A Wide Dynamic Range Multichannel Spectral Estimator

Cuichun Xu and Steven Kay

Dept. of Electrical and Computer Engineering

University of Rhode Island

Kingston, RI 02881

401-874-5804 (voice) 401-782-6422 (fax)

ccxu@ele.uri.edu

EDICS: MDS - SPEC, SSP - SPEC

March 22, 2006

Abstract

A multichannel extension of the AR matrix prewhitened power spectral density estimator, which was originally developed for a single channel, is proposed. In order to make the extension, the Cholesky decomposition of the inverse autocorrelation matrix for a multichannel autoregressive process is discussed and the autoregressive model order selection for a multichannel process based on the exponentially embedded families criterion is introduced. The asymptotic mean and variance of the proposed estimator are derived. Compared to a filter-based autoregressive prewhitened multichannel power spectral estimator, the new estimator has less bias, i.e. higher resolution, and less overall mean square error for short data records due to the amelioration of end effects by the matrix prewhitener. It can serve as an excellent multichannel spectral estimator for processes exhibiting a wide dynamic range. Simulation results are given which show the advantage of the new estimator over a variety of common multichannel power spectral density estimators.

1 Introduction

Multichannel power spectral density (MPSD) estimation techniques are widely used in many applications, such as sonar, radar, geophysics and biomedicine. Similar to single channel power spectral density (PSD) estimation, there are basically two broad categories of MPSD estimators. One is the nonparametric approach, among which the Fourier-based estimators are the most popular [1–3]. The other is the parametric method, which assumes a model for the data. Spectral estimation then becomes a problem of estimating the parameters in the assumed model. The most commonly used model is the autoregressive (AR) model because accurate estimates of the AR parameters can be found by solving a set of linear equations [1,3,5].

Similar to the single channel case for short data records the Fourier-based methods can suffer from significant bias problems while AR model-based methods can suffer from inaccuracies in the model as well as from imprecise model order selection. Furthermore, some effective AR model-based approaches cannot be easily extended to the multichannel case [1,5]. In addition, as pointed out by Jenkins and Watts [2], spurious cross-correlation or spurious cross-spectral content may arise unless a prewhitening filter is applied before MPSD estimation. One such prewhitening filter was suggested by Thomson [6] for single channel PSD estimation. The filter system function is given by $A(z) = 1 + \sum_{k=1}^{p} a[k]z^{-k}$ and the filter parameters $a[1], a[2], \cdots, a[p]$ can be estimated from the data using any AR model-based method. Denoting the output of this FIR filter by $u[n]$, a Fourier-based estimator is then used to generate the PSD estimate

$\hat{P}_u(f)$. Finally, the PSD estimate of the original data is found as [6]

$$\hat{P}_x(f) = \frac{\hat{P}_u(f)}{|1 + \sum_{k=1}^{p} \hat{a}[k] \exp(-j2\pi f k)|^2} \tag{1}$$

where $\hat{a}[1], \hat{a}[2], \cdots, \hat{a}[p]$ are the estimated AR filter parameters. We term this the *AR prewhitened* (ARPW) spectral estimator.

Because of the inconsistency of the definitions in the literature concerning MPSD estimation, the following definitions will be made. A complex multichannel sequence $\mathbf{x}[n]$ is defined as the complex $L \times 1$ vector

$$\mathbf{x}[n] = [x_1[n], x_2[n], \cdots, x_L[n]]^T$$

where $x_i[n]$ represents the data observed at the output of the ith channel and L is the number of channels. For a wide sense stationary (WSS) multichannel random process, the autocorrelation function (ACF) at lag k is defined as the $L \times L$ matrix function

$$\begin{aligned}
\mathbf{R}_x[k] &= E\left[\mathbf{x}[n+k]\mathbf{x}^H[n]\right] \\
&= \begin{bmatrix}
r_{11}[k] & r_{12}[k] & \cdots & r_{1L}[k] \\
r_{21}[k] & r_{22}[k] & \cdots & r_{2L}[k] \\
\vdots & \vdots & \ddots & \vdots \\
r_{L1}[k] & r_{L2}[k] & \cdots & r_{LL}[k]
\end{bmatrix}
\end{aligned} \tag{2}$$

where $E[\cdot]$ is the mathematical expectation, the superscript H denotes conjugate transpose and $r_{ij}[k]$ is the cross-correlation function (CCF) between $x_i[n]$ and $x_j[n]$ at lag k or

$$r_{ij}[k] = E\left[x_i[n+k]x_j^*[n]\right]. \tag{3}$$

For multichannel data of N samples, the sample vector, which is $NL \times 1$, is defined as

$$\mathbf{x} = \left[\mathbf{x}^T[0], \mathbf{x}^T[1], \cdots, \mathbf{x}^T[N-1]\right]^T. \tag{4}$$

The $NL \times NL$ multichannel autocorrelation matrix (ACM) of order N is defined as

$$\begin{aligned}
\mathbf{R}_x &= E\left[\mathbf{x}\mathbf{x}^H\right] \\
&= \begin{bmatrix}
\mathbf{R}_x[0] & \mathbf{R}_x[-1] & \cdots & \mathbf{R}_x[-N+1] \\
\mathbf{R}_x[1] & \mathbf{R}_x[0] & \cdots & \mathbf{R}_x[-N+2] \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{R}_x[N-1] & \mathbf{R}_x[N-2] & \cdots & \mathbf{R}_x[0]
\end{bmatrix}.
\end{aligned} \tag{5}$$

From definition (2) it is seen that $\mathbf{R}_x^H[k] = \mathbf{R}_x[-k]$, so $\mathbf{R}_x$ is hermitian. Because the multichannel process is assumed to be wide sense stationary, $\mathbf{R}_x$ is also block Toeplitz.

The power spectral density matrix or cross-spectral matrix is defined as

$$\mathbf{P}_x(f) = \begin{bmatrix}
P_{11}(f) & P_{12}(f) & \cdots & P_{1L}(f) \\
P_{21}(f) & P_{22}(f) & \cdots & P_{2L}(f) \\
\vdots & \vdots & \ddots & \vdots \\
P_{L1}(f) & P_{L2}(f) & \cdots & P_{LL}(f)
\end{bmatrix}.$$

The diagonal elements $P_{ii}(f)$ are the PSDs of the individual channels or auto-PSDs, while the off-diagonal elements $P_{il}(f)$ for $i \neq l$ are the cross-PSDs between $x_i[n]$ and $x_l[n]$, which are defined as

$$P_{il}(f) = \sum_{k=-\infty}^{\infty} r_{il}[k] \exp(-j2\pi fk). \tag{6}$$

The magnitude squared coherence (MSC) between channels i and j is a quantity that indicates whether the spectral amplitude of the process at a particular frequency in channel i is associated with large or small spectral amplitude at the same frequency in channel j. It is defined as

$$|\gamma_{ij}(f)|^2 = \frac{|P_{ij}(f)|^2}{P_{ii}(f)P_{jj}(f)}. \tag{7}$$

A classic Fourier-based spectral estimator is the periodogram, which is given as the $L \times L$ matrix

$$\hat{\mathbf{P}}_{PER}(f) = \frac{1}{N} \mathbf{X}(f) \mathbf{X}^H(f) \tag{8}$$

where the Fourier transform is the $L \times 1$ vector

$$\mathbf{X}(f) = \sum_{n=0}^{N-1} \mathbf{x}[n] \exp(-j2\pi fn). \tag{9}$$

The multichannel pth order AR model is defined as

$$\mathbf{x}[n] = -\sum_{i=1}^{p} \mathbf{A}[i]\mathbf{x}[n-i] + \mathbf{u}[n] \tag{10}$$

where $\mathbf{A}[1], \mathbf{A}[2], \cdots, \mathbf{A}[p]$ are $L \times L$ AR coefficient matrices and $\mathbf{u}[n]$ is the excitation white noise or

$$\mathbf{R}_u[k] = \boldsymbol{\Sigma}\delta[k]$$

and $\boldsymbol{\Sigma}$ is the $L \times L$ excitation noise covariance matrix with $\delta[\cdot]$ being the discrete delta function.

The ARPW estimator given in (1) is readily extended to the multichannel case. With the notations defined above, the multichannel version of (1) is

$$\hat{\mathbf{P}}_x(f) = \hat{\mathbf{A}}^{-1}(f)\hat{\mathbf{P}}_u(f)\hat{\mathbf{A}}^{-H}(f) \tag{11}$$

where $\hat{\mathbf{A}}(f) = \mathbf{I} + \sum_{i=1}^{p} \hat{\mathbf{A}}[i] \exp(-j2\pi fi)$ and $\hat{\mathbf{A}}[1], \hat{\mathbf{A}}[2], \cdots, \hat{\mathbf{A}}[p]$ are the estimated multichannel AR filter parameters. In addition to reducing spurious cross-spectral content, this prewhitened spectral estimator also gives an auto-PSD spectral estimate with much less bias than a Fourier-based spectral estimator. This is because the prewhitener reduces the dynamic range of the PSD. However, this estimator is still inferior to the method proposed in this paper. Instead of the FIR prewhitening filter, the proposed estimator uses a prewhitening *matrix*, which is essentially a time-varying filter that is less susceptible to end effects. The new estimator for a single channel PSD has been proposed in [4], while in this paper it is extended to MPSD estimation.

Another widely used MPSD spectral estimator is the minimum variance spectral estimator (MVSE) [1,13]. Also called the maximum likelihood spectral estimator, it was first proposed by Capon for single channel spectral estimation and is easily generalized to the multichannel case. This MPSD estimator can

be thought of as motivated by a maximum likelihood (ML) estimation problem. Assume the multichannel data are

$$\mathbf{x}[n] = \mathbf{A}_c \exp\left(j2\pi f_0 n\right) + \mathbf{w}\left[n\right], \qquad n = 0, 1, \cdots, N - 1 \tag{12}$$

where $\mathbf{x}[n]$ is an $L \times 1$ vector, $\mathbf{A}_c$ is an $L \times 1$ complex amplitude to be estimated, f_0 is a known frequency, and $\mathbf{w} = \left[\mathbf{w}^T[0], \mathbf{w}^T[1], \cdots, \mathbf{w}^T[N-1]\right]^T$ is a $NL \times 1$ complex Gaussian noise vector with zero mean and known $NL \times NL$ covariance matrix $\mathbf{R}_w$. The ML estimate of $\mathbf{A}_c$ is [1]

$$\hat{\mathbf{A}}_c = \left(\mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{E}_0\right)^{-1} \mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{x} \tag{13}$$

where $\mathbf{x}$ is the data sample vector given by (4) and $\mathbf{E}_0 = [\mathbf{I}_L, \mathbf{I}_L \exp\left(j2\pi f_0\right), \cdots, \mathbf{I}_L \exp\left(j2\pi f_0\left(N-1\right)\right)]^T$ $(NL \times L)$ with $\mathbf{I}_L$ being an $L \times L$ identity matrix. The $L \times L$ covariance matrix of this estimator is [1]

$$\mathbf{C}_{\hat{A}_c} = \left(\mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{E}_0\right)^{-1}. \tag{14}$$

This covariance can be shown to be the noise output covariance matrix of a narrowband multichannel FIR filter which is constrained to have unity frequency response at $f = f_0$ and adjusts itself to reject noise components with frequencies not near $f = f_0$ [1]. As a result, for a general WSS multichannel random process, not necessarily a sinusoid in noise, the covariance matrix at the output of the filter can be used to indicate the power of the process at $f = f_0$. Therefore, for a general WSS multichannel random process $\mathbf{x}[n]$ the MVSE is defined as

$$\hat{\mathbf{P}}_{MV}\left(f\right) = p(\mathbf{E}^H(f)\hat{\mathbf{R}}_x^{-1}\mathbf{E}(f)^{-1} \tag{15}$$

where $\hat{\mathbf{R}}_x$ is the estimated $pL \times pL$ ACM of $\mathbf{x}$ and $\mathbf{E(f)} = [\mathbf{I}_L, \mathbf{I}_L \exp\left(j2\pi f\right), \cdots, \mathbf{I}_L \exp\left(j2\pi f\left(p-1\right)\right)]^T$ $(pL \times L)$. This spectral estimator is the covariance matrix of $\hat{\mathbf{A}}_c$ divided by $1/p$, which can be thought of as the filter bandwidth [1]. Further details, such as the choice of p, the estimation of $\hat{\mathbf{R}}_x$, etc., can be found in [1].

The paper is organized as follows. The motivation and properties of the proposed estimator are presented in Section 2, while Section 3 describes its implementation. Some computer simulation results are given in Section 4 and finally Section 5 offers some conclusions.

