A WELCH METHOD APPROXIMATION OF THE THOMSON MULTITAPER SPECTRUM ESTIMATOR

Maria Hansson-Sandsten

Mathematical Statistics,
Centre for Math. Sciences, Lund University,
SE-221 00 Lund, Sweden,
email: sandsten@maths.lth.se

ABSTRACT

The Thomson multitaper estimator has become successful for spectrum analysis in many application areas. From the aspect of efficient implementation, the so called Welch or WOSA-Weighted Overlap Segment Averaging, has advantages. In the Welch estimator, the same, time-shifted, window is applied to the data-sequence. In this submission, the aim is to find a Welch estimator structure which has a similar performance as the Thomson multitaper estimator. Such a estimator might be more advantageous from real-time computation aspects as the spectra can be estimated when data samples are available and a running average will produce the subsequent averaged spectra. The approach is to restructure the corresponding covariance matrix of the Thomson estimator to the structure of a Welch estimator and to find a mean square error approximation of the covariance matrix. The resulting window of the Welch estimator should however fulfill the usual properties of a spectrum estimator, such as low-pass structure and well suppressed sidelobes.

Index Terms— Spectrum, Multitaper, Multiple windows, Thomson, Welch, WOSA

1. INTRODUCTION

The concept of **Multiple Windows** or **Multitapers** was invented by Thomson, [1] where the windows are the Discrete Prolate Spheroidal Sequences (DPSS), developed by Slepian and Pollack, which is described in a number of famous papers ending with [2]. The main idea of multitapers is to reduce the variance of the periodogram by averaging several uncorrelated periodograms where the same data sequence is used for all periodograms but the shape of the window change in a way that give uncorrelated periodograms and thereby reduced variance. However, multitapers were actually used much earlier in the form of one window shifted in time, the Welch method or WOSA (Weighted Overlap Segmented Averaging) by Welch, [3, 4]. The time-shifted window by Welch give uncorrelated periodograms as the time-shifted window overlap different data sequences, although the same window is used. Bronez shows that the Thomson method outperforms the Welch method in terms of leakage, resolution and variance, [5]. The multitaper estimators are used in many applications but still the simple idea of one time-shifted window of the Welch method is the most applied and is the most advantageous from real-time implementation aspects. Other suggestions of methods using and optimizing the time-shift of one window are e.g., the circularly shifted taper of the time-division multiple window method (TDMW) by Clark, [6], which has a similar performance as the Thomson estimator using just one circularly time-shifted window and all data samples. Recently, an improved Welch estimator are proposed using circular time-shifts in [7].

Thanks to the Swedish Research council for funding.

In this submission we rely on the usual structure of the Welch estimator which might be more advantageous from real-time computation aspects as the spectra can be estimated when data samples are available and a running average will produce the subsequent averaged spectra. The aim is to find a window for the time-shifted structure so that the resulting Welch-structure estimator has a similar performance as the Thomson estimator. The resulting window should fulfill the usual properties of a spectrum estimator, such as low-pass structure and well suppressed sidelobes.

In section 2 the estimation of the Thomson multitapers is presented. Section 3 gives the computation of the approximative Welch-structure estimator. In section 4, the evaluation is presented and the paper is concluded in section 5.

2. MULTITAPER SPECTRUM ANALYSIS

The real valued stationary discrete-time random process, $x(n)$, is given. We would like to estimate the spectrum $S_x(f)$, from N samples $\mathbf{x} = [x(0) \ \dots \ x(N-1)]^T$ of the process by using the estimator

$$\hat{S}_x(f) = \frac{1}{K} \sum_{k=1}^{K} \hat{S}_k(f) \tag{1}$$

where

$$\hat{S}_k(f) = \left| \sum_{n=0}^{N-1} x(n) h_k(n) e^{-i2\pi f n} \right|^2 . \qquad (2)$$

Equation (2) is a windowed periodogram obtained by using the data window $\mathbf{h}_k = [h_k(0) \ \ldots \ h_k(N-1)]^T$. An average of K periodograms provides the multitaper estimate in Eq. (1).

