Robust power spectral estimation for EEG data

Tamar Melman [a],[*], Jonathan D. Victor [b]

[a] Tri-Institutional Program in Computational Biology in Medicine, Weill Cornell Medical College, 1300 York Avenue, Box 194, New York, NY 10065, United States
[b] Feil Family Brain and Mind Research Institute and Department of Neurology, Weill Cornell Medical College, 1300 York Avenue, New York, NY 10021, United States

HIGHLIGHTS

- We present a method for power spectral estimation based on robust statistics.
- Compared to standard methods, the new approach is resistant to transient artifacts.
- Confidence intervals estimated in a Bayesian fashion have appropriate coverage.
- The approach is computationally efficient.
- Software is provided in the form of a MATLAB toolbox.

ARTICLE INFO

ABSTRACT

Background: Typical electroencephalogram (EEG) recordings often contain substantial artifact. These artifacts, often large and intermittent, can interfere with quantification of the EEG via its power spectrum. To reduce the impact of artifact, EEG records are typically cleaned by a preprocessing stage that removes individual segments or components of the recording. However, such preprocessing can introduce bias, discard available signal, and be labor-intensive. With this motivation, we present a method that uses robust statistics to reduce dependence on preprocessing by minimizing the effect of large intermittent outliers on the spectral estimates.
New method: Using the multitaper method (Thomson, 1982) as a starting point, we replaced the final step of the standard power spectrum calculation with a quantile-based estimator, and the Jackknife approach to confidence intervals with a Bayesian approach. The method is implemented in provided MATLAB modules, which extend the widely used Chronux toolbox.
Results: Using both simulated and human data, we show that in the presence of large intermittent outliers, the robust method produces improved estimates of the power spectrum, and that the Bayesian confidence intervals yield close-to-veridical coverage factors.
Comparison to existing method: The robust method, as compared to the standard method, is less affected by artifact: inclusion of outliers produces fewer changes in the shape of the power spectrum as well as in the coverage factor.
Conclusion: In the presence of large intermittent outliers, the robust method can reduce dependence on data preprocessing as compared to standard methods of spectral estimation.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Electroencephalography (EEG), a technique for recording the electrical activity of the brain via surface electrodes, is a commonly used assay of brain activity in research and clinical settings. Well-recognized advantages of the EEG include its high temporal resolution, noninvasive nature, and ease of use (Bunge and Kahn, 2009). However, it is also highly sensitive to electrical activity from non-neural sources, such as eye movements (Gasser et al., 1992),

muscle activity (Whitham et al., 2007), electrode movement, and electric fields from the environment (Tatum et al., 2011). These sources generate signals that corrupt the underlying neural signal, and are difficult, if not impossible, to avoid.

For many research applications, and increasingly for clinical applications (Schiff et al., 2014), spectral measures are used to analyze EEG characteristics (Mitra and Pesaran, 1999). Since activity in specific frequency bands often has direct biological interpretations (Penfield and Jasper, 1954), the power spectrum is of particular interest. However, since the raw EEG signal is contaminated by non-neural sources, obtaining reliable estimates of the power spectrum that reflect underlying brain activity is not straightforward.

Computation of the power spectrum typically involves segmenting the continuous signal, applying Fourier analysis to each segment, and calculating the mean over segments of the power at each frequency. The data segments, typically of duration 1 s or more, may be determined arbitrarily (e.g., for records of spontaneous EEG), or based on events in a behavioral paradigm (e.g., for event-related potential studies). Fourier components arising from segments contaminated by typical artifacts (e.g., muscle and eye movements) are typically large relative to those of segments that only contain the neural signal, and therefore bias the mean upwards. This problem is usually addressed by removing these artifacts, by a combination of manual identification of artifact-containing segments and automated means, such as independent component analysis (ICA) (Makeig et al., 1996); however this can be labor- and time-intensive, subjective, and can discard portions of usable data.

Here we describe an alternative approach to this outlier problem, via the use of robust statistics. Specifically, we focus on the median and other quantile-based statistics. Via simulations and application to real EEG data, we show that this approach can recover the power spectrum of the underlying signal even in the presence of substantial artifact. Finally, we provide code that extends the Chronux (Bokil et al., 2010; Mitra and Bokil, 2008), toolbox to carry out these computations, including the calculation of Bayesian confidence intervals.

2. Methods

2.1. Algorithm

2.1.1. Modified multitaper method

A power spectrum is typically estimated from a measured time series by cutting the time series into segments, applying Fourier analysis to these segments, and averaging the power in each frequency bin across segments. The true value of the power spectrum is the limit of this process as the length and number of the data segments tend to infinity. However, in practice these segments are finite in length and limited in number, so power spectral estimates are necessarily biased (resulting from spectral leakage due to the finite length of the data segment) and imprecise (due to the finite number of data segments).

The multitaper method (Prieto et al., 2007), a power-spectral estimator that we use as a starting point for our approach, tackles the tradeoff between this bias and variance in a way that is optimal for Gaussian signals. The method minimizes spectral leakage (the artifactual spreading of power from one frequency bin into its neighbors), by windowing each segment by an orthogonal set of functions, the Slepian tapers. For further background on the multitaper method see Thomson (1982), Mitra and Pesaran (1999) and Mitra and Bokil (2008). Chronux is a freely available MATLAB toolbox that provides convenient implementations of the multitaper method, which we extend with an implementation of the robust approach.

The standard multitaper method consists of the following steps: (1) multiplying each data segment by each of the tapers, (2) applying Fourier analysis to these products, (3) averaging over tapers within each segment, and (4) averaging over segments. To formalize this, we denote the original signal by $X(t)$, with B segments cut from the signal denoted as $x_1(t), \ldots, x_b(t), \ldots, x_B(t)$, each of length T. These segments are non-overlapping, but need not be contiguous. We denote the K Slepian tapers by $a_1(t), \ldots, a_k(t), \ldots, a_K(t)$ (the choice of K is driven by the desired spectral resolution and data length; a common choice for 3-s-long segments, and the Chronux default, is $K = 5$). With this notation, the standard multitaper estimate of $S_x(\omega)$, the true spectral power at frequency ω, is defined as:

$$\hat{S}^{\text{standard}}(\omega) = \frac{1}{B} \sum_{b=1}^{B} \sum_{k=1}^{K} \frac{1}{T} \left| \int_0^T x_b(t) a_k(t) e^{-i\omega t} dt \right|^2. \tag{1}$$

