14). For median KS d this was not the case and the trend was reversed to some degree with increased differences between the two methods at longer windows (Figure 15).

The general performance for both methods decreased on all metrics for simulations with higher oscillation amplitude and the new optimized method performed better in the vast majority of instances. The exceptions to this were simulations with two peaks at 10 Hz and 20 Hz and a similar simulation with a small 16 Hz peak added in between to connect

	$P_{episode}$ standard BOSC	$P_{episode}$ robustfit	$P_{episode}$ optimized BOSC
10 Hz	0.44	0.45	0.45
20 Hz	0.42	0.44	0.44
10 Hz and high-power	0.42	0.43	0.44
20 Hz and high-power	0.42	0.43	0.44
10 Hz and 20 Hz	0.36	0.43	0.43
10 Hz and 20 Hz (wide)	0.30	0.32	0.32

Table 1: $P_{episode}$ values for standard BOSC, robust regression and optimized BOSC for the six simulated signals used for the spectra in Figure 11. Peak frequencies are shown in the table rows and other features that match the power spectra so Figure 11A matches with the 10 Hz row, B matches with 20 Hz, C with 10 Hz and high-power, D with 20 Hz and high-power, E with 10 Hz and 20 Hz, and F with 10 Hz and 20 Hz (wide). These $P_{episode}$ values are from the entire signal length and the the true proportion of oscillations (theoretical $P_{episode}$) included with the noise is 0.46. This is due to the noise generation seed remaining constant in the eBOSC code we adopted (Kosciessa et al., 2020). The performance differences between the three methods are similar to those seen in background fit quality with 10 Hz having the smallest difference between the three. Similar performance is seen for robust regression and optimized BOSC for the other simulation types apart from the high-power simulations, where optimized BOSC has the advantage. A visual representation of the values from the simulation with 10 Hz oscillations is shown in Figure 12A.

	$P_{episode}$ standard BOSC	$P_{episode}$ robustfit	$P_{episode}$ optimized BOSC
10 Hz	0.73	0.80	0.80
20 Hz	0.84	0.92	0.92
10 Hz and high-power	0.68	0.73	0.77
20 Hz and high-power	0.84	0.86	0.90
10 Hz and 20 Hz	0.70	0.90	0.89
10 Hz and 20 Hz (wide)	0.75	0.81	0.81

Table 2: $P_{episode}$ values for standard BOSC, robust regression and optimized BOSC for the six simulated signals used for the spectra in Figure 11 but these values are from a small 5-s window with the true proportion of oscillations (theoretical $P_{episode}$) included with the noise being 1.00, again due to the noise generation seed remaining constant. The smaller window retains the differences shown in Table 1 but allows them to be more pronounced, although the 10 Hz + 20 Hz simulation gives a slight edge to robust regression over optimized BOSC in a special case. A visual representation of the values from the simulation with 10 Hz oscillations is shown in Figure 12B.