2 Multichannel AR Matrix Prewhitened Spectral Estimator

2.1 Motivation for the New Estimator

Similar to the derivation of the MVSE discussed in the previous section, the data structure of (12) is used. However, we now assume that $\mathbf{A}_c$ is an $L \times 1$ complex Gaussian *random* vector with zero mean and unknown covariance matrix $\mathbf{C}_A$, which is to be estimated. We also assume that $\mathbf{A}_c$ is independent of the noise $\mathbf{w}[n]$. We note that the basic data assumption here is that the noise samples $\mathbf{w}[0], \mathbf{w}[1], \cdots, \mathbf{w}[N-1]$ are correlated and hence a simple ML result for independent and identically distributed samples does not apply. As derived in Appendix A, the ML estimator of $\mathbf{C}_A$ is

$$\hat{\mathbf{C}}_\mathbf{A} = S1 - S2 \tag{16}$$

where

$$S1 = \left(\mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{E}_0\right)^{-1} \mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{x}\mathbf{x}^H \mathbf{R}_w^{-1} \mathbf{E}_0 \left(\mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{E}_0\right)^{-1}$$

and

$$S2 = \frac{\left(\mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{E}_0\right)^{-1} \mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{x}\mathbf{x}^H \mathbf{R}_w^{-1} \mathbf{E}_0 \left(\mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{E}_0\right)^{-1}}{\mathbf{x}^H \mathbf{R}_w^{-1} \mathbf{E}_0 \left(\mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{E}_0\right)^{-1} \mathbf{E}_0^H \mathbf{R}_w^{-1} \mathbf{x}}.$$

Since $\mathbf{A}_c$ is independent of the noise vector $\mathbf{w}$, the ACM of $\mathbf{x}$ is

$$
\begin{aligned}
E\left(\mathbf{x}\mathbf{x}^H\right) &= E\left(\left(\mathbf{E}_0\mathbf{A}_c + \mathbf{w}\right)\left(\mathbf{A}_c^H\mathbf{E}_0^H + \mathbf{w}^H\right)\right) \\
&= \mathbf{E}_0\mathbf{C}_A\mathbf{E}_0^H + \mathbf{R}_w.
\end{aligned}
$$

As a result, the expectation of $S1$ is

$$
\begin{aligned}
&E\left(\left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{x}\mathbf{x}^H\mathbf{R}_w^{-1}\mathbf{E}_0\left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}\right) &(17)\\
&= \left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}\mathbf{E}_0^H\mathbf{R}_w^{-1}E\left(\mathbf{x}\mathbf{x}^H\right)\mathbf{R}_w^{-1}\mathbf{E}_0\left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}\\
&= \left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}\mathbf{E}_0^H\mathbf{R}_w^{-1}\left(\mathbf{E}_0\mathbf{C}_A\mathbf{E}_0^H + \mathbf{R}_w\right)\mathbf{R}_w^{-1}\mathbf{E}_0\left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}\\
&= \mathbf{C}_A + \left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}.
\end{aligned}
$$

Since from (14), $\left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}$ is the covariance matrix of the noise at the narrowband filter output, $\mathbf{C}_A + \left(\mathbf{E}_0^H\mathbf{R}_w^{-1}\mathbf{E}_0\right)^{-1}$ will be the sinusoid plus noise covariance matrix at the filter output. Furthermore, $S2$ can be thought of as an approximate estimate of the noise covariance matrix at the filter output. So $S1 - S2$ will be an estimate of the sinusoid covariance matrix or $\mathbf{C}_A$. (It can be shown, for a multichannel process $S2$ is a biased estimate of the noise power at the filter output, so $\hat{\mathbf{C}}_A$ is also a biased estimator.) Since the expected value of $S1$ is the overall power at the output of a narrowband filter, it can be used to indicate spectral content on average. Suppose the data does not consist solely of a sinusoid in noise, but more generally a wide sense stationary random process. Then we can define a new spectral estimator based on $S1$ as

$$
\hat{\mathbf{P}}_x\left(f\right) = N\left(\mathbf{E}^H\left(f\right)\hat{\mathbf{R}}_x^{-1}\mathbf{E}\left(f\right)\right)^{-1}\mathbf{E}^H\left(f\right)\hat{\mathbf{R}}_x^{-1}\mathbf{x}\mathbf{x}^H\hat{\mathbf{R}}_x^{-1}\mathbf{E}\left(f\right)\left(\mathbf{E}^H\left(f\right)\hat{\mathbf{R}}_x^{-1}\mathbf{E}\left(f\right)\right)^{-1} \quad (18)
$$

where $\mathbf{E}\left(f\right) = \left[\mathbf{I}_L, \mathbf{I}_L\exp\left(j2\pi f\right), \cdots, \mathbf{I}_L\exp\left(j2\pi f\left(N-1\right)\right)\right]^T$ $(NL \times L)$ and $\hat{\mathbf{R}}_x$ is some suitable estimate of the $NL \times NL$ ACM $\mathbf{R}_x$. Again, the bandwidth of the filter $1/N$ is used to convert the power to power spectral density. We term this the *multichannel AR matrix prewhitened* (MARMPW) spectral estimator for reasons which will become clear shortly. If $L = 1$ or for a single channel, after some algebra (18) can be simplified to

$$
\hat{P}_x\left(f\right) = N\left|\frac{\mathbf{e}^H\left(f\right)\hat{\mathbf{R}}_x^{-1}\mathbf{x}}{\mathbf{e}^H\left(f\right)\hat{\mathbf{R}}_x^{-1}\mathbf{e}\left(f\right)}\right|^2 \quad (19)
$$

where $\hat{\mathbf{R}}_x$ is an estimate of the $N \times N$ ACM of the single channel process and $\mathbf{e}\left(f\right) = \left[1, \exp\left(j2\pi f\right), \cdots, \exp\left(j2\pi f\left(N-1\right)\right)\right]^T$ $(N \times 1)$. This is the ARMPW estimator proposed in [4]. Although the new multichannel spectral estimator is for complex data, it is also applicable to real data. Of course then $\mathbf{E}^H\left(f\right)\hat{\mathbf{R}}_x^{-1}\mathbf{E}\left(f\right)$ will be real since $\hat{\mathbf{R}}_x^{-1}$ is a real symmetric matrix.

2.2 Properties of the MARMPW

The MARMPW spectral estimator is asymptotically unbiased. This means that as $N \to \infty$, we have $E\left[\hat{\mathbf{P}}_x\left(f\right)\right] \to \mathbf{P}_x\left(f\right)$. To see this, first note that if the autocorrelation matrix is a continuous function of a fixed number of parameters, then substituting the consistent estimates of these parameters, we have $\hat{\mathbf{R}}_x \to \mathbf{R}_x$ in probability. If $\mathbf{R}_x$ is invertible and hence is a continuous function of its elements, then $\hat{\mathbf{R}}_x^{-1} \to \mathbf{R}_x^{-1}$ in probability [7]. Finally, by Slutsky's theorem [7] we can replace $\hat{\mathbf{R}}_x$ by $\mathbf{R}_x$ in (18) to yield

$$
\hat{\mathbf{P}}_x\left(f\right) = N\left(\mathbf{E}^H\left(f\right)\mathbf{R}_x^{-1}\mathbf{E}\left(f\right)\right)^{-1}\mathbf{E}^H\left(f\right)\mathbf{R}_x^{-1}\mathbf{x}\mathbf{x}^H\mathbf{R}_x^{-1}\mathbf{E}\left(f\right)\left(\mathbf{E}^H\left(f\right)\mathbf{R}_x^{-1}\mathbf{E}\left(f\right)\right)^{-1}. \quad (20)
$$

5

Taking the expectation we have

$$
\begin{aligned}
E\left[\hat{\mathbf{P}}_x\left(f\right)\right] &= N\left(\mathbf{E}^H\left(f\right)\mathbf{R}_x^{-1}\mathbf{E}\left(f\right)\right)^{-1} \\
&= \left(\frac{\mathbf{E}^H\left(f\right)}{\sqrt{N}}\mathbf{R}_x^{-1}\frac{\mathbf{E}\left(f\right)}{\sqrt{N}}\right)^{-1}.
\end{aligned}
\tag{21}
$$

Since $\mathbf{R}_x$ is a hermitian and block Toeplitz matrix, it is well known that as $N \to \infty$ [9,10]

$$
\mathbf{R}_x^{-1} \to \sum_{i=0}^{N-1}\frac{\mathbf{E}\left(f_i\right)}{\sqrt{N}}\mathbf{P}_x^{-1}\left(f_i\right)\frac{\mathbf{E}^H\left(f_i\right)}{\sqrt{N}}
\tag{22}
$$

where $f_i = i/N$. Therefore we have for $f_k = k/N$ that as $N \to \infty$

$$
\begin{aligned}
\frac{\mathbf{E}^H\left(f_k\right)}{\sqrt{N}}\mathbf{R}_x^{-1}\frac{\mathbf{E}\left(f_k\right)}{\sqrt{N}} &\to \frac{\mathbf{E}^H\left(f_k\right)}{\sqrt{N}}\sum_{i=0}^{N-1}\frac{\mathbf{E}\left(f_i\right)}{\sqrt{N}}\mathbf{P}_x^{-1}\left(f_i\right)\frac{\mathbf{E}^H\left(f_i\right)}{\sqrt{N}}\frac{\mathbf{E}\left(f_k\right)}{\sqrt{N}} \\
&= \mathbf{P}_x^{-1}\left(f_k\right).
\end{aligned}
\tag{23}
$$

This follows from $\mathbf{E}^H\left(f_i\right)\mathbf{E}\left(f_k\right) = N\mathbf{I}_L\delta_{ik}$, where δ_{ik} is the Kronecker delta. Thus, we have finally from (21) and (23) that for $f = f_k = k/N$

$$
E\left[\hat{\mathbf{P}}_x\left(f_k\right)\right] \to \mathbf{P}_x\left(f_k\right) \text{ as } N \to \infty.
$$

By a continuity argument, we have $E\left[\hat{\mathbf{P}}_x\left(f\right)\right] \to \mathbf{P}_x\left(f\right)$ for continuous spectra. Thus, the MARMPW spectral estimator is an asymptotically unbiased estimator. By a similar argument (see Appendix B) it can be shown that as $N \to \infty$, $\mathrm{var}\left(\hat{P}_{ij}\left(f\right)\right) \to P_{ii}\left(f\right)P_{jj}\left(f\right)$. This is the same asymptotic variance as for the multichannel periodogram [2].

3 Implementation of the MARMPW

3.1 Estimation of the Matrix Prewhitener

As explained in [4] the inverse of the ACM $\mathbf{R}_x^{-1}$ in (18) acts as the prewhitener. In order to estimate this autocorrelation matrix prewhitener the Cholesky decomposition of $\mathbf{R}_x^{-1}$ for a complex WSS random process is derived in Appendix C. It is given as

$$
\mathbf{R}_x^{-1} = \mathbf{\Phi}^H\mathbf{P}^{-1}\mathbf{\Phi}
\tag{24}
$$

where $\mathbf{\Phi}$ is the $NL \times NL$ lower triangular matrix

$$
\mathbf{\Phi} = \begin{bmatrix}
\mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{A}_1[1] & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{A}_2[2] & \mathbf{A}_2[1] & \mathbf{I} & \cdots & \mathbf{0} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathbf{A}_{N-1}[N-1] & \mathbf{A}_{N-1}[N-2] & \mathbf{A}_{N-1}[N-3] & \cdots & \mathbf{I}
\end{bmatrix}
\tag{25}
$$

with each block being $L \times L$ and $\mathbf{P}$ is the $NL \times NL$ block diagonal matrix

$$
\mathbf{P} = \mathrm{diag}\left(\mathbf{\Sigma}_0, \mathbf{\Sigma}_1, \cdots, \mathbf{\Sigma}_{N-1}\right).
\tag{26}
$$

6

Here $\mathbf{I}$ is the $L \times L$ identity matrix, $\mathbf{A}_j[i]$ is the ith filter parameter for an AR model of order j, and $\boldsymbol{\Sigma}_j$ is the corresponding excitation noise covariance matrix. This decomposition is valid for any process. However, if the process is an AR process of order p, then we have for $j \geq p$

$$
\begin{aligned}
\mathbf{A}_j[i] &= \begin{cases} \mathbf{A}_p[i] & i = 1, 2, \cdots, p \\ 0 & i > p \end{cases} \\
\boldsymbol{\Sigma}_j &= \boldsymbol{\Sigma}_p.
\end{aligned}
\tag{27}
$$

In order to obtain a statistically stable prewhitening matrix we restrict the order of the maximum AR model to p, where $p \ll N$. The estimation of $\mathbf{R}_x^{-1}$ then reduces to an estimation of an AR model of order p based on $\mathbf{x}[n]$ for $n = 0, 1, \cdots, N-1$. Many multichannel AR methods can automatically provide estimates of all the lower order AR model parameters [1,5]. In the computer simulations of the next section, we use the autocorrelation method which solves the Yule-Walker equations with a biased ACF estimator [1]. From the derivation in Appendix C, it is readily seen, (24) also holds for real data. Of course it is more appropriate to change conjugate transpose $[\cdot]^H$ to transpose $[\cdot]^T$. The same results for real data are also given in [3].

