A COMPLEX CROSS-SPECTRAL DISTRIBUTION MODEL USING NORMAL VARIANCE MEAN MIXTURES

J. A. Palmer[1], S. Makeig[1] and K. Kreutz-Delgado[2]

[1]Swartz Center for Computational Neuroscience
[2]Department of Electrical and Computer Engineering
University of California San Diego, La Jolla, CA 92093

ABSTRACT

We propose a model for the density of cross-spectral coefficients using Normal Variance Mean Mixtures. We show that this model density generalizes the corresponding marginal density of the Complex Wishart distribution for the cross-spectral density. The Maximum Likelihood estimate of parameters in the distribution is derived, and examples are given from alpha brain wave sources in separated EEG data.

Index Terms— cross-spectrum, coherence, phase, estimation

1. INTRODUCTION

We consider the distribution of the cross-periodogram, defined as the conjugate product of the complex short time Fourier transform coefficients at a given frequency. Cross-spectral coefficients are typically averaged to form estimates of the coherence and phase lag between two random processes at a given frequency. To use these estimates in analysis or decision making, we would like to have exact or at least approximate forms of the probability density of the estimates.

Estimation of coherence between signals is important in many signal processing areas, including communications and biological signal processing. We propose a derivation of the distribution of cross-spectral coefficients based on the spectral representation of stationary processes. This derivation leads to the consideration of Normal Variance Mean Mixtures, which are skewed forms of Gaussian Scale Mixtures.

To fix ideas and describe the practical context that motivates the present work, consider the application of electroencephalogram (EEG) analysis [1]. In the EEG setup, an array of electrodes on the scalp records the superposition of ambient electromagnetic field generating sources. The technique of Independent Component Analysis (ICA) [1] has been applied successfully to such recordings to separate them into brain components of interest, and artifacts and interfering signals not of interest.

Figure 1 shows EEG data after it has been separated into independent source activations using ICA, along with the corresponding scalp maps defining the projection of the source activity to the scalp. A prominent rhythm in the recording is the alpha, or 10 Hz, rhythm, and this rhythm is also present in several of the separated components. ICA separates out the alpha activity into a small number of components with spatially distinct scalp maps. These components however have residual dependence that cannot be eliminated by an instantaneous linear unmixing.

The assumption of independence among the brain sources is, however, only made as an analytical expedient. Ultimately we would like to model any residual dependence that might remain after the attempted separation, and indeed such residual dependence is often the primary object of inquiry. A type of signal dependence of central importance in the analysis of EEG and MEG signals in neuroscience is coherence.

Components 6, 12, and 22 each show significant alpha activity at similar times. The scalp maps each project widely, to largely distinct regions. A basic question is whether or not the alpha activity separated by ICA is in indeed independent, or whether it represents aspects of a single underlying dynamic alpha phenomenon involving all of these components, which is responsible for all of the alpha. In the latter case, the alpha signals in the (supposedly) separated components should be phase coherent (as well as amplitude coherent), i.e. the periodic activity washing over them should have a constant phase delay.

2. COMPLEX CROSS SPECTRA

Given two real-valued, zero-mean, discrete time stationary random processes, $X_1(t)$ and $X_2(t)$, the cross-spectrum $S_{12}(\omega)$ is defined by $S_{12}(\omega) \triangleq \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} R_{12}(\tau) \exp(-\jmath\omega\tau)$, where we denote the imaginary number by $\jmath \triangleq \sqrt{-1}$, $E\{\cdot\}$ denotes expectation, and $R_{12}(\tau) = E\{x_1(t+\tau)x_2(t)\}$ is the cross-covariance. Using the spectral representation, the stationary time-series $X_i(t)$ can be represented as a sum of complex exponentials,

$$X_i(t) = \int_{-\pi}^{\pi} \exp(\jmath\omega t)\, dZ_i(\omega)$$

where $Z_i(\omega)$ is an orthogonal increment process. The cross spectrum can also be written,

$$S_{12}(\omega) = E\{dZ_1(\omega)\, dZ_2^*(\omega)\}$$

If we interpret the increment $dZ_i(\omega)$ as the coefficient of $\exp(\jmath\omega t)$ in the $X_i(t)$, then $S_{12}(\omega)$ may be interpreted as the average value of the product of these coefficients of $\exp(\jmath\omega t)$ in the processes X_1 and X_2 [2, p. 659]. Letting $dZ_i(\omega) = |dZ_i(\omega)| \exp(\jmath\phi_i(\omega))$, we see that,
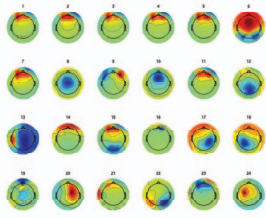
$$S_{12}(\omega) = E\{|dZ_1(\omega)||dZ_2(\omega)| \exp(\jmath(\phi_1(\omega) - \phi_2(\omega)))\}$$