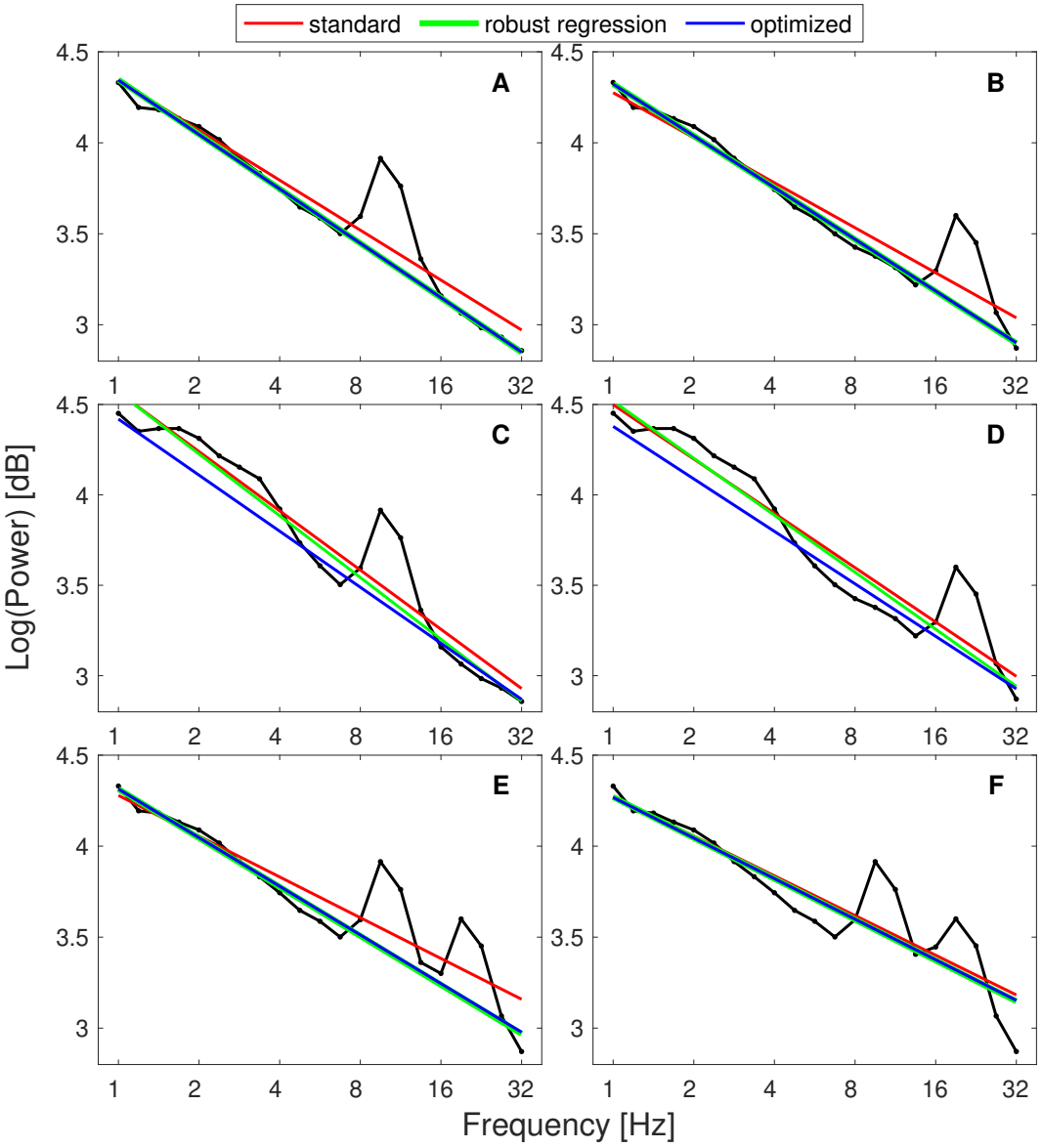

Figure 11: Background fits for the standard BOSC method (red), BOSC with robust regression (green), and the optimized BOSC version (blue) using power spectra of simulated signals. (A) Power spectrum and background fits for a simulated signal with a 10 Hz alpha peak. Robust regression and optimized BOSC overlap almost entirely. (B) A simulated signal with a 20 Hz beta peak. The fit of standard BOSC is tilted more than the 10 Hz peak signal, similar to the partial vs full spectrum. Again, robust regression and optimized BOSC overlap almost entirely. (C) A simulated signal with a 10 Hz beta peak and added high-power values from 1-4 Hz. The fits for original BOSC and robust regression are both affected by the high-power values such that they are tilted the opposite way as the signals without them. (D) A simulated signal similar to (C) but with a 20 Hz peak instead. Again standard BOSC along with robust regression are skewed while optimized BOSC is resilient. (E) A simulated signal with 10 Hz and 20 Hz peaks. Robust regression and optimized BOSC perform very similarly with a very slight edge to robust regression in this special case. (F) A simulated signal with 10 Hz and 20 Hz peaks, along with a smaller 16 Hz peak to reduce the valley in between. Robust regression and optimized BOSC perform similarly again and the difference between them and standard BOSC is much smaller.

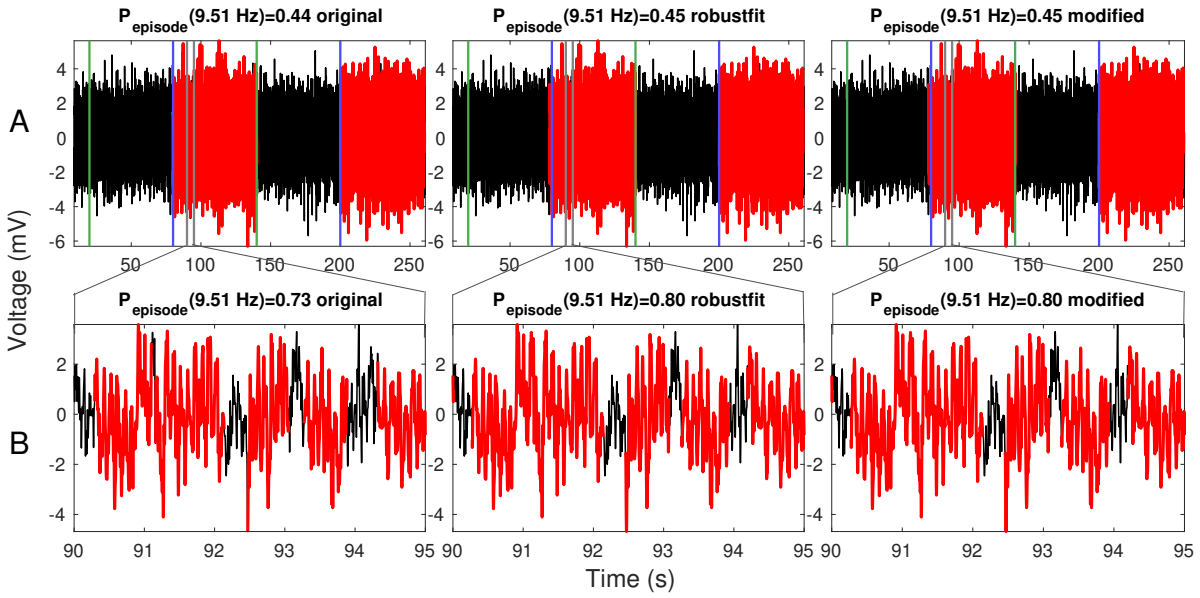

Figure 12: Detection comparison example for simulated signals between standard BOSC, robust regression (robustfit.m) and optimized BOSC again using graphs showing detected episodes (in red) superimposed on the raw signal (black). The green and blue lines indicate where oscillations begin (blue) and end (green) similar to the indicators for participants to close or open their eyes. This particular example corresponds to Figure 11A. The top row (A) shows detections and $P_{episode}$ values for the entire signal and the bottom row (B) is zoomed in to a particular section to highlight greater differences on a small scale.

the two peaks, which skewed both methods. The optimized BOSC method had the clearest advantage with the high-power value simulations. Shorter time windows also resulted in worse performance like the empirical data with the exception of a new outcome measure $|\mu_{slope} - (-1)|$ which remained relatively constant in most instances. This represented the absolute difference between the mean slope and –1, the preset slope of these simulations using just $1/f$ noise so smaller differences indicated a better match to the true slope (Figure 16).

In sum, the simulated signal, for which we know the ground truth, confirmed the superiority of the optimized BOSC method, including its stability at short windows and special advantage with high-power oscillations.

5. Discussion

In this paper we assessed, for the first time, the robustness of the background fit from the standard BOSC method using short time windows within a larger signal. We found that background fitting decreased in performance substantially for windows < 30 s in our ~ 4.5-minute empirical signals for all outcome measures (slope and y-intercept SD, median KS d). However, this does vary depending on the unique qualities of each signal in question. This is exemplified by how performance on all metrics was better at all windows for the synthetic data which was free from gain shifts and generally had much more controlled variation. Also for the first time, we analyzed the robustness of the BOSC method using signals with