3.2 AR model order selection for $\mathbf{R}_x$

We consider first real AR models. The extension to complex is then given. For the selection of the AR model order, even though an arbitrary WSS random process is representable as an AR process of order $p = \infty$, the usual model order selection rules appear to choose reasonable finite order approximations. The commonly used methods are Akaike's information criterion (AIC) [11] and the minimum description length (MDL) criterion [12]. In this paper we propose to use a method recently described in [8], which is based on exponentially embedded families (EEF). It has been shown to have better performance than AIC and the MDL for AR model order estimation. For real data, it proceeds as follows [8]

$$
\hat{p} = \arg \max_{1 \leq i \leq p_{\max}} \left\{ \left(L_{G_i}(\mathbf{x}) - i \left[\ln \left(\frac{L_{G_i}(\mathbf{x})}{i} \right) + 1 \right] \right) u \left(\frac{L_{G_i}(\mathbf{x})}{i} - 1 \right) \right\}
\tag{28}
$$

where $u(x)$ is the unit step function and $L_{G_i}(\mathbf{x})$ is the generalized log-likelihood ratio for an AR model of order i as compared to a reference model, which is taken to be an AR model of order zero (just WGN). For a real multichannel AR process of order i, the generalized log-likelihood ratio becomes [3,8]

$$
L_{G_i}(\mathbf{x}) = (N-i) \ln \left(\frac{\det \left(\hat{\boldsymbol{\Sigma}}_0 \right)}{\det \left(\hat{\boldsymbol{\Sigma}}_i \right)} \right)
\tag{29}
$$

where $\hat{\boldsymbol{\Sigma}}_i$ and $\hat{\boldsymbol{\Sigma}}_0$ are $L \times L$ conditional ML excitation noise covariance matrix estimators of the AR model of order i and the reference model respectively, given as [3]

$$
\begin{aligned}
\hat{\boldsymbol{\Sigma}}_i &= \frac{1}{N-i} \mathbf{x}_i^T \left(\mathbf{I} - \mathbf{H}_i \left(\mathbf{H}_i^T \mathbf{H}_i \right)^{-1} \mathbf{H}_i^T \right) \mathbf{x}_i
\tag{30} \\
\hat{\boldsymbol{\Sigma}}_0 &= \frac{1}{N-i} \mathbf{x}_i^T \mathbf{x}_i.
\tag{31}
\end{aligned}
$$

Here $\mathbf{I}$ is the $(N-i) \times (N-i)$ identity matrix and

$$
\mathbf{x}_i = \begin{bmatrix} \mathbf{x}[i] & \mathbf{x}[i+1] & \cdots & \mathbf{x}[N-1] \end{bmatrix}^T \quad ((N-i) \times L)
\tag{32}
$$

$$\mathbf{H}_i = \begin{bmatrix} \mathbf{x}^T[i-1] & \mathbf{x}^T[i-2] & \cdots & \mathbf{x}^T[0] \\ \mathbf{x}^T[i] & \mathbf{x}^T[i-1] & \cdots & \mathbf{x}^T[1] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}^T[N-2] & \mathbf{x}^T[N-3] & \cdots & \mathbf{x}^T[N-1-i] \end{bmatrix} \quad ((N-i) \times Li) . \tag{33}$$

For large data records computation of (29) may be a serious load. In this case $\hat{\boldsymbol{\Sigma}}_i$ and $\hat{\boldsymbol{\Sigma}}_0$ can be replaced by estimates from the AR model based methods which provide approximate ML estimates. From the derivation of EEF in [8], it is easy to see the above procedure also works for complex data if we change transpose $[\cdot]^T$ to conjugate transpose $[\cdot]^H$ in (30) and (31).

4 Summary of the New Estimator

One last comment is from the observation that both the periodogram of (8) and the MARMPW spectral estimator of (18) result in estimates of the MPSD matrix with rank 1. In fact, this problem exists in all MPSD estimators involving Fourier-based methods. In order to solve this problem and recover the correct rank of $\mathbf{P}_x(f)$, we can break the data into K blocks and average the spectral estimate of each block. For a full rank $\mathbf{P}_x(f)$, at least $K = L$ blocks should be averaged. In the simulations to follow the AR parameters used to construct $\hat{\mathbf{R}}_x^{-1}$ are estimated from the whole data record. However, the MARMPW spectral estimator of (18) is applied to each block and the results are averaged. Of course $\hat{\mathbf{R}}_x^{-1}$ is now a $N'L \times N'L$ matrix, where N' is the data length of each block.

The MARMPW spectral estimator is now summarized.

1. Choose a maximum AR model order p_{max}.

2. Using the EEF model order estimator of (28) obtain $\hat{p}$.

3. Estimate the AR model parameters for the different order AR models $j = 0, 1, \cdots, \hat{p}$ as $\left\{ \hat{\boldsymbol{\Sigma}}_0 \right\}, \left\{ \hat{\mathbf{A}}_1[1], \hat{\boldsymbol{\Sigma}}_1 \right\}, \cdots, \left\{ \hat{\mathbf{A}}_{\hat{p}}[1], \hat{\mathbf{A}}_{\hat{p}}[2], \cdots, \hat{\mathbf{A}}_{\hat{p}}[\hat{p}], \hat{\boldsymbol{\Sigma}}_{\hat{p}} \right\}$. In the simulations to follow we have used the autocorrelation or Yule-Walker method and so the AR model parameters for all the lower order models are available.

4. Segment the data into K equal length blocks, with $K \geq L$.

5. Construct the $N'L \times N'L$ matrix $\hat{\mathbf{R}}_x^{-1}$ using the estimated parameters (see (24–27)).

6. Compute the MARMPW spectral estimate of (18) for each block and average them to get the final estimate.

5 Computer Simulation Results

In this section we present some computer simulations intended to show the bias, i.e. resolution, and mean square error (MSE) advantages of the proposed spectral estimator. We compare the MARMPW spectral estimator with the following estimators:

AR-YW Autocorrelation or Yule-Walker method, which is a typical AR model-based method, denoted here as AR-YW. The EEF method is used to determine the AR model order.

Periodogram A conventional Fourier-based spectral estimator, given in (8).

MARPW Multichannel ARPW spectral estimator, given in (11). The EEF method is used to determine AR model orders and the AR-YW method is used to estimate the AR model parameters.

MVSE Multichannel minimum variance spectral estimator, given in (15), where p is chosen to be N' and $\hat{\mathbf{R}}_x^{-1}$ is the same $\hat{\mathbf{R}}_x^{-1}$ used in the MARMPW spectral estimator.

Except for the AR-YW, which does not require an estimate of the $NL \times NL$ ACM (only $\hat{\mathbf{R}}_x[0], \hat{\mathbf{R}}_x[1], \cdots, \hat{\mathbf{R}}_x[p]$ with $p \ll N$ are needed), all other methods have the same data segmentation and averaging as in the MARMPW spectral estimator.

5.1 Two Complex Sinusoids

We compare a two-channel process with both channels being two closely spaced *complex* sinusoids. The signal of channel 1 is

$$x_1[n] = \exp\left(j\left(2\pi f_1 n + \varphi_1\right)\right) + \exp\left(j\left(2\pi f_2 n + \varphi_2\right)\right) + w_1[n], \qquad n = 0, 1, \cdots, N-1$$

and the signal of channel 2 is

$$x_2[n] = \exp\left(j\left(2\pi f_1 n + \varphi_1 + \pi/3\right)\right) + \exp\left(j\left(2\pi f_2 n + \varphi_2 + \pi/3\right)\right) + w_2[n], \qquad n = 0, 1, \cdots, N-1$$

where $f_1 = 0.13$, $f_2 = 0.14$, φ_1 and φ_2 are independent uniformly distributed random phases, $w_1[n]$ and $w_2[n]$ are two independent complex white Gaussian noise (WGN) processes with covariance matrix $\mathbf{\Sigma} = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix}$. A data record of $N = 256$ points was segmented into 4 blocks with each block being $N' = 64$, which is insufficient for Fourier-based MPSD estimators to resolve the two sinusoids. For this short data record case EEF is unable to give correct AR order which should be 2, since the data is approximately only two complex sinusoids. However, even with the selected $\hat{p}$ being 1, the MARMPW spectral estimator still gives reasonable results. For the periodogram, in order to have maximum resolution, no data windowing is used. In Figures 1–4, are shown the average of $\hat{P}_{11}(f)$ over 500 realizations of the various MPSD estimators. It can be observed that compared to other estimators the MARMPW and MARPW estimators have the least bias or least "leakage", since the spectral estimates not near the peaks are approximately the true value of $10\log_{10}\sigma_w^2 = -20$ dB. It is also seen that except for the MARMPW estimator, no estimator resolves the two sinusoids. This is the case even for the AR-YW estimator, in which there is no block averaging. The failure of the MARPW estimator to resolve the two sinusoids results from the use of an FIR prewhitening filter, which is affected to a great extent by end effects in this short data length situation. For this process the true value of MSC should be 1 at f_1 and f_2 and 0 elsewhere. In Figures 5, 6 are shown the estimators of MSC, where it can be seen that the MARMPW and MARPW estimators provide the most fidelity. Figures 7, 8 show the average angle of $\hat{P}_{12}(f)$. The MARMPW and MARPW estimators again give the best estimates, which should be $-\pi/3 = -60°$ at f_1 and f_2 and $0°$ elsewhere. The estimates of $\hat{P}_{22}(f)$ are similar to that of $\hat{P}_{11}(f)$ and so they are not shown. Finally, it should be noted that the AR-YW spectral estimator and MVSE produce nearly identical results. This is expected since the MVSE is a weighted average of the AR-YW spectral estimator of various orders [1].

5.2 Real ARMA Process

A *real* autoregressive moving average (ARMA) process was chosen next. The process is given as

$$\mathbf{x}[n] = -\begin{bmatrix} -1.36 & 0.35 \\ 0 & -1.51 \end{bmatrix} \mathbf{x}[n-1] - \begin{bmatrix} 0.960 & 0.35 \\ 0 & 0.960 \end{bmatrix} \mathbf{x}[n-2] \qquad (34)$$

$$+ \mathbf{u}[n] + \begin{bmatrix} -0.262 & -0.650 \\ 0 & -0.867 \end{bmatrix} \mathbf{u}[n-1] + \begin{bmatrix} 0.490 & 0.350 \\ 0 & 0.810 \end{bmatrix} \mathbf{u}[n-2]$$

where $\mathbf{u}[n]$ is a two-channel real white Gaussian noise process with covariance matrix $\boldsymbol{\Sigma}_u = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. A data record of $N = 256$ was segmented into 4 blocks with each block having $N' = 64$ points. In order to reduce the bias, the periodogram in this case utilized a Hamming window for each block. The window has sidelobes about 43 dB down, and is matched to the dynamic range of the true $P_{11}(f)$, which exhibits a dynamic range of about 45 dB as shown in Figure 9 as the thick solid line. In 500 realizations, the EEF chose $\hat{p} = 5$ on the average (with $p_{max} = 15$). In Figures 9–12 are shown the mean of $\hat{P}_{11}(f)$ of the various estimators. It is seen from Figures 9 and 11 that the MARMPW, the MARPW and the windowed periodogram spectral estimators have the least leakage among all the estimators. However, the windowing results in widened peaks in the periodogram and the MARPW estimator exhibits seriously smoothed peaks as seen in Figure 10 and 12 respectively. A comparison of the relative MSE of $\hat{P}_{11}(f)$, which is defined as $\left(\text{MSE of } \hat{\mathbf{P}}_{11}(f) \right) / \mathbf{P}_{11}^2(f)$, is made for the MARMPW and MARPW estimator in Figures 13. It can be seen that the MARMPW estimator has much less relative MSE than the MARPW estimator near the peaks, which are generally of most interest in applications, and has similar relative MSE elsewhere. As to the estimation of MSC, from Figures 14 and 15, the results of the MARMPW and MARPW estimators are similar to each other and better than other estimators. The same conclusion can be made for the angle of $\hat{P}_{12}(f)$ from Figures 16 and Figure 17. It is also found from the simulations that the overall MSE of the estimated MSC by the MARMPW estimator is slightly less than that of the estimated MSC by the MARPW estimator. This conclusion also holds for the angle of $\hat{P}_{12}(f)$. Because $P_{22}(f)$ exhibits less spectral dynamic range, all estimators yield good results, and hence the results are omited. As before, it should be noted that the AR-YW spectral estimator and MVSE produce similar results.