Thus the magnitude of the cross-spectrum is approximately the average of the product of the magnitudes of the spectra of X_1 and X_2 at the frequency ω, while the phase of the cross-spectrum is approximately the average *difference* in the phases of X_1 and X_2 at frequency ω [2, pp. 660-661].
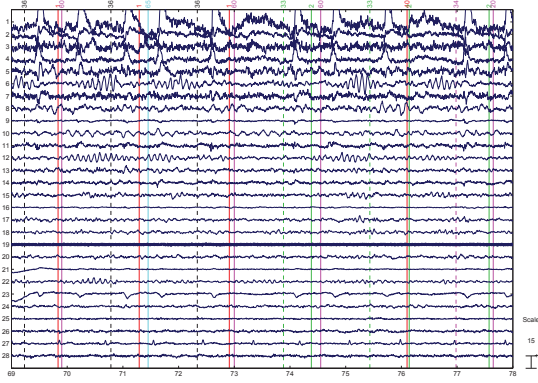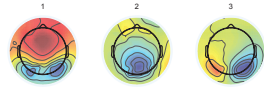
(a) Separated EEG activations

(b) EEG component maps   (c) Alpha components 6, 12, 22

Fig. 1. Separated EEG components. (a) Activation time series of top variance components in EEG unmixed using ICA. (b) Component maps corresponding to the activations in (a). (c) Component maps corresponding to the alpha activations in components 6, 12, and 22 respectively.

The spectral representation and the convenience of the FFT motivate the following estimate for the cross-spectrum. Define the finite Fourier transform of the T-sample segment of process i by,

$$d_{X_i}(\omega) \triangleq \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_i(t) \exp(-\jmath\omega t)$$

A window or taper is commonly used as a convergence factor [3], but we shall suppress this here for simplicity of exposition since the results are largely independent of the taper. The periodogram for process i is defined by [3, p. 120],

$$I_{ii}(\omega) \triangleq (2\pi)^{-1} |d_{X_i}(\omega)|^2$$

The periodogram is used as a statistic for the estimation of the power spectral density. Similarly, the cross-periodogram, defined for processes X_1 and X_2 by,

$$I_{12}(\omega) \triangleq (2\pi)^{-1} d_{X_1}(\omega) d_{X_2}^*(\omega)$$

is used as a statistic for the estimation of the cross-spectrum.

The form of the distribution is motivated by the following simplified model. We shall be interested primarily in estimating the cross-spectra between two series whose individual spectral densities are sharply peaked about a particular frequency. We thus model
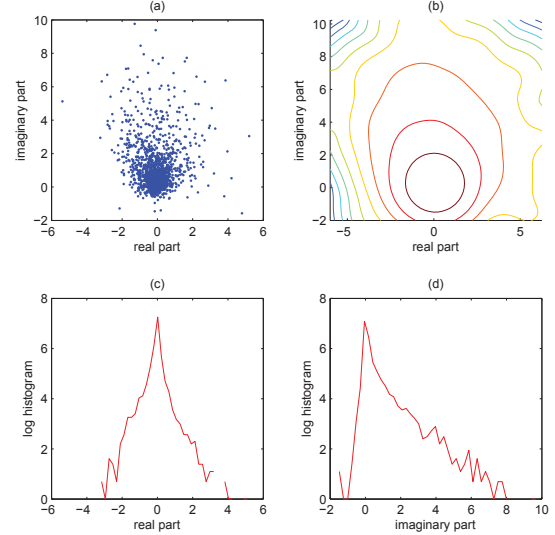

Fig. 2. Complex cross-periodogram of components 6 and 12, at 10Hz (sampling rate 256Hz). 3000 samples using non-overlapping blocks of 750 ms (192 samples). (a) Scatter plot of complex cross-periodogram values. (b) Contour plot of smoothed 2D histogram. (c) Log histogram of the real part. (d) Log histogram of imaginary part. Note the log convexity of the tails.