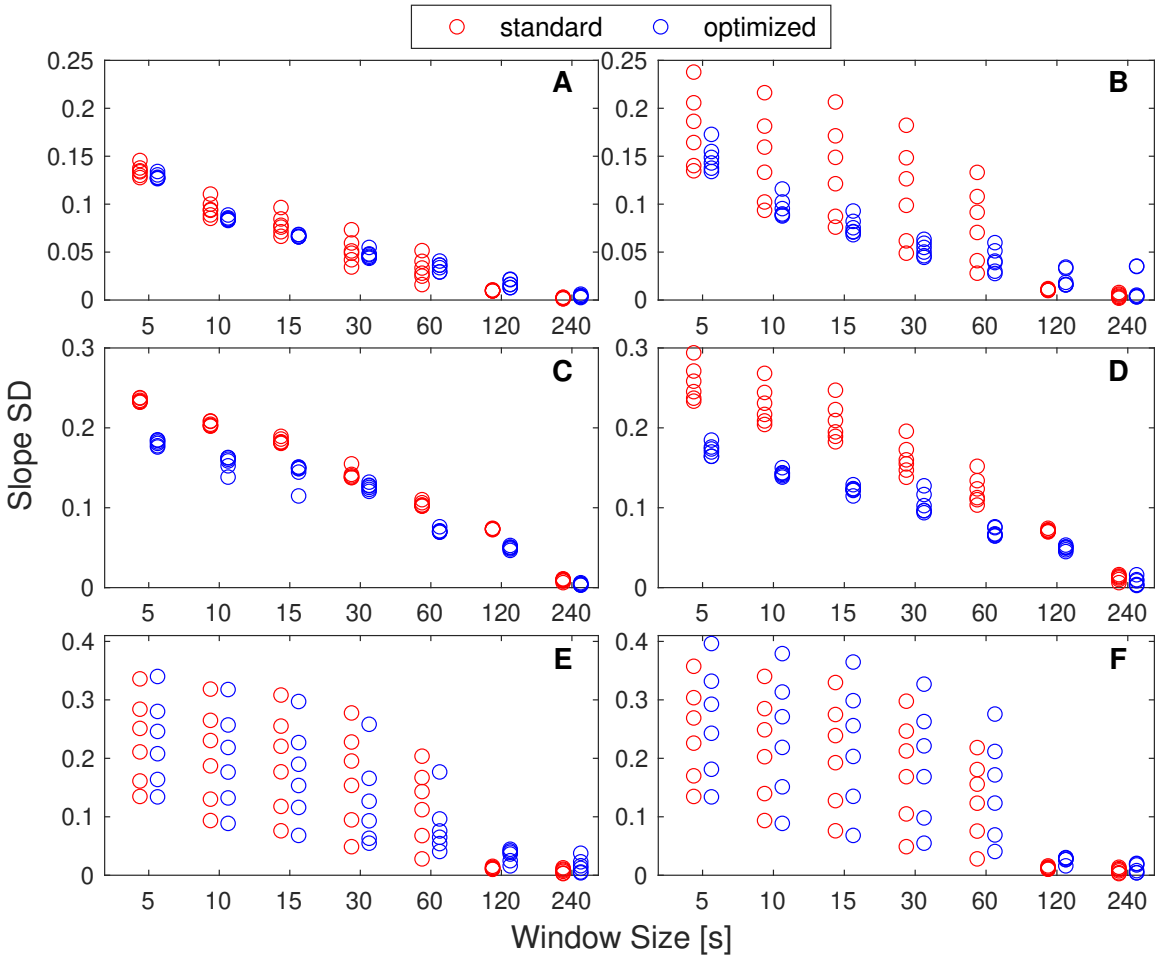

Figure 13: Slope standard deviation (slope SD) for the various kinds of simulation types. The red points are from standard BOSC and the blue points are from optimized BOSC. The x-values (window sizes) for the red and blue points are the same they are just plotted side-by-side, the same is true for Figures 14, 15, and 16. (A) The simulation with a 10 Hz alpha peak. (B) The simulation with a 20 Hz beta peak. The high-power simulations show the biggest difference with (C) being high-power and a 10 Hz peak and (D) being high-power and a 20 Hz peak. (E) shows a simulation with both a 10 Hz and 20 Hz peak and (F) is a similar simulation, but with a smaller 16 Hz peak in between to make the other two peaks wider and connected. These two instances show the lowest difference with the optimized BOSC showing worse performance for (F), at least on this particular metric.