We denote the power estimate for a single sample b and a single taper by $S_{b,k}(\omega)$:

$$S_{b,k}(\omega) = \frac{1}{T} \left| \int_0^T x_b(t) a_k(t) e^{-i\omega t} dt \right|^2. \tag{2}$$

With this notation, the standard spectral estimate takes the form

$$\hat{S}^{\text{standard}}(\omega) = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{K} \sum_{k=1}^{K} S_{b,k}(\omega). \tag{3}$$

Thus, the standard multitaper estimate is a nested mean: first a mean over the K tapers within each segment to obtain the estimate $\hat{S}_b(\omega) = \text{mean}\left(\left\{\hat{S}_{b,k}(\omega)\right\}\right)$, and then a mean over the B segments:

$$\hat{S}^{\text{standard}}(\omega) = \text{mean}\left(\left\{\hat{S}_b(\omega)\right\}\right). \tag{4}$$

Since our goal is to reduce the effect of outlier estimates from each segment, we replace the mean over segments by a robust estimator, resulting in the estimated power spectral quantity $\hat{S}^{\text{robust}}(\omega)$. There are many possible choices for the robust estimator—for example: an estimator based on the hth quantile, a trimmed mean, a Winsorized mean (Huber, 1963), or iterative rejection of outliers. While the present framework applies to all of these choices, estimators based on quantiles are more amenable to computation of Bayesian confidence intervals (see below), and we therefore focus on these, both in the illustrations below and in the MATLAB toolbox. We denote the estimator based on the hth quantile as $\hat{S}^{\text{quantile}\_h}(\omega)$. Note that $h = 1/2$ corresponds to the median; this is the default value in the code.

Even for Gaussian data, the median power of the tapered estimates does not equal the mean power. This is because spectral estimates are approximately distributed as chi-squared, which is positively skewed. As shown in Appendix A, we can take the skewing into account by dividing the median power by a data-independent scale factor. Furthermore, scale factors can be derived that convert not just the median (0.5 quantile), but any quantile, into mean power. Appendix A details the calculation of these scale factors, which is implemented in the MATLAB module analytical_scalefactor_Robust().

Including this scale factor yields our main result, the robust spectral estimate:

$$\hat{S}^{\text{quantile}\_h}(\omega) = \frac{\text{quantile}\_h\left(\left\{\hat{S}_b(\omega)\right\}\right)}{C(h, d, B)}, \tag{5}$$

where $C(h, d, B)$ is the scale factor for quantile h; d is the number of degrees of freedom ($d = 2K$ for typical frequencies, $d = K$ for DC and the Nyquist frequency); and B, as above, is the number of segments.

Notably, the quantile is applied to the B power estimates from each segment (replacing the outer operation in Eq. (3)); within segments, the step of computing the mean over the tapers remains unchanged from the original method. There are two reasons for this choice: (1) if artifact is present in a segment b, it is likely to affect many of the tapered estimates from that segment; and (2) the K Slepian tapers were designed to be used together to capture all of the power within a frequency bin. The toolbox supports the alternative strategies of computing either the one-tiered median across all tapered estimates or the "two-tiered" median across tapers and then across segments, but as we see no principled reason for this, it is not the default. The robust approach is also applicable to other spectral estimation methods, such as Welch windowing.

2.1.2. Confidence interval estimation

Standard nonparametric approaches to confidence interval estimation (Thomson, 2007) are based on resampling strategies, such as the jackknife or the bootstrap. These approaches are appropriate for the mean, which depends smoothly on the data—a necessary condition for the jackknife or bootstrap to be valid. However, since quantile-based estimates do not depend smoothly on the data, an alternative approach is needed.

Our approach is as follows. Let $Q(\omega, h)$ denotes the true value of the hth quantile of estimates at frequency ω. We seek the probability distribution $P(Q(\omega, h)|\text{data})$: the distribution of the true value of the hth quantile, given the observed data. To find this, we use a Bayesian approach with the conservative choice of an uninformative (flat) prior for the power spectral value. Using Bayes' theorem, we can express $P(Q(\omega, h)|\text{data})$ in terms of the probability of drawing the data from a distribution with known h^{th} quantile, or $P(\text{data}|Q(\omega, h))$:

$$P(Q(\omega, h)|\text{data}) = P(\text{data}|Q(\omega, h)) \times \frac{P(Q(\omega, h))}{P(\text{data})}$$

$$\propto P(\text{data}|Q(\omega, h)). \tag{6}$$

To be consistent with our prior reasoning for implementing a two-tiered approach, we implement the Bayesian approach to confidence intervals by considering the data to be the set of spectral estimates $\hat{S}_b(\omega)$ derived from each segment (taking the mean of the tapered estimates within each segment, Eq. (4)).

We then use order statistics to compute $P(\text{data}|Q(\omega, h))$. Specifically, we re-label each $\hat{S}_b(\omega)$ as $Y_1, \ldots, Y_B$, where Y_1 is the smallest ranked value in $\{\hat{S}_b(\omega)\}$. We also denote $Y_0 = -\infty$ and $Y_{B+1} = +\infty$, as this will allow us to account for the possibility of $Q(\omega, h)$ lying below the smallest ranked value or above the largest. The probability that $Q(\omega, h)$ lies between Y_i and Y_{i+1} is equal to the probability that exactly i of the $\hat{S}_b(\omega)$ estimates are below this quantile, and $B - i$ are above it. Since the chance of any single estimate lying below the hth quantile of the estimates is exactly h, this probability is determined by the binomial distribution:

$$P(Y_i < Q(\omega, h) < Y_{i+1}) = \binom{B}{i} h^i (1-h)^{B-i}. \tag{7}$$

Thus, to ensure that the probability that $Q(\omega, h)$ lies between two ordered values, Y_l and Y_m (where $m > l$) is at least $1 - \alpha$, we need to find indices l and m for which

$$\sum_{i=l}^{m-1} P(Y_i < Q(\omega, h) < Y_{i+1}) \geq 1 - \alpha. \tag{8}$$

We choose the intervals in descending order of probability to determine the smallest number of intervals in Eq. (8) for a given coverage $1 - \alpha$. The union of these intervals is the desired confidence interval. Fig. 1 illustrates this procedure for $\alpha = 0.05$ (i.e., 95% confidence intervals).