2.1. The Thomson estimator

The Thomson multitaper estimation method can be considered to be a filtering procedure of white noise in a filter bank of FIR-filters of pre-defined bandwidth B. The impulse responses of the subfilters are $\mathbf{h}_k$ and the corresponding frequency functions are $H_k(f) = \mathbf{h}_k^T \cdot \phi(f)$, where $\phi(f) = [1 \ e^{-i2\pi f} \ldots e^{-i2\pi(N-1)f}]^T$. The white noise spectrum is $S_w(f) = 1$. Given the input signal $x(n)$, the power of the output signal within the frequency interval $(-B/2, B/2)$ is

$$\begin{aligned} P_B &= \int_{-B/2}^{B/2} |H_k(f)|^2 S_w(f) df \\ &= \mathbf{h}_k^T \int_{-B/2}^{B/2} \phi(f) S_w(f) \phi^H(f) df \mathbf{h}_k \\ &= \mathbf{h}_k^T \int_{-1/2}^{1/2} \phi(f) S_B(f) \phi^H(f) df \mathbf{h}_k \\ &= \mathbf{h}_k^T \mathbf{R}_B \mathbf{h}_k, \end{aligned} \qquad (3)$$

where $S_B(f)$ is equal to $S_w(f) = 1$ in the band $(-B/2, B/2)$ and zero for all other frequency values. The Toeplitz $(N \times N)$ covariance matrix $\mathbf{R}_B$ has the elements $r_B(l) = \int_{-\infty}^{\infty} S_B(f) e^{-i2\pi l f} df = B\text{sinc}(\pi B l)$, $0 \leq |l| \leq N-1$, where $\text{sinc}(x) = \frac{sin(x)}{x}$. The K_T window functions, $\mathbf{h}_k$, which maximize P_B are used as multitapers and the optimization is performed subject to total power of a window equals one, i.e.,

$$P_{tot} = \int_{-1/2}^{1/2} |H_k(f)|^2 df = \mathbf{h}_k^T \mathbf{h}_k = 1.$$

The solution with respect to $\mathbf{h}_k$ is the set of eigenvectors of the eigenvalue problem

$$\mathbf{R}_B \mathbf{q}_k = \lambda_k \mathbf{q}_k, \quad k = 1 \ldots N. \qquad (4)$$

The solution is found as the Discrete Prolate Spheroidal Sequences (DPSS). It is also shown that the number of eigenvalues K_T close to one, which is related to the predefined frequency bandwidth B and window length N as

$$K_T \approx N \cdot B - 2.$$

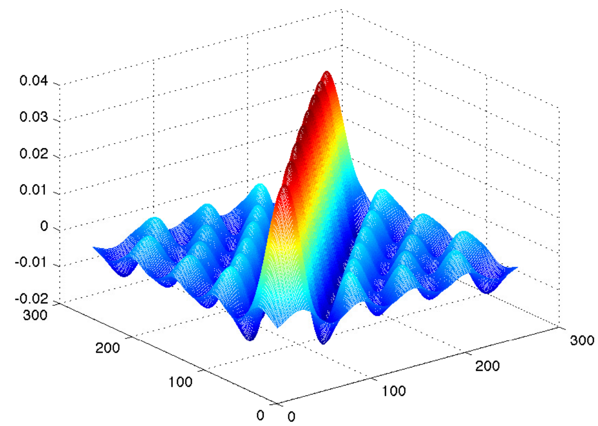

Fig. 1. The resulting covariance matrix $\mathbf{R}_{th}$ of the Thomson method with $N = 256$, $B = 0.04$ and $K_T = 8$.

3. A THOMSON-WELCH STRUCTURE APPROXIMATION

One way to find an approximative Thomson estimator with the structure of a Welch estimator is to approximate the covariance matrix of the Thomson estimator with a similar one with the appropriate structure of a time-shifted window.