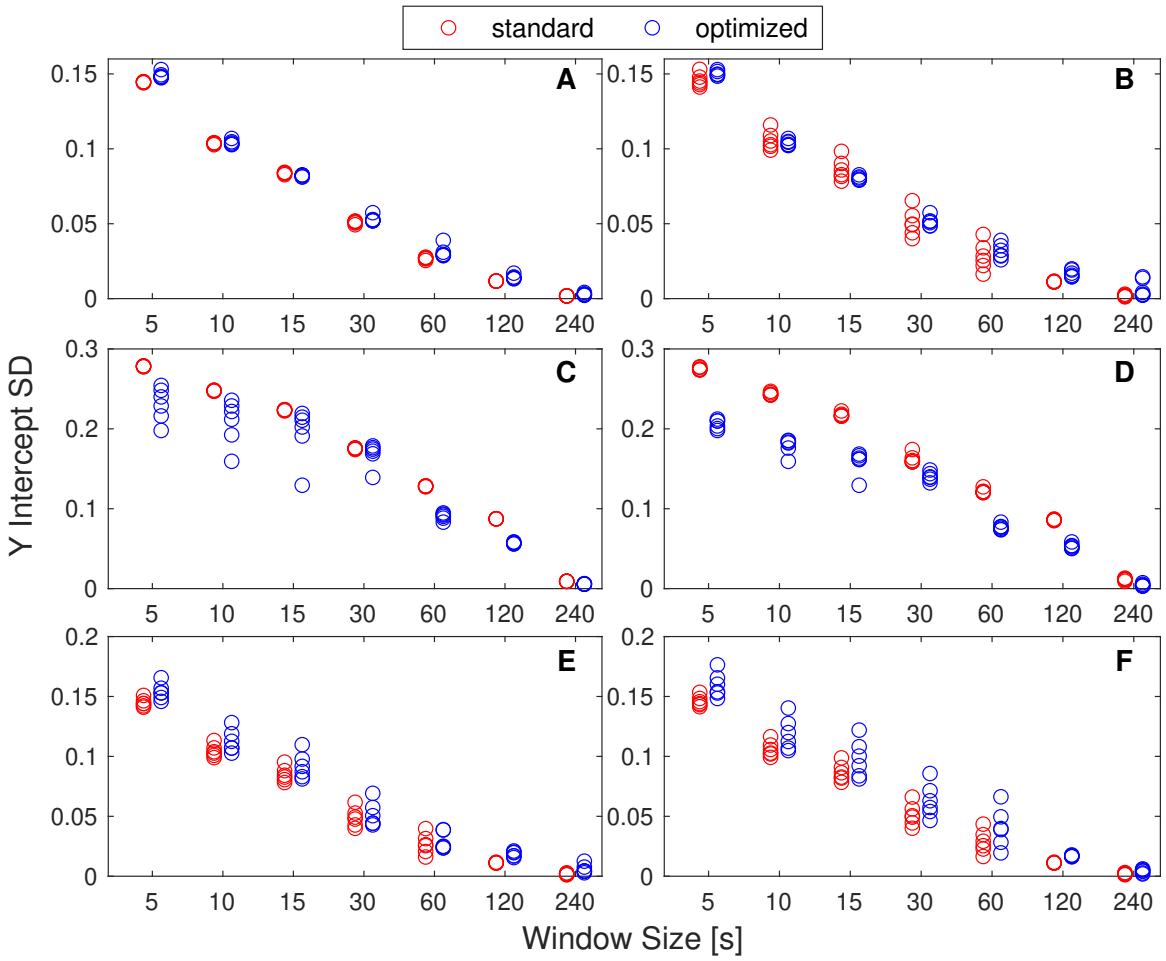

Figure 14: Y-intercept standard deviation (y-intercept SD) for the various kinds of simulation types. The colour coding and subplots match Figure 13. Patterns are generally similar, with the biggest differences in the high-power simulations (C) and (D) and the lowest performance of optimized BOSC in (E) and (F). The clustering in (A) for standard BOSC that is not present in Figure 13A may be due to coincidental convergence of lines with different slopes on similar intercepts.

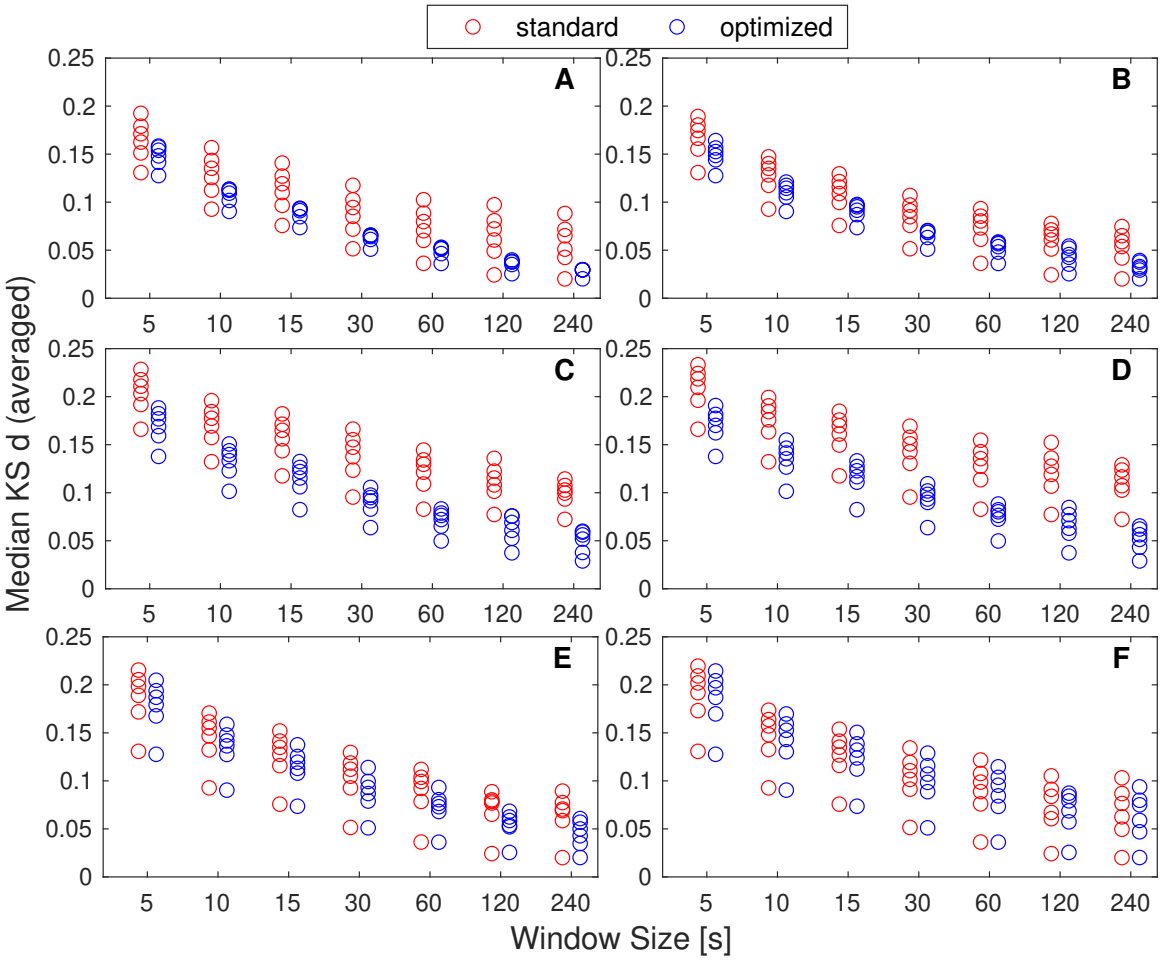

Figure 15: Median KS d for the various kinds of simulation types. The colour coding and subplots match Figure 13. Optimized BOSC performs better on all types but differences are generally less prominent for this particular metric. Again the biggest differences are in the high-power simulations (C) and (D) while (E) and (F) show the smallest differences.