6   Conclusions

The multichannel extension of the ARMPW spectral estimator has been described. Based on a matrix prewhitener it has a higher resolution than conventional Fourier-based and AR model-based PSD estimators and has about the same variance as a periodogram. It is also more accurate than spectral estimators based on a multichannel FIR prewhitening filter, due to the amelioration of end efforts by the matrix prewhitener. In order to construct this matrix prewhitener, the Cholesky decomposition of the inverse autocorrelation matrix of a multichannel AR process has been derived and the AR model order selection using the EEF method has been applied. Because of its excellent properties, the MARMPW spectral estimator is a valuable method for multichannel processes with high dynamic spectral range.

A   ML estimation of $\mathbf{C}_A$

Lemma:

Let $\mathbf{x}$ be an $L \times 1$ complex vector with $\|\mathbf{x}\| > 1$, where $\|\cdot\|$ denotes Euclidean norm, then minimization of $J(\mathbf{C}) = \ln(\det(\mathbf{I} + \mathbf{C})) - \mathbf{x}^H \mathbf{C}(\mathbf{I} + \mathbf{C})^{-1}\mathbf{x}$ over all positive semidefinite $L \times L$ matrices $\mathbf{C}$ yields

$$ J_{\min} = -\left(\mathbf{x}^H \mathbf{x} - \ln\left(\mathbf{x}^H \mathbf{x} \right) - 1 \right) $$

and is achieved uniquely for $\mathbf{C} = \hat{\mathbf{C}} = \mathbf{x}\mathbf{x}^H - \frac{\mathbf{x}\mathbf{x}^H}{\|\mathbf{x}\|^2}$. (If $\|\mathbf{x}\| \leq 1$, then the ML estimator of $\mathbf{C}$ gives the nonuseful result $\hat{\mathbf{C}} = \mathbf{0}$.)

Proof: Consider

$$ J(\mathbf{C}) - J_{\min} = \ln(\det(\mathbf{I} + \mathbf{C})) - \mathbf{x}^H \mathbf{C}(\mathbf{I} + \mathbf{C})^{-1}\mathbf{x} + \mathbf{x}^H \mathbf{x} - \ln\left(\mathbf{x}^H \mathbf{x} \right) - 1. \tag{35} $$

Let $\mathbf{C} = \sum_{i=1}^{L} \lambda_i \mathbf{v}_i \mathbf{v}_i^H = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$, where $\mathbf{V}$ is a unitary matrix, $\lambda_i \geq 0$ for all $1 \leq i \leq L$ and $\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \lambda_2, \cdots, \lambda_L\right)$, then

$$\mathbf{x}^H \mathbf{x} - \mathbf{x}^H \mathbf{C} \left(\mathbf{I} + \mathbf{C}\right)^{-1} \mathbf{x} = \mathbf{x}^H \mathbf{V}\mathbf{V}^H \mathbf{x} - \mathbf{x}^H \mathbf{V}\mathbf{V}^H \mathbf{C} \left(\mathbf{I} + \mathbf{C}\right)^{-1} \mathbf{V}\mathbf{V}^H \mathbf{x}. \tag{36}$$

Let $\mathbf{u} = \mathbf{V}^H \mathbf{x}$ and since $\mathbf{V}^{-1} = \mathbf{V}^H$, the right hand side of (36) can be rewritten as

$$
\begin{aligned}
\mathbf{u}^H \mathbf{u} - \mathbf{u}^H \left[\mathbf{V}^H \mathbf{C} \left(\mathbf{I} + \mathbf{C}\right)^{-1} \mathbf{V}\right] \mathbf{u} &= \mathbf{u}^H \mathbf{u} - \mathbf{u}^H \left[\mathbf{V}^H \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \left(\mathbf{I} + \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H\right)^{-1} \mathbf{V}\right] \mathbf{u} \\
&= \mathbf{u}^H \mathbf{u} - \mathbf{u}^H \mathbf{\Lambda} \left[\mathbf{V}^{-1} \left(\mathbf{I} + \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H\right) \mathbf{V}^{-H}\right]^{-1} \mathbf{u} \\
&= \mathbf{u}^H \mathbf{u} - \mathbf{u}^H \mathbf{\Lambda} \left[\mathbf{I} + \mathbf{\Lambda}\right]^{-1} \mathbf{u} \\
&= \sum_{i=1}^{L} \left(|u_i|^2 - \frac{|u_i|^2 \lambda_i}{1 + \lambda_i}\right) \\
&= \sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right).
\end{aligned}
\tag{37}
$$

Also,

$$
\begin{aligned}
\ln\left(\det\left(\mathbf{I} + \mathbf{C}\right)\right) - \ln\left(\mathbf{x}^H \mathbf{x}\right) &= \ln\left(\prod_{i=1}^{L} \left(1 + \lambda_i\right)\right) - \ln\left(\sum_{i=1}^{L} |u_i|^2\right) \\
&= \sum_{i=1}^{L} \ln\left(1 + \lambda_i\right) - \ln\left(\sum_{i=1}^{L} |u_i|^2\right).
\end{aligned}
\tag{38}
$$

Plugging (37) and (38) into (35), we have

$$
\begin{aligned}
J\left(\mathbf{C}\right) - J_{\min} &= \sum_{i=1}^{L} \ln\left(1 + \lambda_i\right) - \ln\left(\sum_{i=1}^{L} |u_i|^2\right) + \sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right) - 1 \\
&\quad + \ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right)\right) - \ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right)\right) \\
&= \sum_{i=1}^{L} \ln\left(1 + \lambda_i\right) + \ln\left(\frac{\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right)}{\sum_{i=1}^{L} |u_i|^2}\right) + \sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right) - \ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right)\right) - 1 \\
&= \sum_{i=1}^{L} \ln\left(1 + \lambda_i\right) + \ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \frac{1}{1 + \lambda_i}\right)\right) + \sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right) - \ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right)\right) - 1.
\end{aligned}
$$

Let $S_1 = \sum_{i=1}^{L} \ln\left(1 + \lambda_i\right) + \ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \frac{1}{1 + \lambda_i}\right)\right)$ and $S_2 = \sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right) - \ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1 + \lambda_i}\right)\right) - 1$.

Since $\sum_{i=1}^{L} \left(\frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2}\right) = 1$ and $\ln\left(\cdot\right)$ is a concave function, by Jensen's inequality we have

$$\ln\left(\sum_{i=1}^{L} \left(\frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \frac{1}{1 + \lambda_i}\right)\right) \geq \sum_{i=1}^{L} \left(\frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \ln \frac{1}{1 + \lambda_i}\right)$$

11

with equality if and only if $|u_i|^2 = 0$ for all $i \neq k$. Hence we have

$$
\begin{aligned}
S_1 &\geq \sum_{i=1}^{L} \ln\left(1 + \lambda_i\right) + \sum_{i=1}^{L} \left(\frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \ln \frac{1}{1 + \lambda_i} \right) \\
&= \sum_{i=1}^{L} \left(1 - \frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \right) \ln\left(1 + \lambda_i\right) \\
&\geq 0
\end{aligned}
$$

with equality if and only if $|u_i|^2 = 0$ for all $i \neq k$ and $\left(1 - \frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \right) \ln\left(1 + \lambda_i\right) = 0$ for all i. As to S_2, since $a - \ln a - 1 \geq 0$ for all $a > 0$, with equality if and only if $a = 1$, we have $S_2 \geq 0$ with equality if and only if $\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1+\lambda_i} \right) = 1$. Hence we have

$$
J\left(\mathbf{C}\right) - J_{\min} = S_1 + S_2 \geq 0
$$

with equality if and only if $S_1 = 0$ and $S_2 = 0$. In order to have $S_1 = 0$, we need to have $|u_i|^2 = 0$ for all $i \neq k$ and $\left(1 - \frac{|u_i|^2}{\sum_{j=1}^{L} |u_j|^2} \right) \ln\left(1 + \lambda_i\right) = 0$ for all i. If the first condition holds, the second condition is $\lambda_i = 0$ for $i \neq k$. Thus,

$$
\hat{\mathbf{C}} = \lambda_k \mathbf{v}_k \mathbf{v}_k^H \tag{39}
$$

also, since $u_i = 0$ for $i \neq k$, we have that $u_i = \mathbf{v}_i^H \mathbf{x} = 0$ for $i \neq k$. It follows that

$$
\mathbf{v}_k = \frac{\mathbf{x}}{\|\mathbf{x}\|}. \tag{40}
$$

For $S_2 = 0$, we need $\sum_{i=1}^{L} \left(\frac{|u_i|^2}{1+\lambda_i} \right) = \frac{|u_k|^2}{1+\lambda_k} = 1$, so that

$$
\lambda_k = |u_k|^2 - 1. \tag{41}
$$

Replacing u_k with $\mathbf{v}_k^H \mathbf{x}$ and plugging (40) into (41) yields

$$
\begin{aligned}
\lambda_k &= \left|\mathbf{v}_k^H \mathbf{x}\right|^2 - 1 \\
&= \mathbf{x}^H \mathbf{x} - 1. \tag{42}
\end{aligned}
$$

Since it is assumed that $\|\mathbf{x}\| > 1$, we have $\lambda_k > 0$ as required. Substituting (40) and (42) into (39) produces the final result

$$
\begin{aligned}
\hat{\mathbf{C}} &= \left(\mathbf{x}^H \mathbf{x} - 1\right) \frac{\mathbf{x}\mathbf{x}^H}{\|\mathbf{x}\|^2} \\
&= \mathbf{x}\mathbf{x}^H - \frac{\mathbf{x}\mathbf{x}^H}{\|\mathbf{x}\|^2}.
\end{aligned}
$$

Next, to derive the ML estimator we note that, the data stucture of (12) can be rewritten in vector form as

$$\mathbf{x} = \mathbf{E}_0 \mathbf{A}_c + \mathbf{w}. \tag{43}$$

It is now assumed that $\mathbf{A}_c$ is an $L \times 1$ complex Gaussian random vector with zero mean and is independent of the noise $\mathbf{w}$. We want to find the ML estimate of $\mathbf{C}_A$, the covariance matrix of $\mathbf{A}_c$, which is assumed to be positive semidefinite. To simplify the calculation, let $\mathbf{y} = \mathbf{R}_w^{-\frac{1}{2}} \mathbf{x}$ and denote $\mathbf{E}_0$ by $\mathbf{E}$. We then have from (43)

$$\mathbf{y} = \mathbf{R}_w^{-\frac{1}{2}} \mathbf{E} \mathbf{A}_c + \mathbf{R}_w^{-\frac{1}{2}} \mathbf{w}.$$

Let $\mathbf{H} = \mathbf{R}_w^{-\frac{1}{2}} \mathbf{E}$ ($NL \times L$) and $\zeta = \mathbf{R}_w^{-\frac{1}{2}} \mathbf{w}$. Since $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_w)$, where $\sim \mathcal{CN}$ denotes "is distributed according to a complex Gaussian distribution" [14], as a result we have the covariance matrix of ζ

$$
\begin{aligned}
\mathbf{C}_\zeta &= \mathbf{R}_w^{-\frac{1}{2}} \mathbf{R}_w \mathbf{R}_w^{-\frac{H}{2}} \\
&= \mathbf{I}.
\end{aligned}
$$

Hence the covariance matrix of $\mathbf{y}$ is

$$\mathbf{C}_y = \mathbf{H} \mathbf{C}_A \mathbf{H}^H + \mathbf{I}.$$

Then the probability density function of the random vector $\mathbf{y}$ is

$$p_y(\mathbf{y}) = \frac{1}{\pi^{NL} \det(\mathbf{C}_y)} \exp\left(-\mathbf{y}^H \mathbf{C}_y^{-1} \mathbf{y}\right).$$

The ML estimator of $\mathbf{C}_A$ is given by minimizing [14]

$$\ln(\det(\mathbf{C}_y)) + \mathbf{y}^H \mathbf{C}_y^{-1} \mathbf{y} \tag{44}$$

over $\mathbf{C}_A$.