the two processes $X_1(t)$ and $X_2(t)$ as follows. Suppose we have B segments, or blocks, of each process. We model the realization in the kth segment by,

$$
\begin{aligned}
X_1(t) &= \xi_k^{1/2} A_1 \cos(\omega t + \theta_k) + \nu_1(t) \\
X_2(t) &= \xi_k^{1/2} A_2 \cos(\omega t + \theta_k + \phi) + \nu_2(t)
\end{aligned}
$$

where $\nu_1(t)$ and $\nu_2(t)$ are independent stationary noise processes with spectra $N_1(\omega)$ and $N_2(\omega)$ respectively. The random non-negative amplitude coefficient ξ_k reflects a common power source modulating the processes X_1 and X_2, scaled respectively by the positive amplitudes A_1 and A_2, which are taken to be constant. The segments are taken to be chosen such that the phase θ_k is uniformly random over $(-\pi, \pi)$, with the phase lag between X_1 and X_2 at frequency ω given by ϕ, which is taken to be constant.

We now consider the distribution of the cross-periodogram statistic for $X_1(t)$ and $X_2(t)$. The Fourier coefficients, $d_{X_i}(\omega)$ are given by,

$$
\begin{aligned}
d_{X_1}(\omega) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} X_1(t) \exp(-\jmath\omega t) \\
&= \tfrac{1}{2}\sqrt{T} A_1 \exp(\jmath\theta_k) \xi_k^{1/2} + d_{\nu_1}(\omega)
\end{aligned}
$$

Similarly,

$$d_{X_2}(\omega) = \tfrac{1}{2}\sqrt{T} A_2 \exp(\jmath\theta_k) \exp(\jmath\phi) \xi_k^{1/2} + d_{\nu_2}(\omega)$$

The cross-periodogram statistic is thus given by,

$$2\pi\, I_{12}(\omega) = \tfrac{1}{4} A_1 A_2 \exp(-\jmath\phi)\, T\xi_k + Y_k\, (T\xi_k)^{1/2} + W_k \quad (1)$$

3570

where,

$$Y_k = \tfrac{1}{2}A_1 \exp(\jmath\theta_k) d^*_{\nu_2}(\omega) + \tfrac{1}{2}A_2 \exp(-\jmath\theta_k)\exp(-\jmath\phi)d_{\nu_1}(\omega)$$

and,

$$W_k = d_{\nu_1}(\omega)\, d^*_{\nu_2}(\omega)$$

Y_k is the sum of independent complex Normal random variables, and is thus complex Normal (and circular). W_k is a complex Gaussian scale mixture, which is also circular. We assume that the noise processes have relatively small variance so that the distribution of W_k is concentrated around zero and does not significantly influence the form of the distribution of $I_{12}(\omega)$. From (1), we see that the joint distribution of the real and imaginary parts of $I_{12}(\omega)$ has the form of a Normal Variance Mean Mixture [4],

$$\mathbf{x} = \xi^{1/2}\mathbf{z} + \xi\boldsymbol{\beta} \tag{2}$$

convolved with the distribution of W_k. The non-negative scale mixing random variable is $T\xi$, and the drift (or skew, or bias) vector $\boldsymbol{\beta} = \tfrac{1}{4}A_1 A_2 \exp(-\jmath\phi)$.

3. COMPLEX GAUSSIAN SCALE MIXTURES AND NORMAL VARIANCE MEAN MIXTURES

Real valued scalar Gaussian scale mixtures (GSMs) [4, 5] have the form $p(x) = \int_0^\infty \mathcal{N}(x; 0, \xi)\, dF(\xi)$, where $\mathcal{N}(x; \mu, \sigma^2)$ denotes the Gaussian density with mean μ and variance σ^2. A GSM distributed random variable, x, can be represented as a product $x = \xi^{1/2}z$, where $p(x|\xi) = \mathcal{N}(0, \xi)$. Vector GSMs, e.g. $\mathbf{x} = \xi^{1/2}\mathbf{z}$, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \Sigma)$, produce ellipsoidally symmetric densities.

Perhaps the most familiar GSM is Student's t distribution for the mean of a Gaussian sample scaled by square root of its sample variance. Since the sample mean and variance are independent, and the sample mean is Gaussian, the normalized sample mean is a Gaussian scale mixture, with mixing density distributed as inverse χ^2.