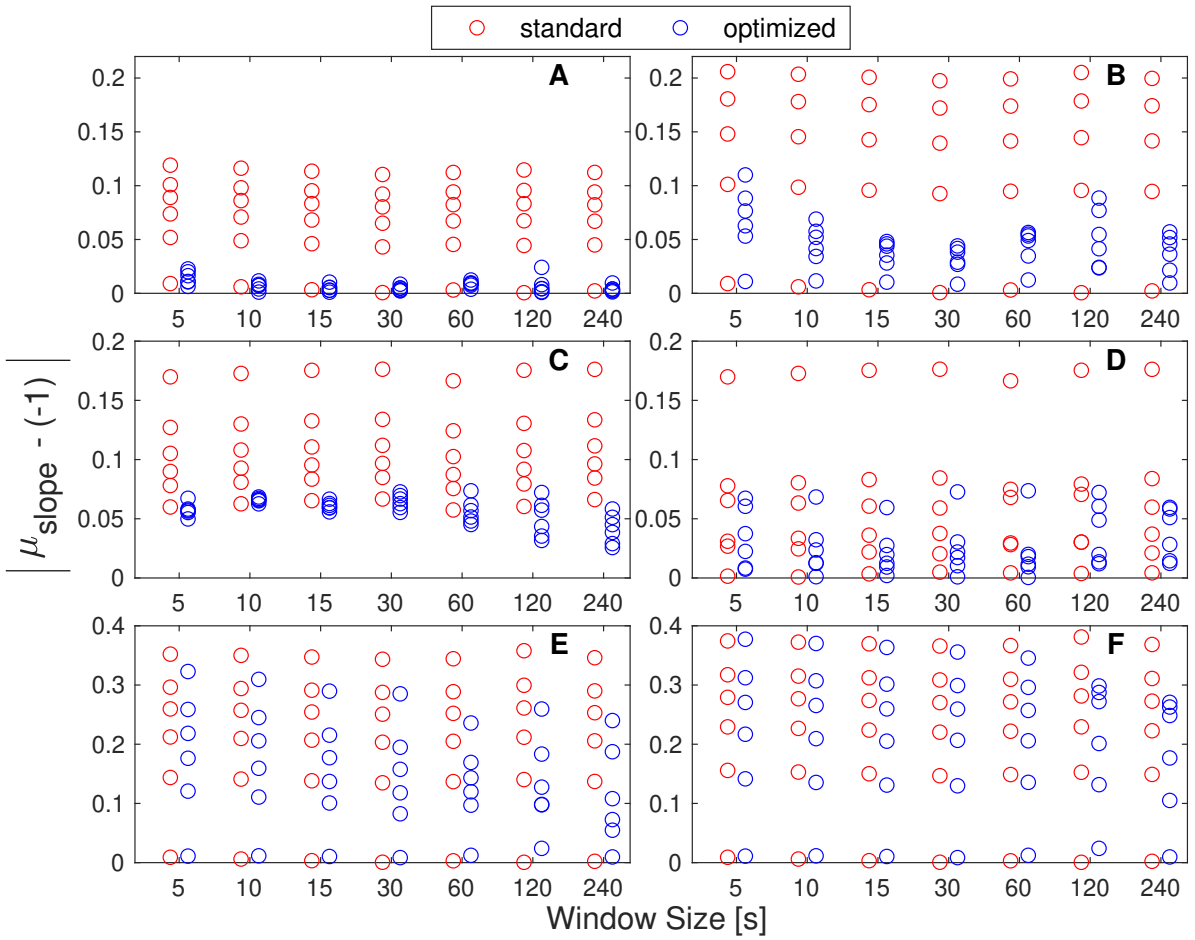

Figure 16: $|\mu_{slope} - (-1)|$, (absolute values) for the various kinds of simulation types. The colour coding and subplots match Figure 13. Like with median KS d, optimized BOSC performs better in all instances; however, differences in (A) and (B) are more pronounced to the point where they are comparable to the high-power simulations (C) and (D). (E) and (F) once again show the smallest differences.

oscillations near the highest analyzed frequency (32 Hz). We found that these scenarios are quite devastating to the background fit and this becomes exacerbated with larger peaks in the power spectrum. Even if the peak is in the middle of the spectrum it can still be problematic if the peak is extremely large. We designed a new version of the BOSC method to optimize performance in these areas in order to address these vulnerabilities.

5.1. Our optimized BOSC method

Based on the results from the two tests, we modified the BOSC method to: remove high-power values, use medians rather than means of power when performing the background fit, and use robust regression in place of least-squares regression. While more focus was given to the sliding-window test, the use of the two tests helps complement the limitation of each as an operationalization of general robustness. A good example is how the robust regression only made a large difference in the partial-spectrum test and its value may not have been clear if we had only used the sliding-window test. As explained earlier, the partial-spectrum test also relates to the later tests we did with simulated signals with peaks and very high/low frequencies by emulating a signal with a peak at the highest analyzed frequency. Therefore, the findings from the partial-spectrum test relate to and are expanded upon with those from the later tests with simulated signals. For the sliding-window test, optimized BOSC provided a strong advantage over standard BOSC with respect to the consistency and alignment of background estimates with theoretical standards such that optimized BOSC had better performance at windows 15 s or smaller compared to 30 s windows using standard BOSC (out of the $\sim$ 4.5 minute length for the empirical signals). The overall pattern of decreased performance for windows under 30 s seen with standard BOSC did remain with optimized BOSC but the differences were less prominent. While the consistency (standard deviation of slopes and intercepts) of background estimates was first tested, the (median KS d) refers to the degree of alignment between the power values and the theoretical $\chi^2(2)$ distribution.

5.2. Overall comparison with standard BOSC

Having evaluated background fit quality and detection using both empirical and simulated signals, it is clear that the optimized BOSC method is most helpful when a signal has spectral peaks at very high or low frequencies along with various artifacts like high-power values. Because optimized BOSC is designed to provide a fit based on theoretical assumptions, it has more ways of dealing with artifacts that may not be identifiable visually or as statistical outliers. That said, the detection performance does not differ substantially between the standard and optimized BOSC methods for signals with no peaks or a peak in the center of the power spectrum (like alpha range), along with having little anomalies and artifacts. This suggests that the standard BOSC method performs sufficiently well in these cases, and even though a spectral peak may skew the regression of the standard BOSC method slightly upwards, the line is not tilted. The effect is similar to an increase in the power thresholds, making oscillation detection a bit more conservative at all frequencies. At the opposite end, the optimized BOSC method also has some remaining limitations as

shown with the more complex simulations. It cannot account for all artifacts, especially of large size and quantity so in those extreme cases manual examination is still required.

5.3. Recommendations

For signals with no prominent peak or a peak near the middle of a power spectrum, the standard BOSC method performs well. For signals with spectral peaks near the edges of the analyzed range of frequencies, artifacts like excessive high-power values or with short windows within a larger signal, the optimized BOSC method offers superior performance. The presence of influential high-power values was certainly not uncommon in our empirical data so mitigating the effects of them is definitely helpful. Even with standard BOSC a comparison of power values to the theoretical $\chi^2(2)$ distribution using KS tests can be beneficial since the influence of high-power values is not always apparent. If there are high KS d values in non-peak frequencies then it may be worth using the optimized method over the original. Our consistent focus on the expected $\chi^2(2)$ distribution of background power during the development of the optimized BOSC method provided us with the opportunity to perform evaluations with objective criteria rather than assuming the closest fit is most aligned with the true background. Despite the success of optimized BOSC in the above cases, there are still situations where a particular signal may contain so many large oscillations or artifacts that neither method will work well enough. These rare cases would still have to be examined manually and handled as special cases.

Finally, although we focused on BOSC, a method that was designed to view oscillations as discretely turning on and off (although potentially varying in amplitude), the lessons learned may be useful for methods of analyzing spectral features of EEG that stop short of thresholding, if they also rely on estimates of the coloured-noise background.