Using the identity

$$\det(\mathbf{I}_{MM} + \mathbf{A}_{MN} \mathbf{B}_{NM}) = \det(\mathbf{I}_{NN} + \mathbf{B}_{NM} \mathbf{A}_{MN})$$

where $\mathbf{A}_{MN}$ denotes the $M \times N$ matrix $\mathbf{A}$, we have that

$$
\begin{aligned}
\det(\mathbf{C}_y) &= \det\left(\mathbf{H} \mathbf{C}_A \mathbf{H}^H + \mathbf{I}_{NL}\right) \\
&= \det\left(\mathbf{C}_A \mathbf{H}^H \mathbf{H} + \mathbf{I}_L\right) \\
&= \det\left(\mathbf{C}_A \mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} + \mathbf{I}_L\right) \\
&= \det\left(\left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E}\right)^{\frac{1}{2}}\right) \det\left(\mathbf{C}_A \mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} + \mathbf{I}_L\right) \det\left(\left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E}\right)^{-\frac{1}{2}}\right) \\
&= \det\left(\left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E}\right)^{\frac{1}{2}} \mathbf{C}_A \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E}\right)^{\frac{1}{2}} + \mathbf{I}_L\right).
\end{aligned}
$$

Let $\mathbf{C} = \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E}\right)^{\frac{1}{2}} \mathbf{C}_A \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E}\right)^{\frac{1}{2}}$, then

$$\det(\mathbf{C}_y) = \det(\mathbf{C} + \mathbf{I}_L). \tag{45}$$

13

Using the matrix inversion lemma, we have

$$
\begin{aligned}
\mathbf{y}^H \mathbf{C}_y^{-1} \mathbf{y} &= \mathbf{y}^H \left(\mathbf{I} + \mathbf{H} \mathbf{C}_A \mathbf{H}^H \right) \mathbf{y} \\
&= \mathbf{y}^H \left(\mathbf{I} - \mathbf{H} \mathbf{C}_A \left(\mathbf{H}^H \mathbf{H} \mathbf{C}_A + \mathbf{I} \right)^{-1} \mathbf{H}^H \right) \mathbf{y}.
\end{aligned}
$$

Omitting the term $\mathbf{y}^H \mathbf{y}$, which is not dependent on $\mathbf{C}_A$, we have

$$
\begin{aligned}
&-\mathbf{y}^H \mathbf{H} \mathbf{C}_A \left(\mathbf{H}^H \mathbf{H} \mathbf{C}_A + \mathbf{I} \right)^{-1} \mathbf{H}^H \mathbf{y} \\
=\ &-\mathbf{y}^H \mathbf{H} \mathbf{C}_A \left(\left(\mathbf{H}^H \mathbf{H} \right)^{\frac{1}{2}} \left(\left(\mathbf{H}^H \mathbf{H} \right)^{\frac{1}{2}} \mathbf{C}_A \left(\mathbf{H}^H \mathbf{H} \right)^{\frac{1}{2}} + \mathbf{I} \right) \left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \right)^{-1} \mathbf{H}^H \mathbf{y}.
\end{aligned}
$$

But $\mathbf{C} = \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{\frac{1}{2}} \mathbf{C}_A \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{\frac{1}{2}} = \left(\mathbf{H}^H \mathbf{H} \right)^{\frac{1}{2}} \mathbf{C}_A \left(\mathbf{H}^H \mathbf{H} \right)^{\frac{1}{2}}$, so that

$$
\begin{aligned}
&-\mathbf{y}^H \mathbf{H} \mathbf{C}_A \left(\left(\mathbf{H}^H \mathbf{H} \right)^{\frac{1}{2}} \left(\mathbf{C} + \mathbf{I} \right) \left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \right)^{-1} \mathbf{H}^H \mathbf{y} \\
=\ &-\mathbf{y}^H \mathbf{H} \mathbf{C}_A \left(\mathbf{H}^H \mathbf{H} \right)^{\frac{1}{2}} \left(\mathbf{C} + \mathbf{I} \right)^{-1} \left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \mathbf{H}^H \mathbf{y} \\
=\ &-\mathbf{y}^H \mathbf{H} \left(\left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \mathbf{C} \right) \left(\mathbf{C} + \mathbf{I} \right)^{-1} \left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \mathbf{H}^H \mathbf{y} \\
=\ &- \left(\left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \mathbf{H}^H \mathbf{y} \right)^H \mathbf{C} \left(\mathbf{C} + \mathbf{I} \right)^{-1} \left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \mathbf{H}^H \mathbf{y}.
\end{aligned}
$$

Letting $\mathbf{q} = \left(\mathbf{H}^H \mathbf{H} \right)^{-\frac{1}{2}} \mathbf{H}^H \mathbf{y} = \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{-\frac{1}{2}} \mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{x}$, the above expression is equivalent to

$$
-\mathbf{q}^H \mathbf{C} (\mathbf{I}_L + \mathbf{C})^{-1} \mathbf{q}.
$$

Plugging this and (45) into (44), the ML estimator of $\mathbf{C}_A$ is obtained by minimizing

$$
\ln \left(\det \left(\mathbf{I}_L + \mathbf{C} \right) \right) - \mathbf{q}^H \mathbf{C} \left(\mathbf{I}_L + \mathbf{C} \right)^{-1} \mathbf{q}
$$

where $\mathbf{C} = \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{\frac{1}{2}} \mathbf{C}_A \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{\frac{1}{2}}$ and $\mathbf{q} = \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{-\frac{1}{2}} \mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{x}$. From the lemma we have

$$
\hat{\mathbf{C}} = \mathbf{q} \mathbf{q}^H - \frac{\mathbf{q} \mathbf{q}^H}{\|\mathbf{q}\|^2}.
$$

After some algebra, we have

$$
\hat{\mathbf{C}}_{\mathbf{A}} = \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{-1} \mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{x} \mathbf{x}^H \mathbf{R}_w^{-1} \mathbf{E} \left(\mathbf{E}^H \mathbf{R}_w^{-1} \mathbf{E} \right)^{-1} \left(1 - \frac{1}{\|\mathbf{q}\|^2} \right). \tag{46}
$$

Note that $\hat{\mathbf{C}}_{\mathbf{A}}$ is rank one and hence is positive semidefinite.

B Asymptotic Variance of MARMPW Estimator

As already described in Section 2.2, as $N \rightarrow \infty$, we can replace $\hat{\mathbf{R}}_x^{-1}$ by $\mathbf{R}_x^{-1}$. Let $\mathbf{y} = \sqrt{N} \left(\mathbf{E}^H (f) \mathbf{R}_x^{-1} \mathbf{E} (f) \right)^{-1} \mathbf{E}^H (f) \mathbf{R}_x^{-1} \mathbf{x}$ then the MARMPW estimator in (20) can be rewritten as

$$
\hat{\mathbf{P}}_x (f) = \mathbf{y} \mathbf{y}^H. \tag{47}
$$

Therefore the covariance of $\hat{P}_{ij}(f)$ and $\hat{P}_{kl}(f)$ is

$$\text{cov}\left(\hat{P}_{ij}(f),\hat{P}_{kl}(f)\right) = \text{cov}\left(y_i y_j^*, y_k y_l^*\right) \tag{48}$$

where y_i is the ith element of $\mathbf{y}$.

Since $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_x)$, as a result we have that

$$\mathbf{y} \sim \mathcal{CN}\left(0, N\left(\mathbf{E}^H(f)\mathbf{R}_x^{-1}\mathbf{E}(f)\right)^{-1}\right).$$

Denoting $N\left(\mathbf{E}^H(f)\mathbf{R}_x^{-1}\mathbf{E}(f)\right)^{-1}$ by $\mathbf{C}_y$, from the discussion in Section 2.2, as $N \to \infty$, $\mathbf{C}_y \to \mathbf{P}_x(f)$. It also follows from $E[x_1^* x_2 x_3^* x_4] = E[x_1^* x_2] E[x_3^* x_4] + E[x_1^* x_4] E[x_2^* x_3]$ for $\mathbf{x} = [x_1, x_2, x_3, x_4]^T$ a zero mean complex Gaussian random vector [14] that (48) can be evaluated as

$$
\begin{aligned}
\text{cov}\left(\hat{P}_{ij}(f),\hat{P}_{kl}(f)\right) &= E\left[y_i y_j^* y_k^* y_l\right] - E\left[y_i y_j^*\right] E\left[y_k^* y_l\right] \\
&= E\left[y_i y_j^*\right] E\left[y_k^* y_l\right] + E\left[y_l y_j^*\right] E\left[y_i y_k^*\right] - E\left[y_i y_j^*\right] E\left[y_k^* y_l\right] \\
&= E\left[y_i y_k^*\right] E\left[y_l y_j^*\right] \\
&= [\mathbf{C}_y]_{i,k} [\mathbf{C}_y]_{l,j} \\
&\to P_{ik}(f) P_{lj}(f).
\end{aligned}
\tag{49}
$$

As a special case, we have for $k = i$ and $l = j$

$$\text{var}\left(\hat{P}_{ij}(f)\right) \to P_{ii}(f) P_{jj}(f).$$

C The Cholesky Decomposition of the Inverse Autocorrelation Matrix

Given a multichannel random process $\mathbf{x}[n]$ for $0 \leq n \leq N-1$, we can use linear prediction to obtain the parameters $\mathbf{A}_j[1], \mathbf{A}_j[2], \cdots, \mathbf{A}_j[j]$ and the corresponding excitation noise covariance matrix $\mathbf{\Sigma}_j$ by minimizing the one-step prediction error power or equivalently by solving the Yule-Walker equations [1]. Let $\mathbf{u}[0] = \mathbf{x}[0]$ and $\mathbf{u}[n] = \mathbf{x}[n] - \sum_{i=1}^{n}(-\mathbf{A}_n[i]\mathbf{x}[n-i])$ for $1 \leq n \leq N-1$ or

$$
\mathbf{u} = \begin{bmatrix} \mathbf{u}[0] \\ \mathbf{u}[1] \\ \mathbf{u}[2] \\ \vdots \\ \mathbf{u}[N-1] \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_1[1] & \mathbf{I} & \cdots & \mathbf{0} \\ \mathbf{A}_2[2] & \mathbf{A}_2[1] & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{N-1}[N-1] & \mathbf{A}_{N-1}[N-2] & \cdots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}[0] \\ \mathbf{x}[1] \\ \mathbf{x}[2] \\ \vdots \\ \mathbf{x}[N-1] \end{bmatrix}. \tag{50}
$$

Then from the linear vector space viewpoint of Yule-Walker equations, $\mathbf{u}[n]$ for $0 \leq n \leq N-1$ are uncorrelated or in other words the ACM of $\mathbf{u}$ is block diagonal [1]. Thus, we have

$$
\begin{aligned}
\mathbf{R}_u &= E\left[\mathbf{u}\mathbf{u}^H\right] \\
&= \begin{bmatrix} \mathbf{R}_u[0] & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_u[1] & \vdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}_u[N-1] \end{bmatrix}
\end{aligned}
$$

15

$$= \begin{bmatrix} \boldsymbol{\Sigma}_0 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_1 & \vdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{N-1} \end{bmatrix}. \tag{51}$$

Denoting

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_1[1] & \mathbf{I} & \cdots & \mathbf{0} \\ \mathbf{A}_2[2] & \mathbf{A}_2[1] & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{N-1}[N-1] & \mathbf{A}_{N-1}[N-2] & \cdots & \mathbf{I} \end{bmatrix}$$

by $\boldsymbol{\Phi}$ and plugging (50) to (51) we have

$$\begin{aligned} \mathbf{R}_u &= E\left[\boldsymbol{\Phi}\mathbf{x}\mathbf{x}^H\boldsymbol{\Phi}^H\right] \\ &= \boldsymbol{\Phi}E\left[\mathbf{x}\mathbf{x}^H\right]\boldsymbol{\Phi}^H \\ &= \boldsymbol{\Phi}\mathbf{R}_x\boldsymbol{\Phi}^H. \end{aligned}$$

It follows that

$$\mathbf{R}_x^{-1} = \boldsymbol{\Phi}^H\mathbf{R}_u^{-1}\boldsymbol{\Phi}.$$

References

[1] S. M. Kay, *Modern Spectral Estimation Theory and Application.* Englewood Cliffs, NJ: Prentice-Hall, 1988.

[2] Gwilym M. Jenkins and Donald G. Watts, *Spectral Analysis and its Applications.* Holden-Day, 1968.