Note that the confidence interval provided by the above procedure for a coverage $1 - \alpha$ typically also applies to coverage factors somewhat larger than $1 - \alpha$. This is because the upper and lower confidence intervals are tethered to discrete values (the observations Y_b), so the confidence bounds that satisfy Eq. (8) typically also satisfy it for smaller values of α as well. The relationship of the predicted coverage factor to the number of samples is shown in Fig. 2A (for $\alpha = 0.05$ and $h = 0.5$). Note also that if the number of segments is sufficiently small, then it may be necessary to include the intervals $(-\infty, Y_1]$ and $[Y_B, \infty)$ in order to satisfy Eq. (8). Fig. 2B shows the minimum number of segments required to have finite confidence intervals, as a function of h.

For comparison purposes, we also calculated confidence intervals via the jackknife procedure, which is the standard Chronux approach. Specifically, the jackknife confidence interval is computed by pooling together all estimates from all tapered segments (for a total of BK estimates) and generating BK subsets of size $BK - 1$ by dropping one tapered segment from each. The standard or robust estimator of central tendency (mean for the standard method; quantile for robust) is then applied to each subset. The standard deviation is calculated and significance is determined according to the t-distribution with $BK - 1$ degrees of freedom. The provided Chronux extension includes jackknife-based confidence intervals as well as bootstrap-based confidence intervals for the robust estimators; however these are not intended for routine use, only for purposes of comparison with the Bayesian confidence intervals.

Finally, we note that the above approach to confidence limit estimation is distinct from the naïve strategy of choosing the $\alpha/2$ and $1 - \alpha/2$ quantiles of the spectral estimates $\hat{S}_b(\omega)$, rather than the ranked values whose indices are identified by Eq. (8). With the naïve strategy, confidence intervals do not narrow as the amount of data increases. Moreover, the impact of outliers is not removed if the outlier fraction in either tail is at least $\alpha/2$.

2.2. Method validation

We applied the above procedures to (1) a synthetic signal of a known power spectral distribution corrupted by artifact, and (2) an EEG record from a human subject, with the typical artifacts of clinical recordings.

2.2.1. Simulated data

The simulated data consisted of a Gaussian signal of known power spectral distribution, to which we added a controlled amount of simulated artifact. The signal was synthesized from random-phase, Gaussian-distributed Fourier components, whose mean power was proportional to $1/\omega$ (over the range 1/3 to 100 Hz in steps of 1/3 Hz), where ω is the frequency. That is, at each frequency ω in the above range, we set the Fourier component of the signal, $\tilde{x}(\omega)$, to have real and imaginary values each independently drawn from a normal distribution of variance equal to $1/\omega$: $\text{Re}(\tilde{x}(\omega)) \sim N(0, \omega^{-1/2})$ and $\text{Im}(\tilde{x}(\omega)) \sim N(0, \omega^{-1/2})$. We then inverted the transform to create the time-domain signal $x(t)$ in that segment. Artifacts, which were added in the time domain, consisted of bursts of Gaussian signal with a flat (constant) power spectrum. The artifact power per unit bandwidth was 3.3 times greater than the signal power per unit bandwidth at the lowest frequency, 1/3 Hz—so the maximum signal:noise was 1/3.3. The burst length was 0.5 s (while samples were simulated to be 3 s in duration), and bursts were added to the signal $x(t)$ at Poisson-distributed intervals, with an average of 0.25 bursts per segment. We studied the performance of the two methods for data with, and without noise added. Sample data, both with and without artifact, are shown in Fig. 3. Since the power spectrum of the underlying
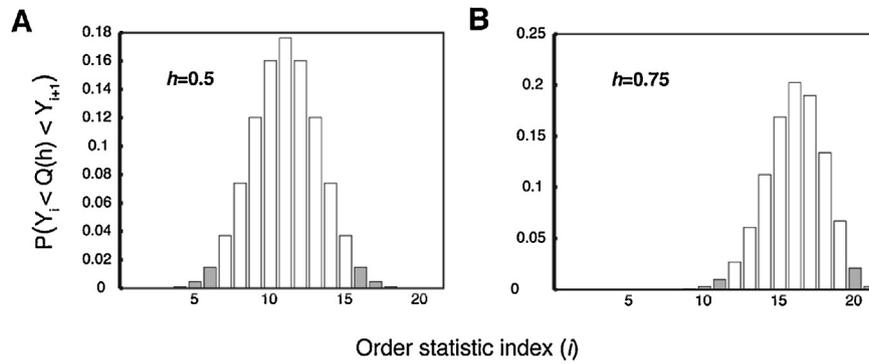
Fig. 1. Procedure for determination of confidence intervals based on order statistics. The bar graphs show the distribution of probabilities of the true h^{th} quantile falling between the ith and $(i+1)^{th}$ order statistic for a set of $B = 20$ values. The first and last bars ($i = 0$ and $i = B$) indicate the probability of the true quantile value falling in the intervals $(-\infty, Y_1]$ and $[Y_B, \infty,)$, respectively, where $Y_1, \ldots, Y_B$ are the order statistics corresponding to the B spectral estimates. The white bars indicate inter-order-statistic intervals whose probabilities sum up to $1 - \alpha$, representing the region between the $1 - \alpha$ confidence intervals.
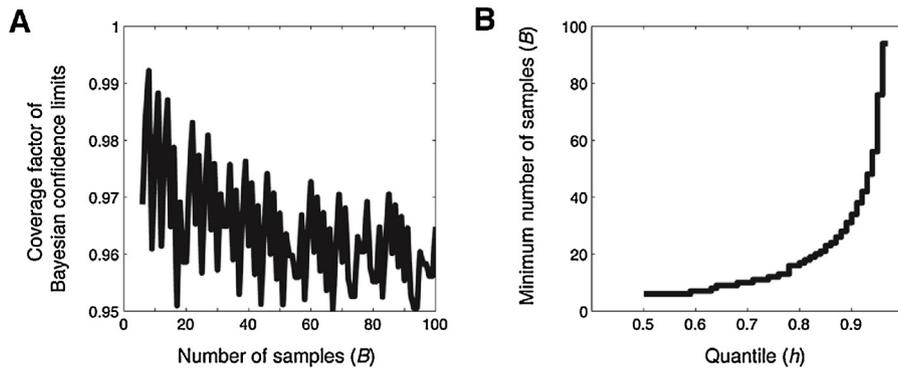
Fig. 2. (A) The nominal coverage factor of a Bayesian confidence interval exceeds $1 - \alpha$, and this excess depends on the number of samples. Here this relationship is shown for the median ($h = 0.5$) estimator. (B) The minimum number of samples necessary in order to obtain finite 95% confidence intervals for a range of quantile values.
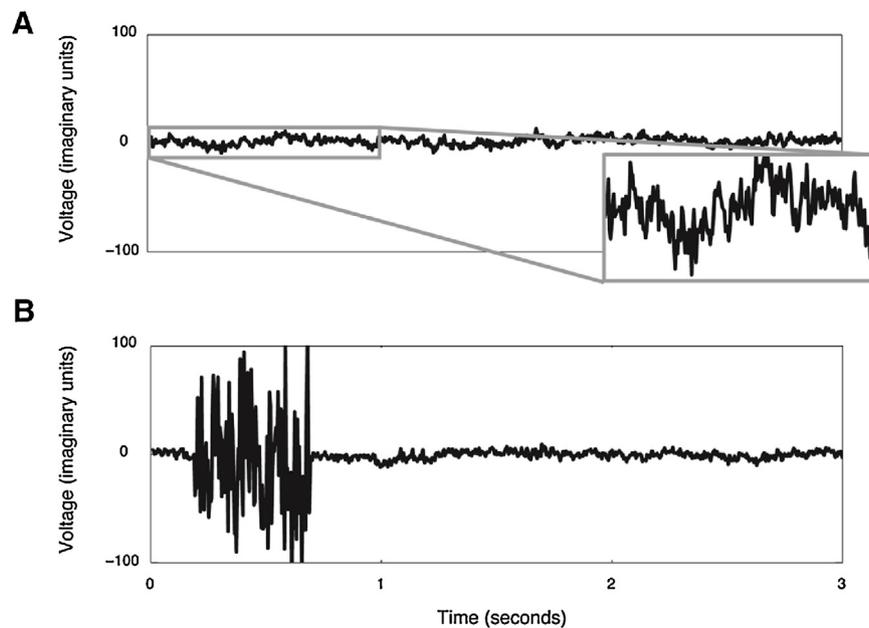
Fig. 3. Representative simulated signal segments. (A) Three seconds of a simulated signal with a $1/\omega$ power spectrum. Inset enlarges a portion of the trace. (B) Three seconds of simulated artifact-containing signal, containing a white noise burst of 0.5-s duration.