5.4. Limitations

We took a systematic approach to assessing modifications of the background-fitting method but restricted our tests to a single data set focusing on posterior alpha oscillations related to visual inattention. Additionally, robustness is very general so it can be operationalized and tested using many different metrics and while our tests highlighted various strengths and weaknesses of BOSC, they certainly were not all-encompassing. While we used simulated signals with oscillations at various or multiple frequencies aside from alpha, oscillations in non-human animals with different characteristics may benefit more or less from our optimized BOSC method. However, BOSC has been successfully applied to invasive recordings from humans and non-human animals (e.g., Caplan et al., 2001; Ekstrom et al., 2005; Watrous et al., 2011; Jutras et al., 2013; Hughes et al., 2012). Moreover, Whitten et al. (2011) and Hughes et al. (2012) found consistency between human scalp-recorded alpha rhythms and rat invasive hippocampal recordings of theta and slow oscillations in that the standard BOSC method showed similar resilience to the precise power thresholds. There is therefore good reason to expect that the improvements considered here will produce similar results in different types of recordings.

Another limitation is that while this optimized BOSC method deals with high-power values well, it is limited in the sense that while high-power values can be removed, absent

low-power values are not filled in. Therefore, if the absence of low-power is the reason power distributions do not fit a $\chi^2(2)$ distribution it cannot be adjusted. Also regarding the nature of the $1/f^\alpha$ background itself, a growing body of literature suggests that the α parameter can vary over ranges of frequency and the exact causal mechanisms behind the background noise have yet to be determined (He, 2014; Gyurkovics et al., 2021; Samaha and Cohen, 2022). The FOOOF algorithm (Donoghue et al., 2020) addresses this by estimating a non-linear background spectrum form by incorporating a "knee" parameter into the regression. This was combined with BOSC to make fBOSC (Seymour et al., 2022) and was expanded upon in the SPRiNT (Spectral Parameterization Resolved in Time) algorithm (Wilson et al., 2022) to address possible time-variance in the $1/f^\alpha$ background. An alternative to FOOOF, Greenberg et al. (2016) replaced the linear regression in BOSC with a quadratic regression to account for curved spectra. These methods do not remove less-obvious artifacts like high-power values, however, which raises the question of how the modifications implemented here may synergize with nonlinear fitting approaches, or if further adaptation is needed.

5.5. Conclusions

In this study we optimized the BOSC method such that it can, in many cases, automatically account for the influence of not only spectral peaks but also high-power values that have a less obvious influence when estimating the $1/f^\alpha$ background. This is because the selected modifications work to bring the power value distribution at each frequency as close as possible to the theoretical $\chi^2(2)$ distribution in addition to downweighting the influence of outliers in power spectra. While limited to linear fitting, this method uses a theoretical standard for assessing the quality of a background estimate rather than assuming that the tightest fit to non-peak frequencies is the best estimate, and this is what distinguishes it from other methods. In terms of application, the tests we carried out show that this optimized BOSC method has clear advantages over the original when using shortened windows within a larger signal along with when spectral peaks are present near the edges of a power spectrum and when a signal has high-power artifacts. However, for analysis of whole signals that are relatively artifact-free and have an alpha or theta peak or none at all, the difference is negligible and the standard BOSC method performs well. Since the influence of high-power values is often not obvious, comparing the power distribution to the $\chi^2(2)$ distribution with KS d can certainly be beneficial when using standard BOSC to check if it is more appropriate to switch to the more robust yet more complex optimized BOSC method or if the standard BOSC method is well suited to the case.

6. Data and Code Availability

The core scripts for the optimized BOSC method along with those for generating simulated data can be found at: https://github.com/kapawluk/optimized-BOSC

The simulated data was generated using scripts derived from the eBOSC project (Kosciessa et al., 2020), with the relevant code found at: https://github.com/jkosciessa/eBOSC/

7. Author Contributions

Kieran A. Pawluk: formal analysis, investigation, software, conceptualization, methodology, visualization, writing draft, revision. Tamari Shalamberidze: data curation, resources, validation, reviewing and editing. Jeremy B. Caplan: conceptualization, funding acquisition, methodology, software, administration and supervision, reviewing and editing.

8. Funding

Partly supported by the Natural Sciences and Engineering Research Council of Canada and the Undergraduate Research Initiative at the University of Alberta.

9. Declaration of Competing Interests

The authors have no conflicts of interests to declare.

References

Berger, H., 1929. Über das elektrenkephalogramm des menschen. Archiv Für Psychiatrie Und Nervenkrankheiten 87, 527–570.

Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304, 1926–1929.

Caplan, J.B., Madsen, J.R., Raghavachari, S., Kahana, M.J., 2001. Distinct patterns of brain oscillations underlie two basic parameters of human maze learning. Journal of Neurophysiology 86, 368–380.

Donoghue, T., Haller, M., Peterson, E.J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A.H., Wallis, J.D., Knight, R.T., Shestyuk, A., Voytek, B., 2020. Parameterizing neural power spectra into periodic and aperiodic components. Nature Neuroscience 23, 1655–1665.

Ekstrom, A.D., Caplan, J.B., Ho, E., Shattuck, K., Fried, I., Kahana, M.J., 2005. Human hippocampal theta activity during virtual navigation. Hippocampus 15, 881–889.

Greenberg, A., Whitten, T.A., Dickson, C., 2016. Stimulating forebrain communications: Slow sinusoidal electric fields over frontal cortices dynamically modulate hippocampal activity and cortico-hippocampal interplay during slow-wave states. NeuroImage 133, 189–206.

Grossmann, A., Morlet, J., 1985. Decomposition of Functions into Wavelets of Constant Shape, and Related Transforms, in: Mathematics + Physics. World Scientific, pp. 135–165.

Gyurkovics, M., Clements, G.M., Low, K.A., Fabiani, M., Gratton, G., 2021. The impact of 1/f activity and baseline correction on the results and interpretation of time-frequency analyses of EEG/MEG data: A cautionary tale. NeuroImage 237, 118192.

He, B., 2014. Scale-free brain activity: past, present, and future. Trends in Cognitive Sciences 18, 480–487.

Hughes, A.M., Caplan, J.B., Dickson, C.T., Whitten, T.A., 2012. BOSC: A better oscillation detection method, extracts both sustained and transient rhythms from rat hippocampal recordings. Hippocampus 22, 1417–1428.

Jutras, M.J., Fries, P., Buffalo, E.A., 2013. Oscillatory activity in the monkey hippocampus during visual exploration and memory formation. Proceedings of the National Academy of Sciences of the USA 110, 13144–13149.

Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews 29, 169–195.