[3] Gregory C. Reinsel, *Elements of Multivariate Time Series Analysis.* Springer-Verlag,1993.

[4] S. M. Kay and Cuichun Xu, "A wide dynamic range spectral estimator," Submitted to *IEEE Trans. Signal Processing,* 2005.

[5] Taiknang ning and C. L. Nikias, "Multichannel AR spectrum estimation: the optimum approach in the reflection coefficient domain," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-34, no. 5, pp. 1139–1152, Oct. 1986.

[6] Thomson, D.J., "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE,* Vol. 70, pp. 1055–1096, 1982.

[7] R.J. Serfling, *Approximation Theorems of Mathematical Statistics.* J.Wiley, 1980, New York.

[8] S. M. Kay, "Exponentially embedded families: A new approach to model order estimation", *IEEE Trans. Aero. and Elect.,* vol. 41, pp. 333–345, Jan. 2005.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory.* Prentice-Hall, Englewood Cliffs, NJ, 1998.

[10] W.A.Fuller, *Introduction to Statistical Time Series.* J. Wiley, 1996, New York.

[11] H. Akaike, "A new look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.,* vol. AC-19, pp. 716–723, 1974.

[12] J. Rissanen, "Modeling by shortest data description," *Automatica,* vol. 14, pp. 465–478, 1978.

[13] J. Benesty, J. Chen, and Y. Huang, "A Generalized MVDR Spectrum," *IEEE Signal Proc. Letters.* vol. 12, no. 12, pp 827–830, Dec. 2005.

[14] S. M. Kay, *Fundamentals of Statistical Signal Processing : Estimation Theory.* Prentice-Hall, Englewood Cliffs, NJ, 1993.
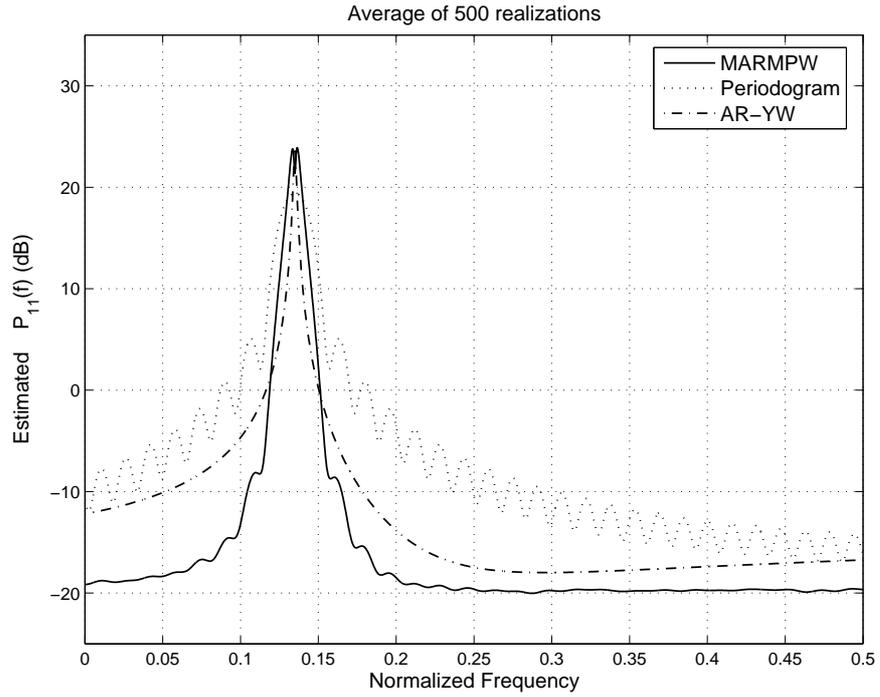
Figure 1: Comparison of $\hat{P}_{11}(f)$ of MARMPW to periodogram and AR-YW for two sinusoids in WGN with frequencies $f_1 = 0.13$ and $f_2 = 0.14$.
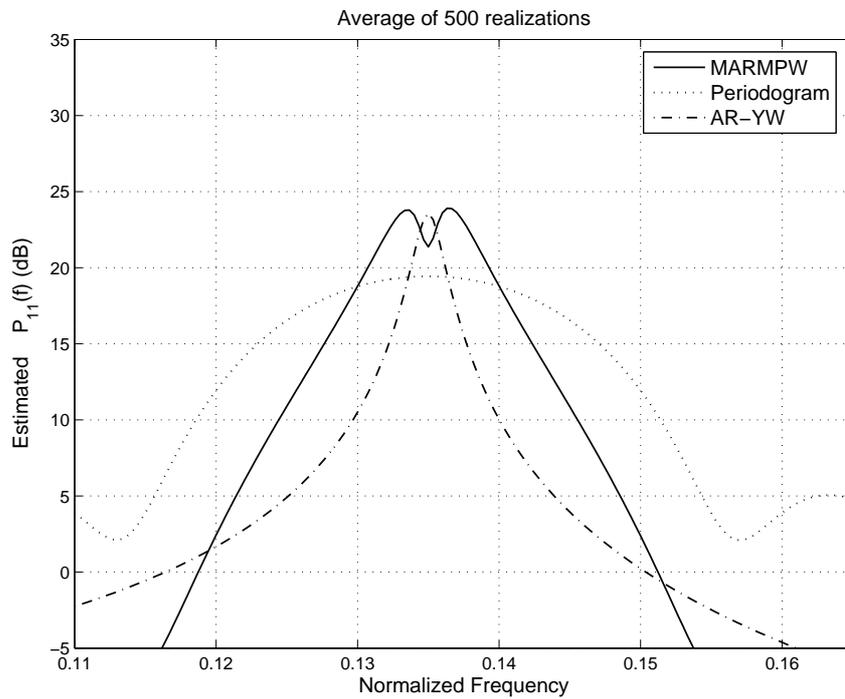
Figure 2: Expanded version of Figure 1.