signal was $1/\omega$, signal:noise decreases further with increasing frequency, from its maximum of 0.3 (at 1/3 Hz). Thus, we anticipate that at sufficiently high frequencies, the noise bursts will lead to unacceptable corruption of the spectral estimates.

2.2.2. Human data

The human EEG data were obtained from a healthy control subject (23 year old male). Data were recorded with an FS128 headbox and an XLTEK acquisition system (Natus Medical, Pleasanton, CA
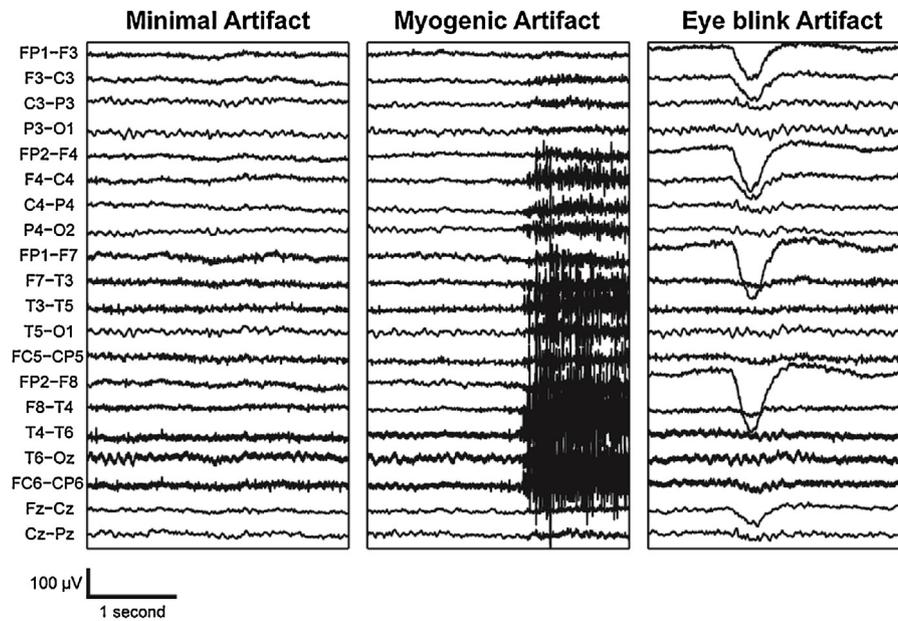
Fig. 4. Typical segments of recorded human EEG. The first panel shows an example of a minimal-artifact segment, taken to consist predominantly of neural signal. The second and third panels show examples of segments containing significant artifact: myogenic and eye blink artifacts.

94566) using an augmented 10/20 montage. The sampling frequency was 250 Hz. Input impedances were $\leq$5 kΩ. The subject was awake during testing, and generated spontaneous movement artifacts and eye blinks. Low-pass and high-pass filters of 0.01 and 100 Hz, respectively, and a 60 Hz notch filter, were applied. For analysis, 40 segments (either 1 or 3 s in duration) were selected from a recording of 8 min: 20 of these segments, labeled "significant artifact," contained EMG, eye-blinks, or other artifacts, as determined by visual inspection of the EEG and simultaneously-recorded video (carried out by an experienced EEG analyst). 20 other segments were labeled "minimal artifact," as they contained little or no visible artifact. Fig. 4 shows one example of a "minimal artifact" segment and two examples of "significant artifact" segments. To generate datasets with 25% artifact, we drew 5 significant-artifact segments and 15 minimal-artifact segments randomly from these two subgroups.

Human subject participation was approved by the Institutional Review Board, and was consistent with the Declaration of Helsinki.

3. Results

We first compare the standard multitaper method and the robust method for simulated signals. Because the true underlying spectrum is known, this allows rigorous assessment of accuracy and the coverage factors of the estimated confidence intervals. We then apply the standard and robust methods to a sample of human EEG, and show that the robust method is less sensitive to typical EEG artifact encountered in clinical recordings.