Kosciessa, J., Grandy, T. Garrett, D., Werkle-Bergner, M., 2020. Single-trial characterization of neural rhythms: Potential and challenges. NeuroImage 206, 116331.

Lee, P.M., 2012. Bayesian statistics: an introduction. John Wiley & Sons Inc.

Percival, D.B., Walden, A.T., 1993. Spectral analysis for physical applications: multitaper and conventional univariate techniques. Cambridge University Press.

Samaha, J., Cohen, M., 2022. Power spectrum slope confounds estimation of instantaneous oscillatory frequency. NeuroImage 250, 118929.

Schiff, S.J., Aldroubi, A., Unser, M., Sato, S., 1994. Fast wavelet transformation of EEG. Electroencephalography and Clinical Neurophysiology 91, 442–455.

Seymour, R., Alexander, N., Maguire, E., 2022. Robust estimation of 1/f activity improves oscillatory burst detection. European Journal of Neuroscience 56, 5836–5852.

Shlesinger, M., West, B., 1998. 1/f versus $1/f^{\alpha}$ noise, in: Random Fluctuations and Pattern Growth: Experiments and Models. Kluwer Dordrecht, pp. 320–324.

The MathWorks Inc., 2024. Robustfit. https://www.mathworks.com/help/stats/robustfit.html#mw_2b70e6c4-a7ef-48a6-9ccf-dfc7b15eb5b9.

van Vugt, M.K., Sederberg, P.B., Kahana, M.J., 2007. Comparison of spectral analysis methods for characterizing brain oscillations. Journal of Neuroscience Methods 162, 49–63.

Watrous, A.J., Fried, I., Ekstrom, A.D., 2011. Behavioral correlates of human hippocampal delta and theta oscillations during navigation. Journal of Neurophysiology 105, 1747–1755.

Whitten, T.A., Hughes, A.M., Dickson, C.T., Caplan, J.B., 2011. A better oscillation detection method robustly extracts EEG rhythms across brain state changes: The human alpha rhythm as a test case. NeuroImage 54, 860–874.

Wilson, L.E., da Silva Castanheira, J.S., Baillet, S., 2022. Time-resolved parameterization of aperiodic and periodic brain activity. eLife 11, e77348.

Appendix A.

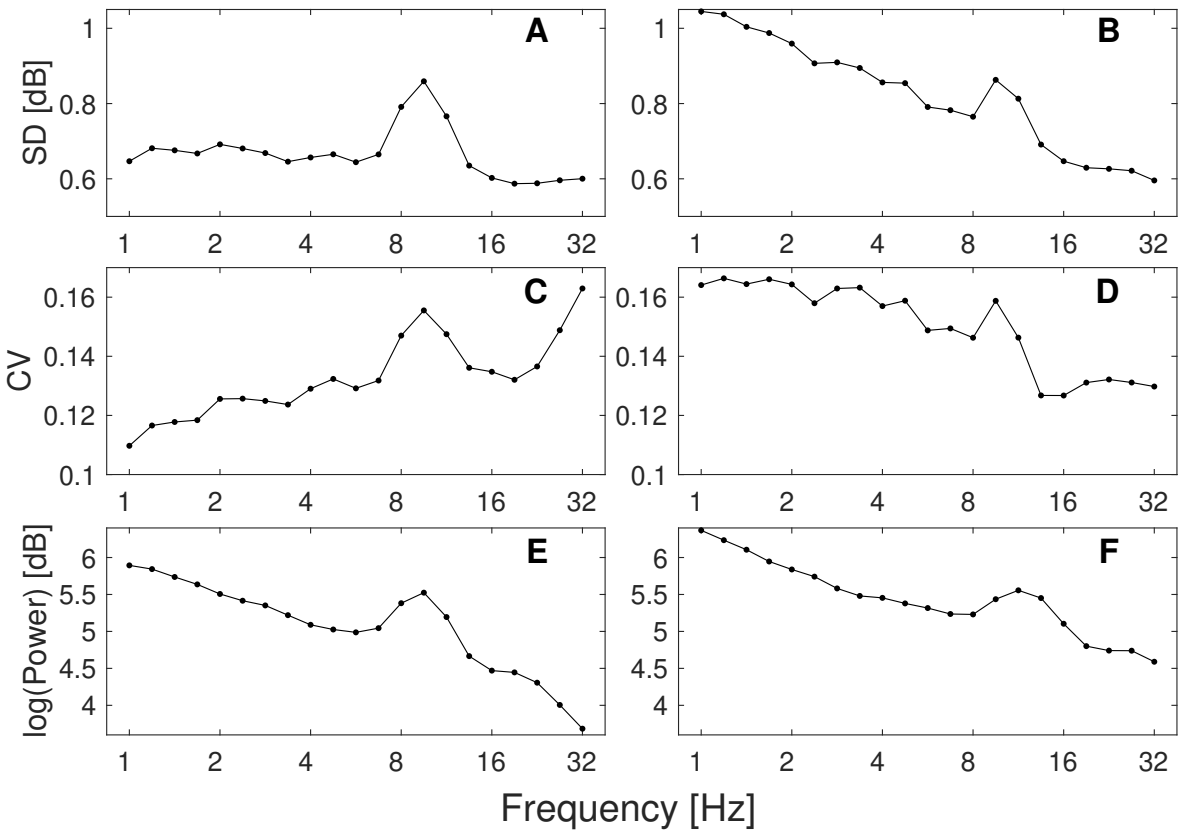

Figure A1: SD(f), CV(f) and respective power spectra from two empirical datasets (left and right column, respectively). SD(f) is shown by (A) and (B) and while the plot in (A) has the highest values at the peak frequencies for the respective power spectrum (E), this is not consistent with (B) and the respective power spectrum in (F). CV(f) is shown by (C) and (D) and neither plot has the highest or lowest values corresponding to the peak frequencies in either power spectrum.