18
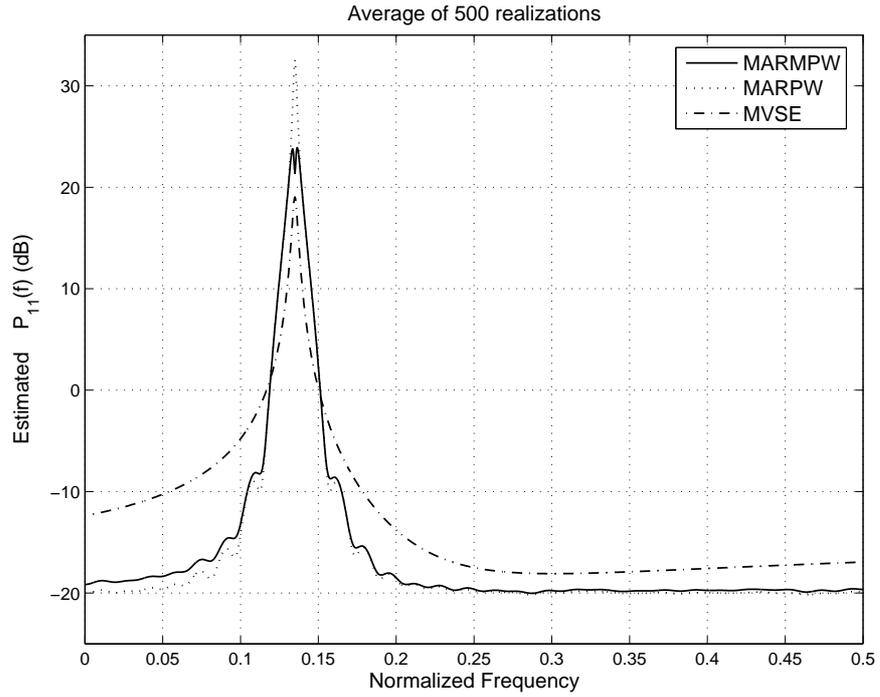
Figure 3: Comparison of $\hat{P}_{11}(f)$ of MARMPW to MARPW and MVSE for two sinusoids in WGN with frequencies $f_1 = 0.13$ and $f_2 = 0.14$.
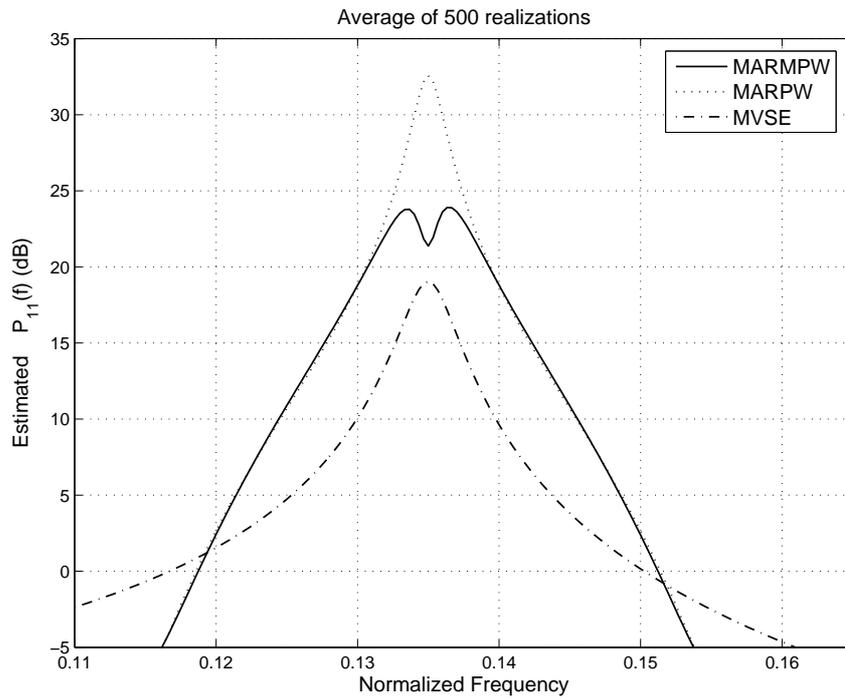
Figure 4: Expanded version of Figure 3.
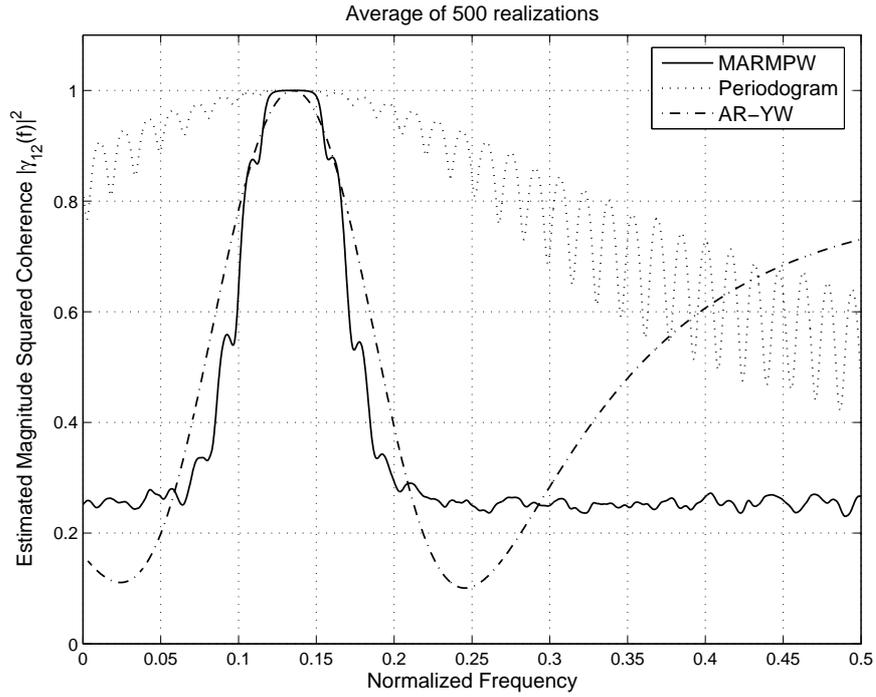
Figure 5: Comparison of MSC estimates of MARMPW to periodogram and AR-YW for two sinusoids in WGN with frequencies $f_1 = 0.13$ and $f_2 = 0.14$.
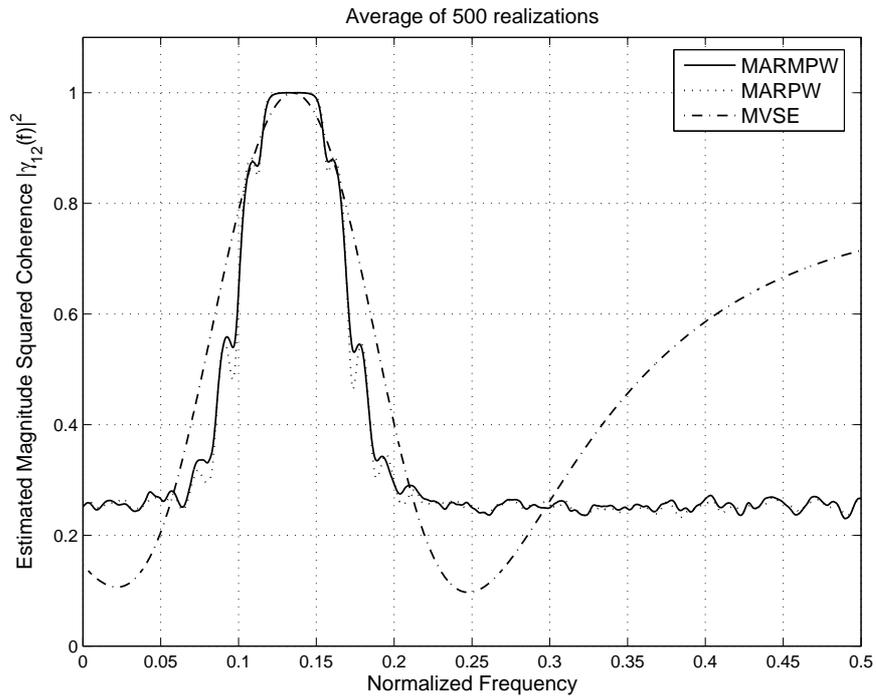
Figure 6: Comparison of MSC estimates of MARMPW to MARPW and MVSE for two sinusoids in WGN with frequencies $f_1 = 0.13$ and $f_2 = 0.14$.
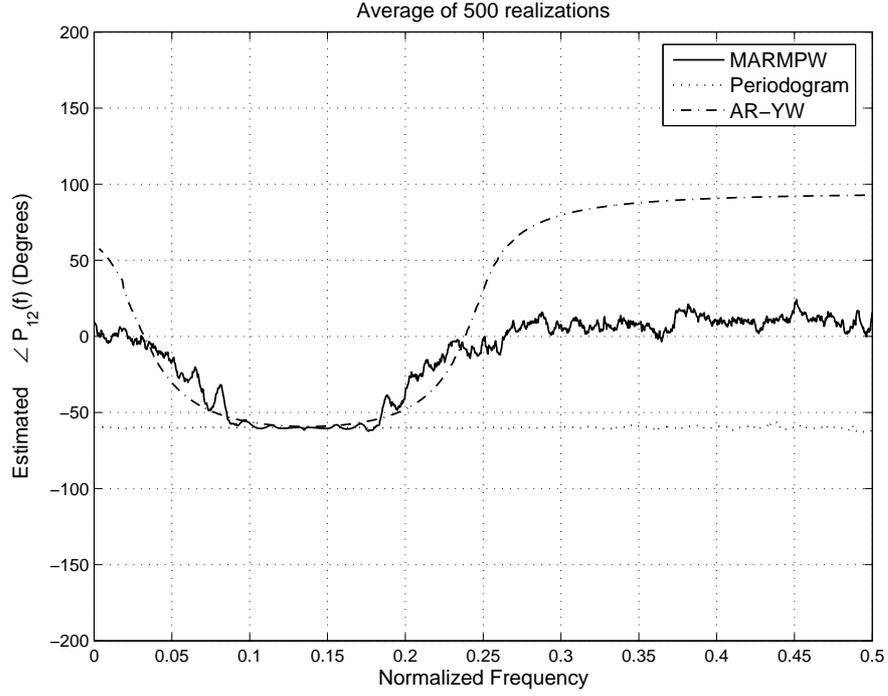
Figure 7: Comparison of $\angle \hat{P}_{12}(f)$ of MARMPW to periodogram and AR-YW for two sinusoids in WGN with frequencies $f_1 = 0.13$ and $f_2 = 0.14$.
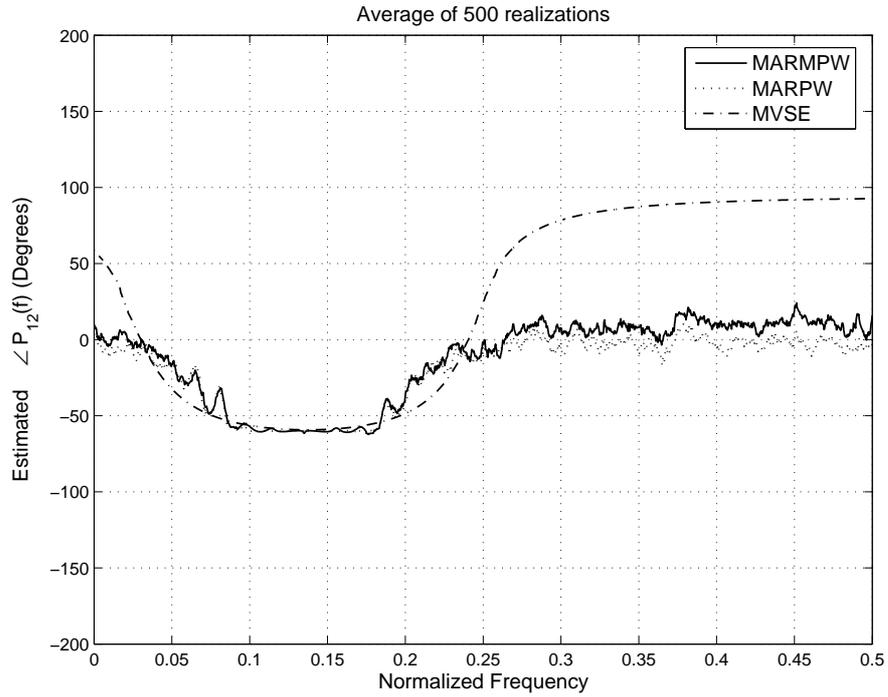
Figure 8: Comparison of $\angle \hat{P}_{12}(f)$ of MARMPW to MARPW and MVSE for two sinusoids in WGN with frequencies $f_1 = 0.13$ and $f_2 = 0.14$.
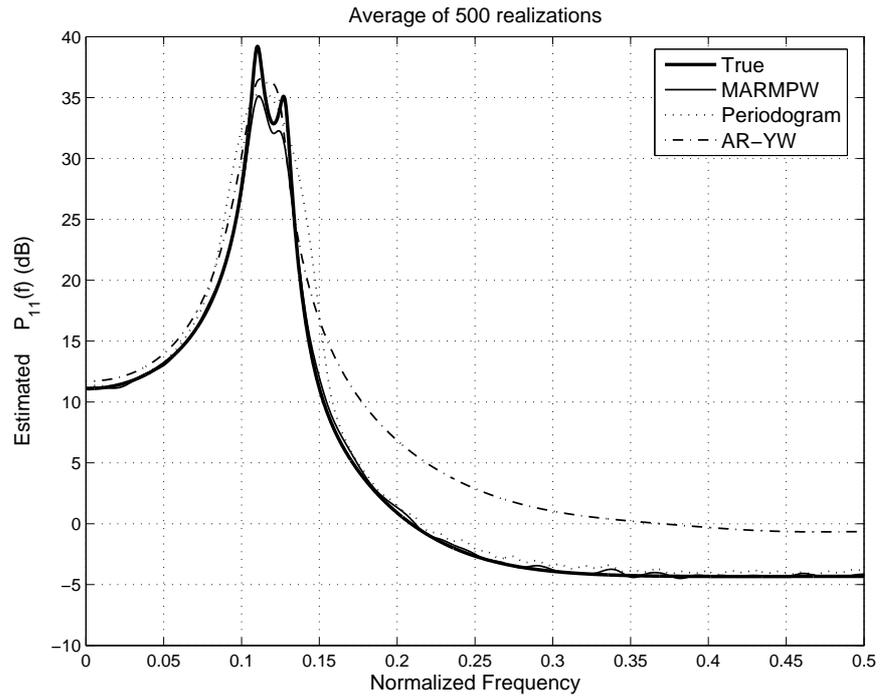
Figure 9: Comparison of $\hat{P}_{11}(f)$ of MARMPW to periodogram and AR-YW for an ARMA process.
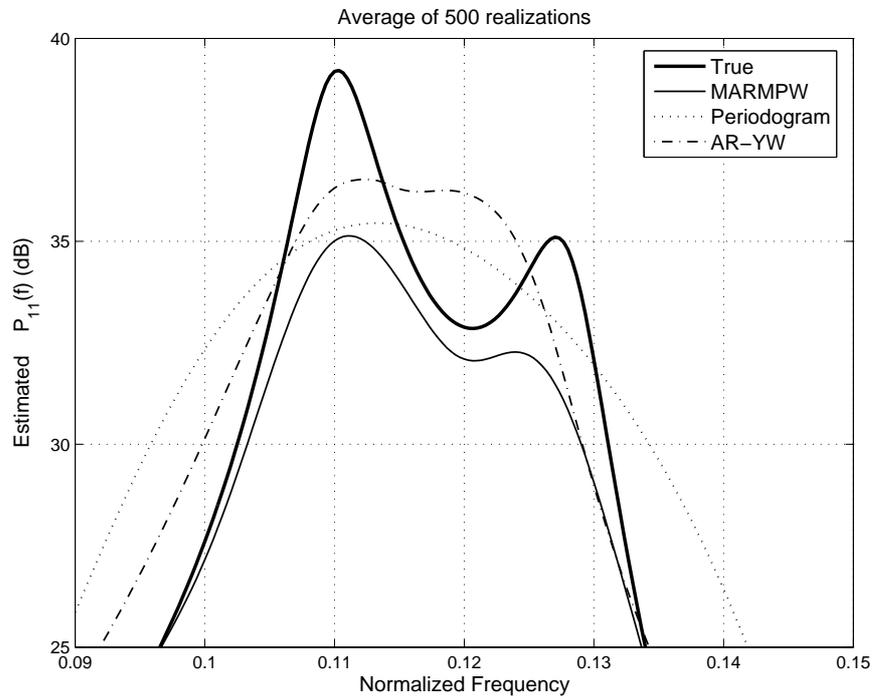
Figure 10: Expanded version of Figure 9.

22
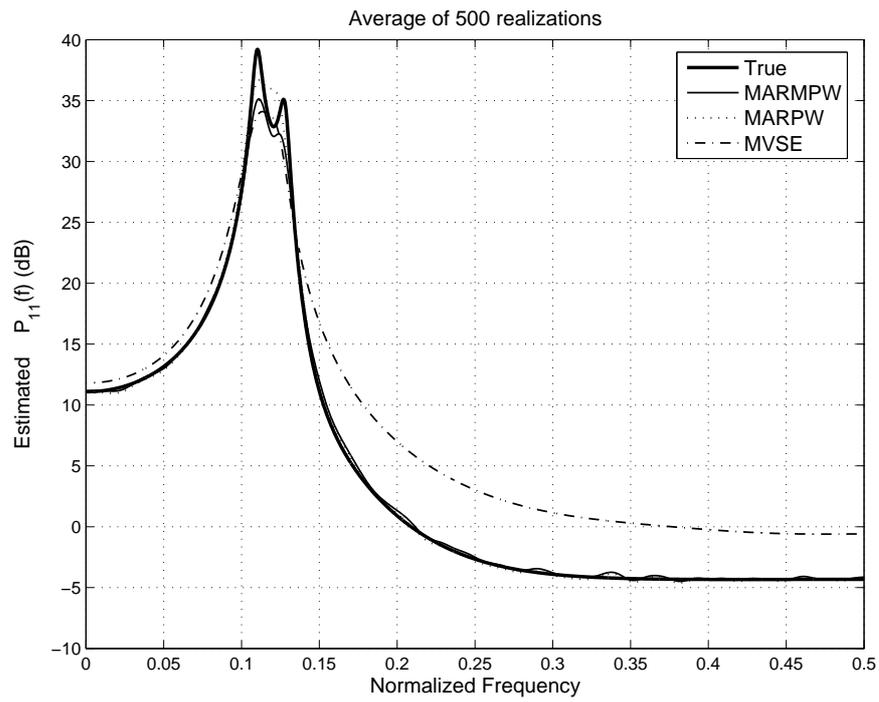
Figure 11: Comparison of $\hat{P}_{11}(f)$ of MARMPW to MARPW and MVSE for an ARMA process.
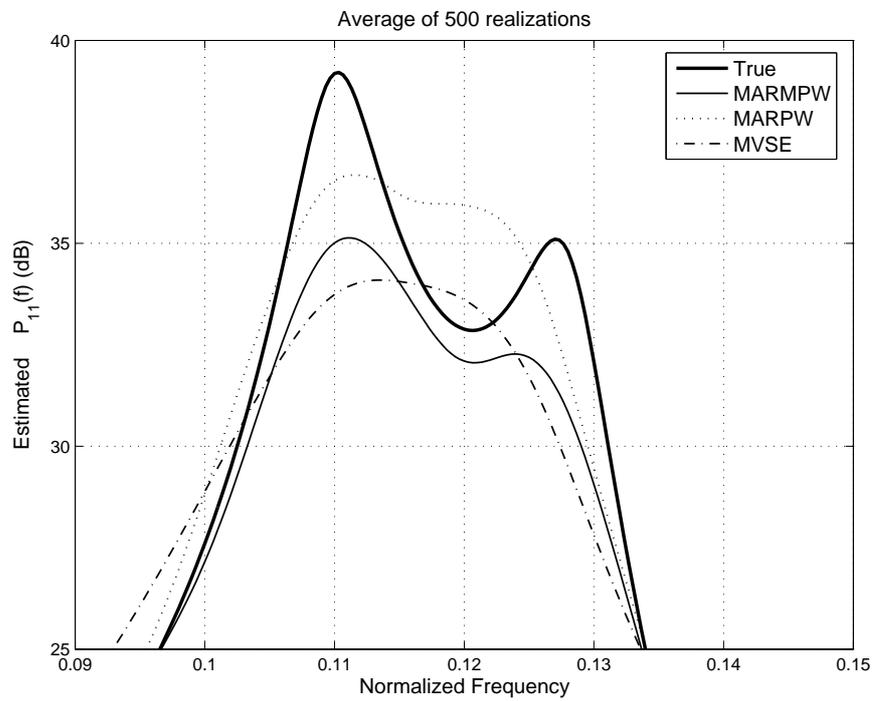
Figure 12: Expanded version of Figure 11.