3.1. Simulated EEG results

Fig. 5 compares power spectral estimates via the standard and robust multitaper methods on a simulated EEG signal. Each method was applied to two data sets that differed in the average number of noise-contaminated segments per data set: a clean dataset, and a dataset with an average of 25% artifact-containing segments, respectively (see Section 2 for details). For the standard method, when artifacts were present the expected high-frequency decline of the power spectrum is corrupted by the flat spectrum of the noise bursts. This shows that the artifact significantly affects the

estimated spectrum. In comparison, results from the robust method shown in the right panel reveal that even over frequencies at which the power spectrum of the standard estimate is dominated by the noise bursts, the robust estimate reflects the underlying signal's spectrum. In sum, in a data set where outliers significantly affect spectral estimates from the standard method, the estimate from the robust method can capture the underlying spectrum.

The simulated data allowed for an assessment of confidence interval estimation methods (Fig. 6), and for a comparison of the Bayesian confidence intervals to the confidence intervals computed by the standard Chronux approach. The first column of the figure shows results for the standard approach, i.e., the standard multitaper estimates with jackknife confidence intervals. As expected, when no noise is present (top panel), the spectral estimates are close to the true value, and confidence interval coverage is approximately 95%. When noise is added, the spectrum is upwardly biased by the noise, and the coverage factors drop. The third column of the figure shows that for the robust method, the confidence interval coverage determined by the Bayesian procedure described in Section 2 remains at approximately 95%, even when noise is added.

The middle column shows that the jackknife approach fails when there are outliers, even when the robust method is used. Specifically, when a jackknife procedure is applied to the robust method, the confidence intervals have a lower-than-veridical coverage factor, and they are also more irregular than the Bayesian confidence intervals. As mentioned in Section 2, the failure of confidence intervals based on resampling is not surprising, since the median depends in a non-smooth fashion on the data. However, even under these circumstances – and when artifacts are not explicitly removed – the Bayesian confidence intervals are approximately veridical.

3.2. Human EEG results

Fig. 7 shows that the robust method can be successfully applied to human EEG data. Fig. 7A compares spectra determined by standard analysis of minimal-artifact (hand-cleaned) EEG segments with a set of EEG segments for which 25% contained significant artifact. As expected, there were substantial deviations of the spectra obtained when significant artifact was present. These included
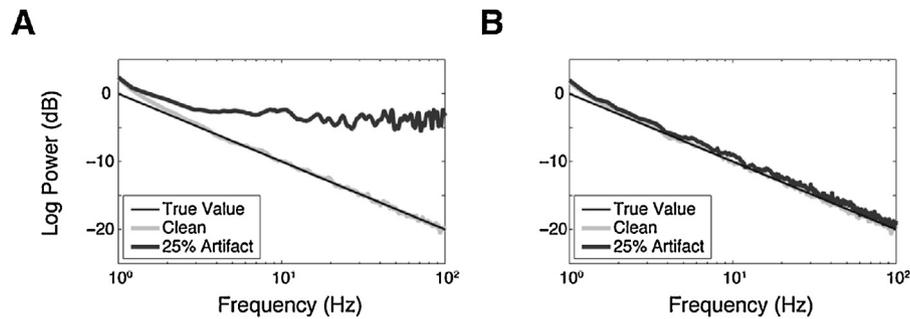
Fig. 5. Power spectral estimates (left: standard method, right: robust method) for simulated datasets with 0% and 25% artifact-containing segments, respectively. The thin sloping black line is the true value of the power spectrum ($1/\omega$). Spectra were estimated with 5 tapers and a time-bandwidth product of 3.
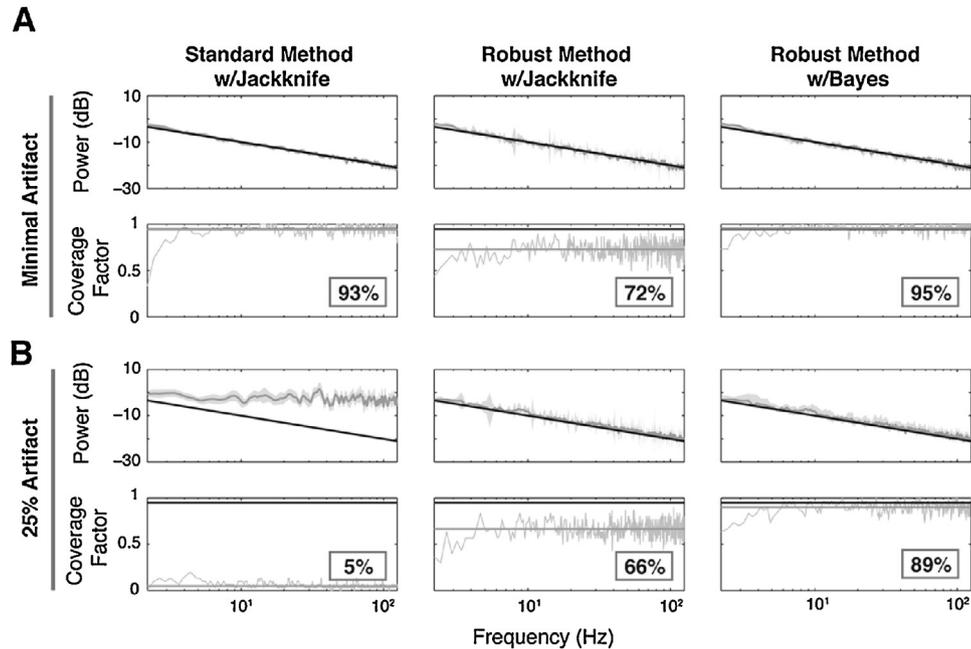
Fig. 6. Coverage factors of confidence intervals obtained via several methods, for simulated data. Simulations as in Fig. 3: spectra are estimated for data sets of 20 samples with 5 tapers and a time-bandwidth product of 3. (A) Minimal-artifact data; (B) 25% of segments on average contaminated by significant artifact. Left column is the standard multitaper estimate of power spectrum, with jackknife confidence intervals. Right column is the robust multitaper estimate of power spectrum with Bayesian confidence intervals. Middle column is the jackknife with the robust method for comparison (see Section 2 for details). In the power spectral plots, the thin black line represents the $1/\omega$ signal spectrum, and the gray line with light gray error bars represents the method's estimate of the power spectrum. In the plots of coverage factor vs. frequency, the thin gray line represents the empirical coverage factor at each frequency, averaged over 30 simulations; the thick gray line represents the mean empirical coverage for each method (also indicated by the number displayed on each graph). Note that when significant artifact is present, the robust method with Bayesian confidence intervals is the only one for which the empirical coverage factor approaches the nominal coverage of 95% (thick horizontal black line).

deviations at low frequencies in frontal channels, presumably due to eye movement artifact, and deviations and at high frequencies in frontal and posterior channels, presumably due to myogenic artifact. Fig. 7B shows that these deviations, especially at high frequencies, are largely eliminated when the robust method is used.