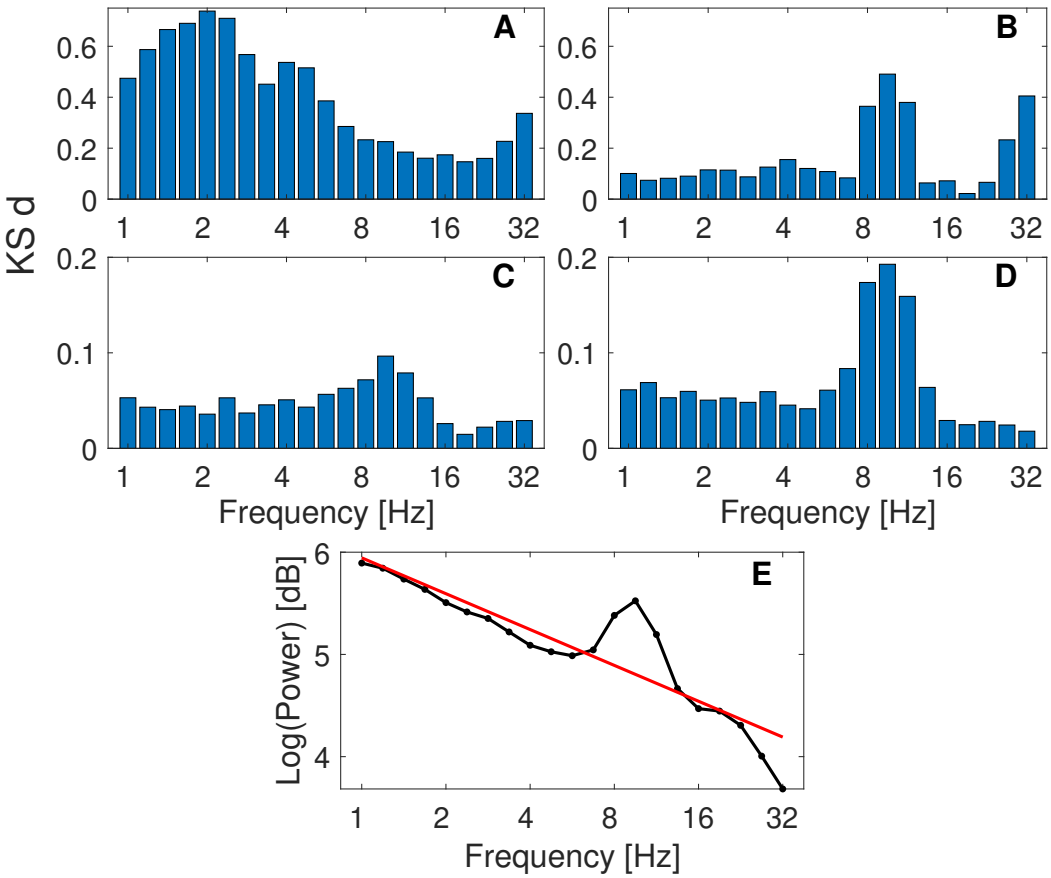

Figure A2: Various methods of calculating KS d vs frequency for the same dataset, shown by the power spectrum (E). (A) KS d calculated using the mean of power values at each frequency. (B) KS d calculated using regression values at each frequency. (C) KS d calculated with the method using regression values, a 99.9% threshold and subsample mean. (D) KS d calculated using the median of power values at each frequency.

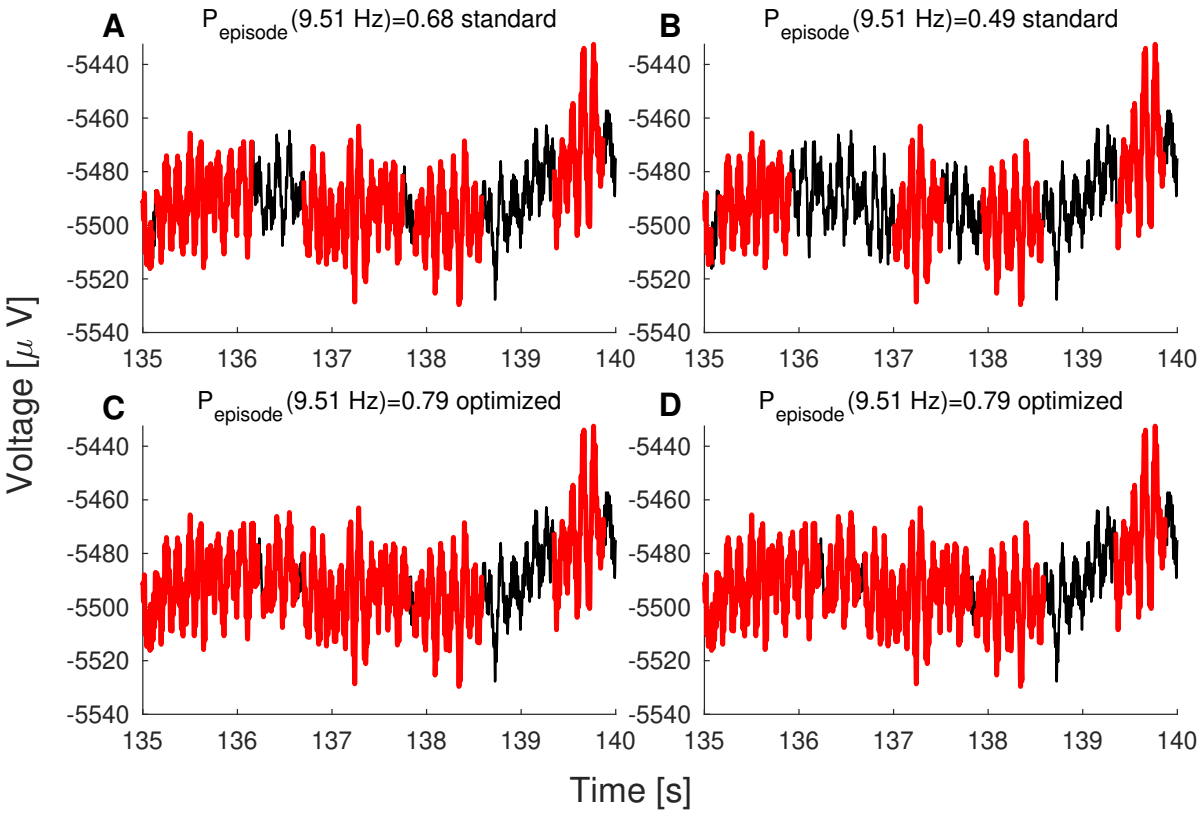

Figure A3: Zoomed in 5 s sections from Figure 10 (from 135 – 140 s in the signal). (A) corresponds to Figure 10C (standard BOSC with whole spectrum) while (B) comes from standard BOSC with the partial spectrum in Figure 10D. (C) and (D) do the same but with optimized BOSC (corresponding to Figure 10E and F respectively) and they are identical with no loss of detection while (A) and (B) show clear diffeences in detection. This zoomed-in section provides an example of how optimized BOSC can conserve detection while only using a partial spectrum that would lead to a loss of detections in standard BOSC.