The covariance matrix of the Thomson estimator is found as

$$\mathbf{R}_{th} = \sum_{k=1}^{K_T} \lambda_k \mathbf{q}_k \mathbf{q}_k^T . \qquad (5)$$

An example of the covariance matrix of the Thomson estimator is seen in Figure 1. The covariance matrix can also be written as

$$\mathbf{R}_{th} = \mathbf{Q} \mathbf{\Lambda}^2 \mathbf{Q}^T = \mathbf{A} \mathbf{A}^T, \qquad (6)$$

with $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}$ and

$$\begin{aligned} \mathbf{Q} &= [\mathbf{q}_1 \ldots \mathbf{q}_{K_T}], \\ \mathbf{\Lambda} &= \text{diag}\left[\sqrt{\lambda_1} \ldots \sqrt{\lambda_{K_T}}\right]. \end{aligned} \qquad (7)$$

The covariance matrix of the Thomson-Welch-approximation is found as

$$\mathbf{R}_{TW} = \mathbf{H}\mathbf{H}^T, \qquad (8)$$

where

$$\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \ldots \mathbf{h}_{K_W}], \qquad (9)$$

with $\mathbf{h}_k$, $k = 1 \ \ldots \ K_W$ defined from

$$\mathbf{h}_k = [\underbrace{0 \ \ldots \ 0}_{(k-1)L_W} \ \mathbf{h}^T \ \overbrace{0 \ \ldots \ 0}^{(N-(k+1))L_W} \ ]^T, \quad k = 1 \ \ldots \ K_W, \qquad (10)$$

where $\mathbf{h} = [h(0) \ \ldots \ h(N_W - 1)]^T$ is the unknown window to be estimated. The number of windows K_W of length N_W is the largest integer that fulfill $K_W \leq \frac{N-N_W}{L_W} + 1$, where L_W is the time-shift. The parameters K_W, L_W and N_W are assumed to be pre-defined.

A number of important criteria of the window $\mathbf{h}$ can be found, e.g., smoothness and low sidelobes in the frequency plane. Therefore a straightforward optimization might not always give a desirable result. The approach taken here is based on finding a new $(N_W \times K_T K_W)$ matrix $\mathbf{B}$ from submatrices of $\mathbf{A}$ according to

$$\mathbf{B} = [\mathbf{B_1} \ \ \mathbf{B_2} \ \ \ldots \ \ \mathbf{B_{K_W}}], \tag{11}$$

where

$$\mathbf{B}_k = \begin{bmatrix} a_{(k-1)L_W+1,1} & \cdots & a_{(k-1)L_W+1,K_T} \\ a_{(k-1)L_W+2,1} & \cdots & a_{(k-1)L_W+2,K_T} \\ \vdots & & \vdots \\ a_{(k-1)L_W+N_W,1} & \cdots & a_{(k-1)L_W+N_W,K_T} \end{bmatrix},$$

and $a_{n,m}$ are the elements of the matrix $\mathbf{A}$. Studying the structure of $\mathbf{B}$ show that if all columns would be equal to $\mathbf{h}$, the resulting estimator would be of Welch-structure. This is of course just possible in the non-overlapping window case, i.e., $L_W = N_W$. However, the elements of $\mathbf{A}$ and thereby also $\mathbf{B}$ are specified and to find a possible approximative window vector $\mathbf{h}$, the singular value decomposition of $\mathbf{B}$ are computed as

$$\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{12}$$

and the first K_B singular vectors of the $(N_W \times N_W)$ matrix $\mathbf{U}$ are used as basis functions for further analysis. These singular vectors will contain the most of the common structure of the vectors in the matrix $\mathbf{B}$ and could possibly serve as a basis in the optimization. Then the aim is to find a final window as

$$\mathbf{h} = \sum_{k=1}^{K_B} \alpha_k \mathbf{u}_k, \tag{13}$$

where $\mathbf{u}_k$, $k = 1 \ \ldots K_B$, are the singular vectors of the matrix $\mathbf{U}$ and α_k parameters to be optimized. The optimization is performed using the mean square error of the corresponding covariance matrices,

$$\min_{\alpha_k} \sum_{n=1}^{N} \sum_{m=1}^{N} (r_{th}(n,m) - r_{TW}(n,m))^2, \tag{14}$$

for $k = 1 \ \ldots \ K_B$, where $r_{th}(n,m)$ and $r_{TW}(n,m)$ are the elements of the matrices $\mathbf{R}_{th}$ and $\mathbf{R}_{TW}$ respectively. For known parameters α_k, the final window for the time-shifted approximation is computed from Eq. (13). For the computation, the Nelder-Mead simplex method of Matlab (fminsearch) with equal initial starting point $\alpha_k = 1/K_B$ is used. Further investigations are needed to evaluate the convergence and possible local minima.