Fig. 7C shows that on cleaned data the robust method gives results that are very close to that of the standard method. Thus, the resistance of the robust method to corruption by artifact is not at the expense of a distortion of the result. Fig. 7D demonstrates these same findings when the data are cut into 1-s segments.

4. Discussion

Above we have shown that a simple modification of the standard approach to spectral estimation – substituting a robust estimator for the mean across segments – can substantially improve spectral estimates of EEG signals in the presence of artifact. The basic rationale is that robust estimators are insensitive to outliers, and many

sources of artifact behave as outliers. When combined with the multitaper method (Thomson, 1982), key advantages of the latter are retained: spectral leakage is minimized, and reliable confidence intervals can be estimated.

The proposed data-driven approach reduces the reliance on removal of artifact by other means. This has several advantages: with less preprocessing, fewer data will be discarded, potentially enabling the capture of subtle EEG dynamics. A reduced reliance on preprocessing methods also has the benefit of reducing the dependence on ad hoc or subjective methods of artifact identification, and may also accelerate the data-processing pipeline.

It is worth noting that the robust methods described here are computationally efficient. Since the MATLAB implementations of median() and quantile() (used for the power spectrum) and sort() (used for Bayesian confidence intervals) have approximately linear runtime even for 10^8 segments, the robust method retains the linear asymptotic runtime of the standard multitaper approach.
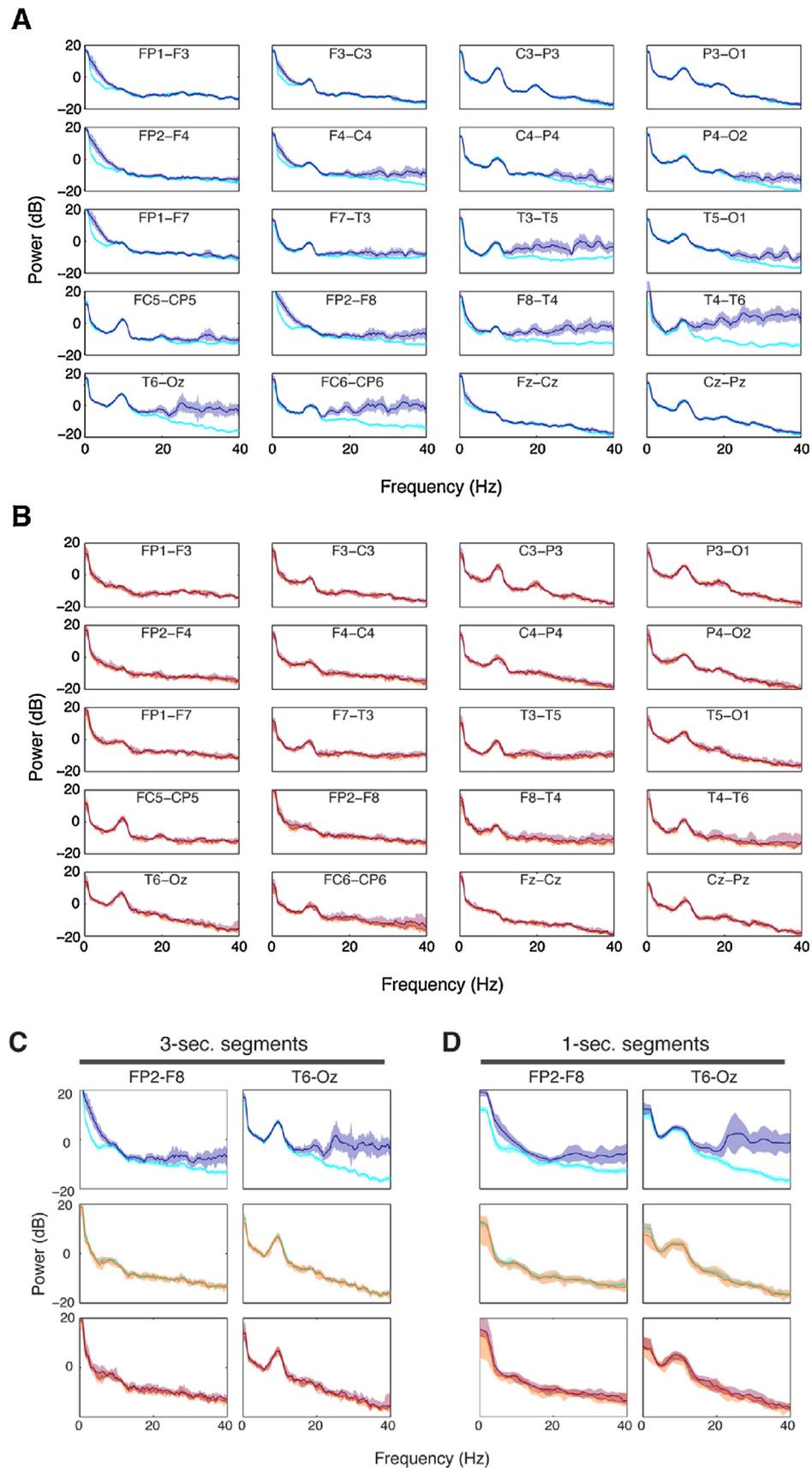
Fig. 7. Power spectral estimates ((A) standard method, (B) robust method) for 20 channels of EEG from a human subject. Spectra were calculated on data sets consisting of 20 data segments of 3-s duration, using 5 tapers and a time-bandwidth product of 3. Light color indicates analyses of minimal-artifact segments; dark color indicates analyses of a mixture of 75% minimal-artifact segments and 25% significant-artifact segments. Blue indicates standard spectral estimates; red indicates robust spectral estimates. Color bands indicate 95% confidence intervals. (C) Detailed comparison at channels FP2-F8 and T6-Oz using 3-s-segments. Color scheme same as in (A) and (B). (D) Same as (C) but with 1-s segments. Note that for minimal-artifact data (middle row of (C) and (D)), the two methods give similar results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Caveats

Because the robust method works by separating signals that are pervasive (e.g., those reflecting background state) from those that are infrequent, it may discard an intermittent signal of neural origin – such as paroxysmal activity – as artifact. The robust method works specifically because it removes outliers. Outliers include many sources of artifact but can also include neural-origin EEG activity that is infrequent.