23
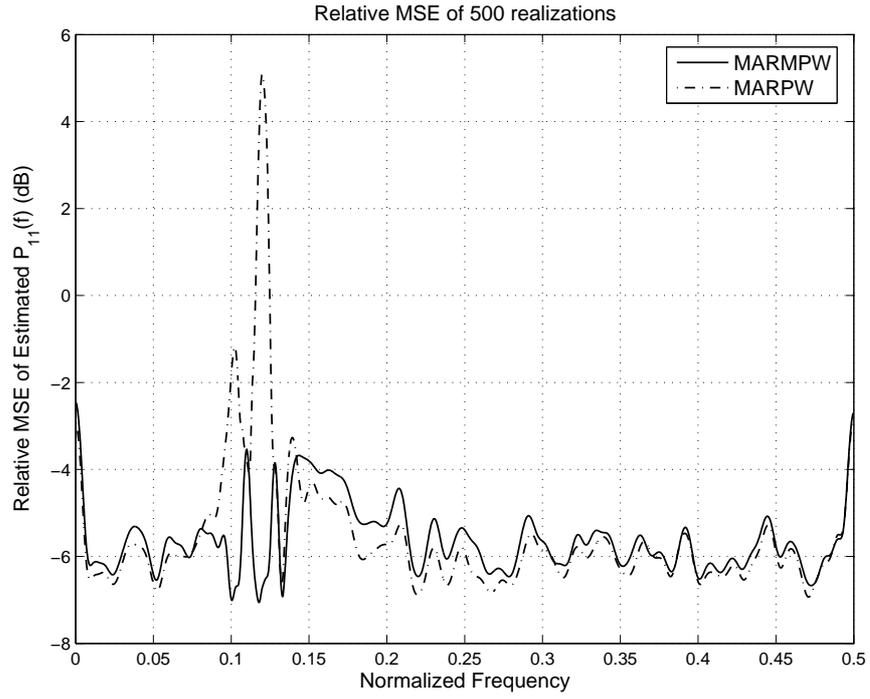
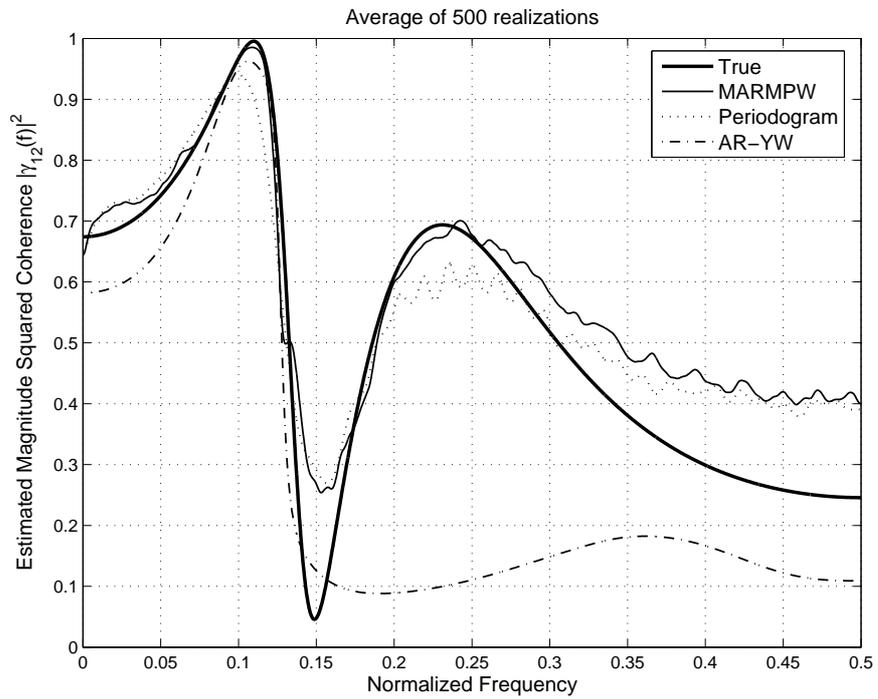Figure 13: Comparison of the relative MSE of $\hat{P}_{11}(f)$ for MARMPW and MARPW spectral estimators.

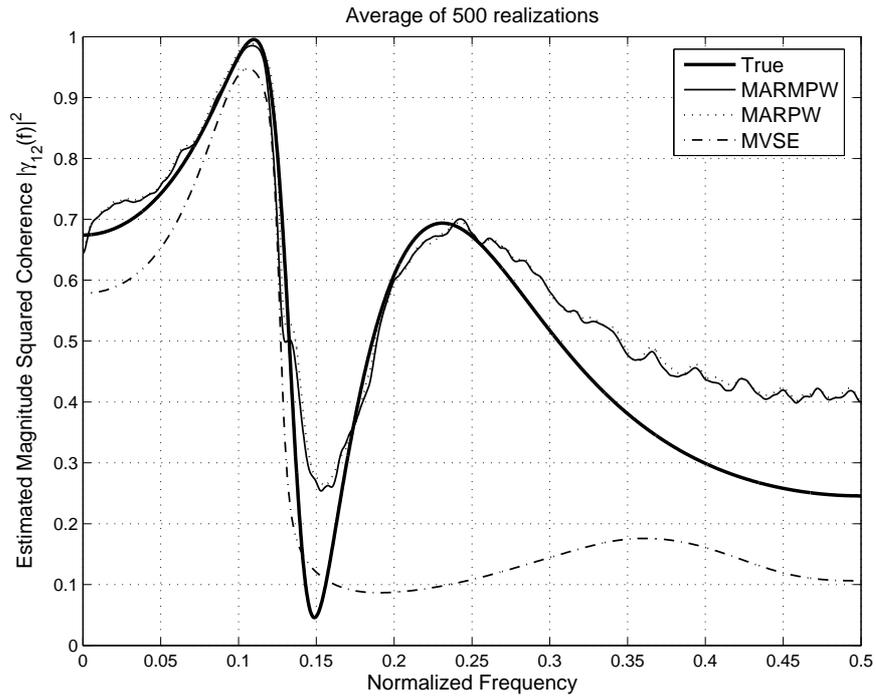Figure 14: Comparison of MSC estimates of MARMPW to periodogram and AR-YW for an ARMA process.

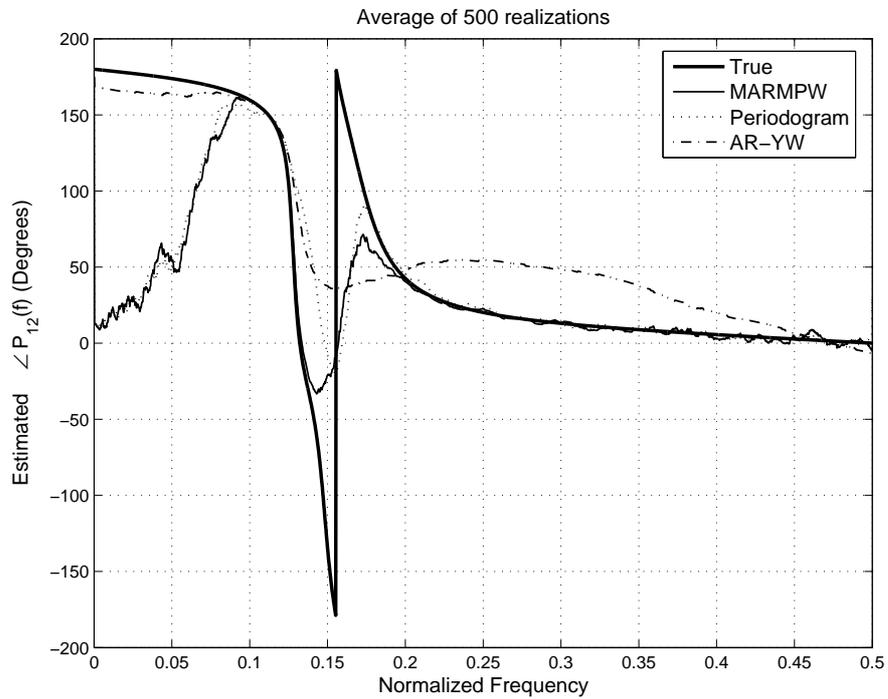Figure 15: Comparison of MSC estimates of MARMPW to MARPW and MVSE for an ARMA process.

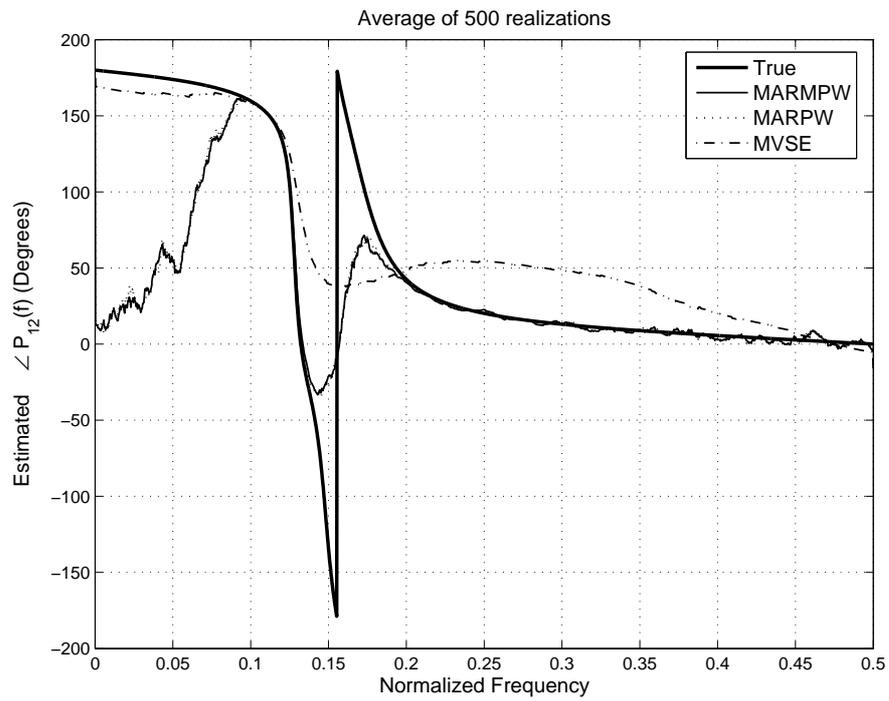Figure 16: Comparison of $\angle \hat{P}_{12}(f)$ of MARMPW to periodogram and AR-YW for an ARMA process.

Figure 17: Comparison of $\angle \hat{P}_{12}(f)$ of MARMPW to MARPW and MVSE for an ARMA process.