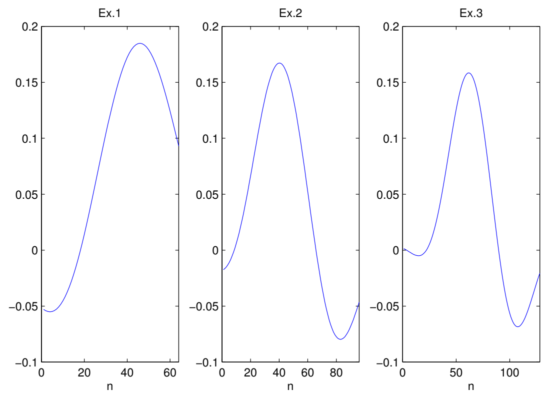

Fig. 2. The window $\mathbf{h}$ computed for different parameter settings: (Case 1) $N_W = 64$, $L_W = 16$ and $K_W = 13$, (Case 2) $N_W = 96$, $L_W = 16$ and $K_W = 11$, (Case 3) $N_W = 128$, $L_W = 16$ and $K_W = 9$.

4. EVALUATION

In the examples and evaluations $N = 256$, $B = 0.04$ and $K_T = 8$, are used for the Thomson method. We show three cases of pre-defined parameter settings for approximation where the window $\mathbf{h}$ is computed for different parameter settings: (1) $N_W = 64$, $L_W = 16$ and $K_W = 13$, (2) $N_W = 96$, $L_W = 16$ and $K_W = 11$, (3) $N_W = 128$, $L_W = 16$ and $K_W = 9$. The resulting windows are presented in Figure 2. We see that the resulting window will have a better performance regarding leakage for a larger value of N_W. This is natural as it is difficult to fulfill the suppression with a shorter window. Some similarity of the shape can be seen, especially for the case 2 compared to case 3, where it seems as the edges of case 3 have been cut for the shorter window in case 2. The resulting covariance matrices $\mathbf{R}_{TW}$ are computed and are displayed in Figures 3, 4 and 5, where we can see that the structure of $\mathbf{R}_{th}$ is essentially kept for the diagonal and sub-diagonal elements.

4.1. Examples

The performance is computed for three different cases of processes, white Gaussian noise, a band-pass process and an AR(4)-process where we compare the result of the original Thomson estimator with the approximative estimators. Note that we not compare which estimator that has the best performance for the specific process, we just compare the performance with the original Thomson estimator. For the white noise spectrum, the process covariance matrix $\mathbf{R}_x = \mathbf{I}$ and for the other processes the process covariance matrices are given from the corresponding spectra. The bandpass process is computed from the FIR-filter of order 100 and cut-off frequencies 0.1 and 0.2 using a Hamming-window and the AR(4)-process are given by the poles $p_{1,2} = 0.95e^{\pm i2\pi 0.1}$

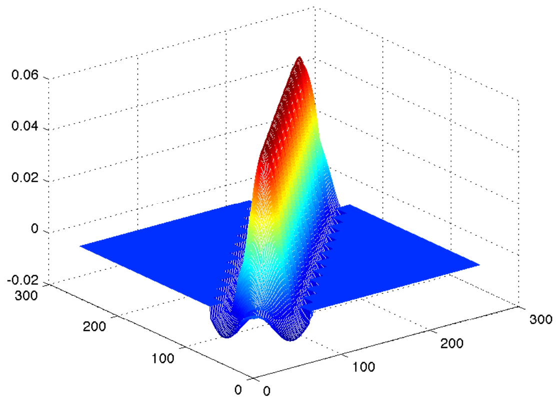

Fig. 3. The resulting covariance matrix $\mathbf{R}_{TW}$ for case 1 with parameters $N_W = 64$, $L_W = 16$ and $K_W = 13$.

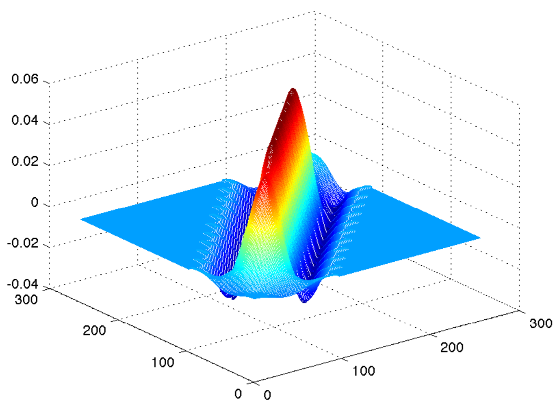

Fig. 4. The resulting covariance matrix $\mathbf{R}_{TW}$ for case 2 with parameters $N_W = 96$, $L_W = 16$ and $K_W = 11$.