Although this method greatly improves spectral estimates for certain data sets, it should not be treated as a panacea for all analysis-limiting noise. Since quantile estimators have a breakdown point of ≤50%, this method may not show any improvement over the standard analysis pipeline for constant or frequent noise that affects the same frequency range in most or all segments, such as 60 Hz line noise from the environment or frequent muscle tics. In these cases, alternatives such as notch filtering or artifact removal by hand must be used. We also note that the proposed approach will not remove pervasive low-level EMG, which can bias spectral estimates in the gamma range (Whitham et al., 2007).

For clarity we tested the method here in the absence of other artifact removal techniques. However, there are benefits to other techniques for removing outliers, such as ICA, automated outlier rejection, or hand-cleaning guided by video assessment of the subject's movements. Combining artifact removal techniques with the robust method may be more effective than either approach on its own.

4.2. Epoched data

While we have illustrated the method for a continuous recording of spontaneous EEG, it is also applicable to an event-related-potential (ERP) experiment. In this scenario, the data segments can be determined by the epochs or event markers themselves, and the spectra calculated as described above. However, we note that the scale factor calculation (Appendix A) assumes that the spectral estimates are distributed like chi-squared, which in turn means that there is no time-locked component, as must be the case for spontaneous EEG. If, in an ERP experiment, a time-locked component is significant because of events synchronized to the epoch boundaries, caution should be exercised: (a) the power spectrum is no longer rigorously defined, and (b) the scale factor will be frequency-dependent, and biased toward 1.

4.3. Extension to multichannel analysis

The utility of robust estimators as applied to the EEG spectrum suggests that robust methods will also be useful in the multichannel domain. In this context, robust estimators of shape, such as the minimum volume ellipse (MVE) (Rousseeuw and Leroy, 1987) or the minimum-covariance determinant (MCD) (Hubert and Debruyne, 2010), could be used to estimate cross-spectra, much as the quantile-based estimators here characterize power. While this specific robust approach appears to be as yet unexplored, a previous study has shown that using the median instead of the mean improves multi-taper coherence estimates (Wong et al., 2011). A full-fledged robust estimator of shape could provide phase information as well.

Acknowledgments
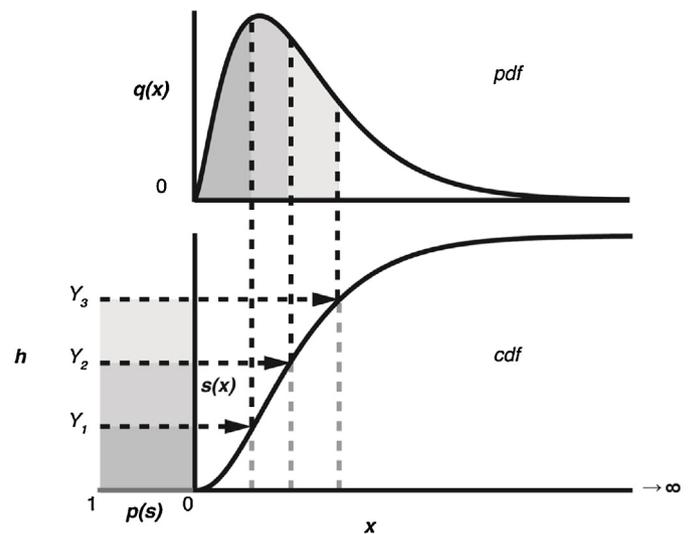
Fig. 8. Determining the order statistics of samples drawn from a non-uniform distribution whose PDF is given by $q(x)$. First, $s(x)$, the CDF associated with $q(x)$, is used to map samples drawn from a uniform distribution on the [0, 1] interval into samples drawn from $q(x)$. The function $s(x)$ is necessarily monotonic, so this mapping does not change the rank order of samples. Order statistics for the uniform distribution, Y_i, are then mapped via $s(x)$ into the domain for $q(x)$.

Appendix A. Scale factor

As mentioned in the main text, the raw quantile of a set of spectral estimates is expected to be proportional to the power, but not equal to it. The proportionality constant is dependent on the quantile h, the number of degrees of freedom d of the underlying χ^2 distribution describing the expected distribution of spectral estimates for Gaussian data, and the number of samples (segments), B. Here we derive this scale factor $C(h, d, B)$ by determining the hth quantile of the tapered spectral estimates, $\hat{S}_b(\omega)$, for a Gaussian signal of unit power.

For most frequencies ω, with the exception of $\omega = 0$ or the Nyquist frequency, Fourier estimates are complex numbers. When ω is greater than the bandwidth of the tapers, the real and imaginary components are approximately independent and of equal variance (Percival and Walden, 1993, p. 360), so for $\{\hat{S}_b(\omega)\}$ the power is distributed as the sum of $2K$ squares of Gaussian-distributed quantities, where K is the number of tapers. For $\omega = 0$ or the Nyquist frequency the Fourier estimates are real, so the power of the $\{\hat{S}_b(\omega)\}$ is distributed as sum of K such quantities. Therefore $\{\hat{S}_b(\omega)\}$ at a particular frequency ω is distributed as χ^2/d where χ^2 has d degrees of freedom, and $d = K$ or $d = 2K$. The proportionality between the quantiles of this distribution and the mean is the factor that converts the quantile estimate into an estimate of power.

To compute the expected value of a quantile, we use the strategy shown in Fig. 8. We first find a monotonic transformation from the uniform distribution on the interval [0,1] into the chi-squared distribution of spectral estimates. Because the transformation is monotonic, the rank-order of the samples drawn from the chi-squared distribution corresponds to the rank-order of the corresponding values in the uniform [0,1] distribution. Therefore, we can take the expected distribution of the hth quantile in the uniform distribution, and transform it back into the chi-squared distribution to determine $C(h, d, B)$.