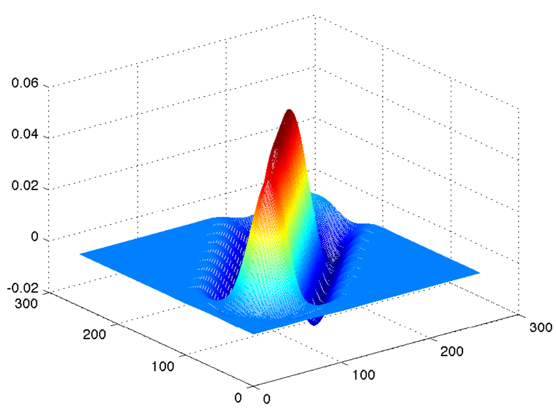

Fig. 5. The resulting covariance matrix $\mathbf{R}_{TW}$ for case 3 with parameters $N_W = 128$, $L_W = 16$ and $K_W = 9$.

and $p_{3,4} = 0.95e^{\pm i2\pi 0.2}$. The expected value of the spectrum estimate is computed as

$$\begin{aligned} E[\hat{S}_x(f)] &= E[\frac{1}{K}\sum_{k=1}^{K}\mathbf{h}_k^T\mathbf{\Phi}^H(f)\mathbf{x}\mathbf{x}^T\mathbf{\Phi}(f)\mathbf{h}_k] \\ &= \frac{1}{K}\sum_{k=1}^{K}\mathbf{h}_k^T\mathbf{\Phi}^H(f)\mathbf{R}_x\mathbf{\Phi}(f)\mathbf{h}_k, \end{aligned} \quad (15)$$

where $\mathbf{\Phi}(f) = \text{diag}[1 \;\; e^{-i2\pi f} \;\; \ldots \;\; e^{-i2\pi(N-1)f}]$ is the Fourier transform matrix. The variance of the spectrum estimate is given by all combinations of the different periodogram covariances,

$$\text{Variance } \hat{S}_x(f) = \frac{1}{K^2}\sum_{k=1}^{K}\sum_{l=1}^{K}\text{cov}(\hat{S}_k(f)\hat{S}_l(f)). \quad (16)$$

Denoting $\mathbf{h}_k^T\mathbf{\Phi}^H(f)\mathbf{x} = C_k$ and assuming $\mathbf{x}$ to be Gaussian gives the covariance as

$$\begin{aligned} &\text{cov}(\hat{S}_k(f)\hat{S}_l(f)) = \text{cov}(C_kC_k^HC_lC_l^H) \\ =\; &E[C_k^HC_kC_lC_l^H] - E[C_k^HC_k]E[C_lC_l^H] \\ =\; &E[C_k^HC_l]E[C_kC_l^H] + E[C_k^HC_l^H]E[C_kC_l] \\ \approx\; &|\mathbf{h}_k^T\mathbf{\Phi}^H(f)\mathbf{R}_x\mathbf{\Phi}(f)\mathbf{h}_l|^2 \end{aligned} \quad (17)$$

according to Walden et al, [8].

Comparing the results for the white Gaussian noise case, the original Thomson method gives the expected value $E[\hat{S}_x(f)] = 1$ where the approximative method for the different cases give: (1) 0.906, (2) 0.837, (3) 0.809. The variance of the Thomson method is Variance $\hat{S}_x(f) = 1/K_T = 0.125$ where the approximative method for the different cases give: (1) 0.101, (2) 0.113, (3) 0.143. The smaller variance of course comes from the large number of time-shifted windows, e.g., case 1, where $K_W = 13$ and case 2, where $K_W = 11$.

For the bandpass process, the results of the expected case 3 give the best sidelobe suppression (black line) compared to the Thomson estimator (blue line), see Figure 6a). The variance performance is similar to the white noise case.

For the AR(4)-process, all methods perform very similar to the Thomson estimator, see Figure 7. The best performance, studying the expected value is given from case 1 (green line) where the performance most similar to the Thomson method is given by the approximative method of case 2 (red line) or possibly case 3 (black line). This is the typical trade-off, when the initial parameters are chosen, between the window length N_W and the number of windows K_W.

5. CONCLUSION

In this submission an approximative Welch-structure spectrum estimator of the Thomson multitaper method is proposed. The window-length and number of windows of the

Welch-estimator are specified and the window, to be shifted, is computed from the mean square error of the covariance matrices of the proposed method and the Thomson estimator. Such an estimator might be more advantageous from real-time computation aspects as the spectra can be estimated when data samples are available and a running average will produce the subsequent averaged spectra. The performance of the new estimator is compared to the Thomson estimator for some examples.

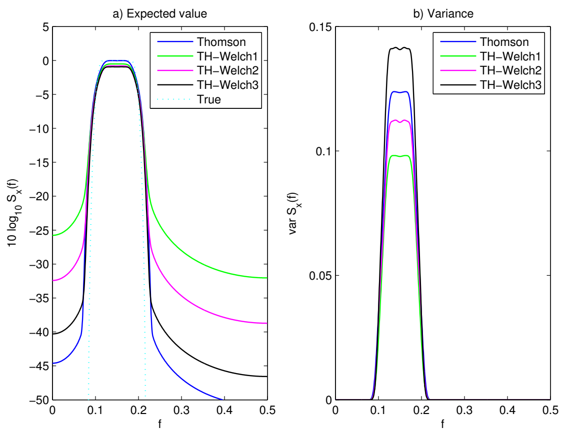

Fig. 6. A comparison of the a) expected values b) variances for the three cases of the approximative method and the Thomson estimator for a band-pass process.

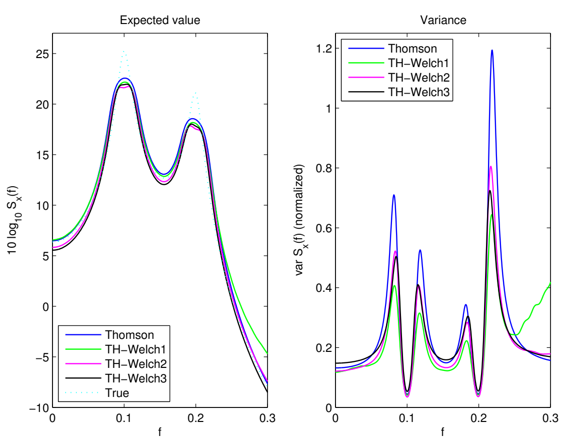

Fig. 7. A comparison of the a) expected values b) variances (normalized with true spectrum for a better view) for the three cases of the approximative method and the Thomson estimator for an AR(4)-process.

6. REFERENCES

[1] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sept 1982.

[2] D. Slepian, "Prolate spheriodal wave functions, Fourier analysis and uncertainty-v: The discrete case," *Bell System Journal*, vol. 57, no. 5, pp. 1371–1430, May-June 1978.

[3] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. on Audio Electroacoustics*, vol. AU-15, no. 2, pp. 70–73, June 1967.

[4] A. T. Walden, "A unified view of multitaper multivariate spectral estimation," *Biometrika*, vol. 87, no. 4, pp. 767–788, 2000.

[5] T. P. Bronez, "On the performance advantage of multitaper spectral analysis," *IEEE Trans. on Signal Processing*, vol. 40, no. 12, pp. 2941–2946, Dec 1992.

[6] M. P. Clark and C. T. Mullis, "Quadratic estimation of the power spectrum using orthogonal time-division multiple windows," *IEEE Trans. on Signal Processing*, vol. 41, no. 1, pp. 222–231, Jan 1993.

[7] K. Barbé, R. Pintelon, and J. Schoukens, "Welch method revisited: Nonparametric power spectrum estimation via circular overlap," *IEEE Trans. on Signal Processing*, vol. 58, no. 2, pp. 553–565, Sept 2010.

[8] A. T. Walden, E. McCoy, and D. B. Percival, "The variance of multitaper spectrum estimates for real gaussian processes," *IEEE Trans. on Signal Processing*, vol. 42, no. 2, pp. 479–482, Feb 1994.