To determine this transformation, we note that for an arbitrary distribution with probability density function (PDF) $q(x)$, the cumulative distribution function (CDF) $s(x)$ is given by

$$s(x) = \int_{-\infty}^{x} q(z)\, dz.$$

This can be rewritten as

$$\frac{ds(x)}{dx} = q(x).$$

The cumulative distribution function, by definition, is uniformly distributed between 0 and 1. With $q(x) = d\chi_d^2(x)$, then $x = (1/d)\text{chi2inv}(s, d)$ is the desired transformation between a uniformly-distributed quantity, s, and the spectral estimates, where $\text{chi2inv}(s, d)$ is the inverse cumulative chi-squared probability density function at s with d degrees of freedom.

We now apply order statistics to the variable s, which is uniformly distributed on $[0\ 1]$. The distribution of the $(k + 1)$th-ranked value for $N + 1 = B$ draws from the uniform distribution is given by

$$p_{k,N}(s) = \frac{1}{\text{beta}(k + 1, N - k + 1)} s^k (1 - s)^{N-k},$$

where $\text{beta}(u, v)$ is the beta-function,

$$\text{beta}(u, v) = \int_0^1 t^{u-1}(1 - t)^{v-1}\, dt$$

(David and Nagaraja, 2003).

Transforming back to the distribution of spectral estimates, we find:

$$q_{k,N}(x)\, dx = p_{k,N}(s)\, ds,$$

where

$$q_{k,N}(x) = p_{k,N}(s)\frac{ds}{dx} = \frac{1}{\text{beta}(k + 1, N - k + 1)} s^k (1 - s)^{N-k}\frac{ds}{dx}.$$

The expected value of this quantity is therefore:

$$\langle x \rangle = \int_0^\infty x(s)\, q_{k,N}(x)\, dx$$

$$= \frac{1}{\text{beta}(k + 1, N - k + 1)} \int_0^\infty x s^k (1 - s)^{N-k}\left(\frac{ds}{dx}\right) dx$$

$$= \frac{1}{\text{beta}(k + 1, N - k + 1)} \int_0^1 \frac{1}{d}\text{chi2inv}(s, d) s^k (1 - s)^{N-k}\, ds$$

When the quantile h falls exactly on a sample, i.e., when $h = (k + 1)/(N + 1) = (k + 1)/B$, this is $C(h, d, B)$. When the quantile h falls between two samples, $C(h, d, B)$ is determined by interpolating this value between two adjacent values of k. By default, the code uses the MATLAB convention for quantile interpolation, i.e., a weighted average of the values at the two adjacent values of k.

As $B = N + 1$ increases, the above result takes a simple asymptotic form, since the integrand factor $s^k(1 - s)^{N-k}$ becomes concentrated at $s = (k + 1)/N + 1 = h$. In this limit,

$$\lim_{B \to \infty} C(h, d, B) = \frac{1}{d} \times \text{chi2inv}(h, d).$$

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2016.04.015.

References

Bokil, H., Andrews, P., Kulkami, J.E., Mehta, S., Mitra, P.P., 2010. Chronux: a platform for analyzing neural signals. J. Neurosci. Methods 192 (September (1)), 146–151.

Bunge, S., Kahn, I., 2009. Cognition: an overview of neuroimaging techniques. Encycl. Neurosci. 2, 1063–1067.

David, H.A., Nagaraja, H.N., 2003. Order Statistics, 3rd ed. Wiley.

Gasser, T., Ziegler, P., Gattaz, W.F., 1992. The deleterious effect of ocular artefacts on the quantitative EEG, and a remedy. Eur. Arch. Psychiatry Clin. Neurosci. 241 (July (6)), 352–356.

Huber, P.J., 1963. Robust estimation of a location parameter. Ann. Math. Stat. 35 (1), 73–101.

Hubert, M., Debruyne, M., 2010. Minimum covariance determinant. Wiley Interdiscip. Rev. Comput. Stat. 2 (January (1)), 36–43.

Makeig, S., Bell, A.J., Jung, T.-P., Sejnowski, T.J., 1996. Independent component analysis of electroencephalographic data. Adv. Neural Inf. Process. Syst., 145–151.

Mitra, P.P., Bokil, H., 2008. Observed Brain Dynamics. Oxford University Press, New York, NY.

Mitra, P.P., Pesaran, B., 1999. Analysis of dynamic brain imaging data. Biophys. J. 76 (2), 691–708.

Penfield, W., Jasper, H., 1954. Epilepsy and the Functional Anatomy of the Human Brain. Little, Brown and Company, Boston, MA.

Percival, D.B., Walden, A.T., 1993. Spectral Analysis for Physical Applications, 1st ed. Cambridge University Press, Cambridge, UK.

Prieto, G.A., Parker, R.L., Thomson, D.J., Vernon, F.L., Graham, R.L., 2007. Reducing the bias of multitaper spectrum estimates. Geophys. J. Int. 171 (October (3)), 1269–1281.

Rousseeuw, P.J., Leroy, A.M., 1987. Robust Regression and Outlier Detection. Wiley.

Schiff, N.D., Nauvel, T., Victor, J.D., 2014. Large-scale brain dynamics in disorders of consciousness. Curr. Opin. Neurobiol. 25 (April), 7–14.

Tatum, W.O., Dworetzki, B.A., Schomer, D.L., 2011. Artifact and recording concepts in EEG. J. Clin. Neurophysiol. 28 (June (3)), 252–263.

Thomson, D.J., 1982. Spectrum estimation and harmonic analysis. Proc. IEEE 70 (9), 1055–1096.

Thomson, D.J., 2007. Jackknifing multitaper spectrum estimates. Signal Process. Mag. IEEE 24 (4), 20–30.

Whitham, E.M., Pope, K.J., Fitzgibbon, S.P., Lewis, T., Clark, C.R., Loveless, S., Broberg, M., Wallace, A., DeLosAngeles, D., Lillie, P., Hardy, A., Fronsko, R., Pulbrook, A., Willoughby, J.O., 2007. Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. Clin. Neurophysiol. 118 (August (8)), 1877–1888.

Wong, K.F.K., Mukamel, E.A., Salazar, A.F., Pierce, E.T., Harrell, P.G., Walsh, J.L., Sampson, A., Brown, E.N., Purdon, P.L., 2011. Robust time varying multivariate coherence estimation: application to electroencephalogram recordings during general anesthesia. Conf. Proc. IEEE Eng. Mol. Biol. Soc., 4725–4728.