Normal Variance Mean Mixtures (NVMMs) are a generalization of GSMs, and can be represented in the form, $\mathbf{x} = \xi^{1/2}\mathbf{z} + \xi\boldsymbol{\beta}$. The resulting density is non-ellipsoidally symmetric (non-circular), and skewed in the direction of $\boldsymbol{\beta}$, in a cone like shape. Note that the particular combination of exponents $\xi^{1/2}$ and ξ is required for the analytical tractability.

3.1. Complex Fourier coefficients of stationary time series

Fourier coefficients of a stationary time series are complex valued random variables, possessing a (circular) *complex Normal* distribution [3, 6, 7], denoted, $\mathbf{x} \sim \mathcal{N}_C(\mathbf{c}, \Sigma)$. The (real domain) joint probability density over the real and imaginary parts, $(\mathbf{x}_R, \mathbf{x}_I)$ can be written in terms of complex quantities $p(\mathbf{x}) \triangleq p(\mathbf{x}_R, \mathbf{x}_I)$ as,

$$p(\mathbf{x}; \mathbf{c}, \Sigma) = \pi^{-n}(\det \Sigma)^{-1} \exp\left(-(\mathbf{x} - \mathbf{c})^H \Sigma^{-1}(\mathbf{x} - \mathbf{c})\right)$$

In particular, the univariate complex normal distribution is given by

$$\mathcal{N}_C(x; \mu, \sigma^2) = \pi^{-1}\sigma^{-2}\exp(-\sigma^{-2}|x - \mu|^2)$$

where x_R and x_I are independent $\mathcal{N}(\mu_R, \sigma^2/2)$ and $\mathcal{N}(\mu_I, \sigma^2/2)$ respectively. As in the real valued case, scale mixtures of complex Normals and complex NVMMs may also be constructed. The scalar multiplier is real valued, and multiplies a complex random vector. NVMMs are also extendable to complex random vectors with complex drift.

3.2. Distribution of cross spectra

If $\mathbf{x}$ is complex Normal, then the distribution of $\sum_{t=1}^N \mathbf{x}_t\mathbf{x}_t^H$ is complex Wishart [3, 6]. The distribution is defined over the set of Hermitian, positive semi-definite $\mathbf{S} \in \mathbf{C}^{n \times n}$,

$$p(\mathbf{S}; \Sigma) = C(n, N)^{-1}|\Sigma|^{-N}|\mathbf{S}|^{N-n}\exp\left(-\operatorname{trace}(\Sigma^{-1}\mathbf{S})\right)$$

where $N \geq n$, and $C(n, N) = \pi^{n(n-1)/2}\prod_{i=1}^N \Gamma(N - i + 1)$. The distribution of the cross-periodogram for complex Normal random variables when $N \geq n$ may be obtained by calculating the marginal distribution of an off-diagonal element of the complex Wishart density. But the distribution of the cross-periodogram values themselves (without summing over multiple blocks) has $N = 1$, and thus even for the minimal dimension of 2 which is necessary for a cross-periodogram, this form of complex Wishart distribution fails to describe the distribution of the raw cross-periodogram values. We may also determine the raw cross-periodogram distribution directly, as follows.

Proposition 1. *Let x_1 and x_2 be complex Normal distributed random variables with complex covariance Σ. Then the random variable $s = x_1 x_2^*$ is a complex Normal Variance Mean Mixture with χ^2 distributed mixing density.*

Proof. Define the Cholesky factorization $\Sigma = \mathbf{T}^H\mathbf{T}$, where

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}$$

with $T_{11}, T_{22} > 0$. Then $\mathbf{x} = [x_1\, x_2]^T$ can be represented as $\mathbf{x} = \mathbf{T}\mathbf{z}$, where $\mathbf{z} = [z_1\, z_2]^T$ with z_1 and z_2 complex Normal and independent. Then we have,

$$\begin{aligned}
x_1 x_2^* &= (T_{11}z_1 + T_{22}z_2)T_{22}z_2^* \\
&= T_{11}z_1 z_2^* + T_{12}T_{22}|z_2|^2 \\
&= |z_2|T_{11}e^{-\jmath\phi_2}z_1 + |z_2|^2 T_{12}T_{22}
\end{aligned}$$

Thus $x_1 x_2^*$ is distributed as a Normal Variance Mean mixture distribution with mixing variable $\xi = |x_2|^2$, which has a χ^2 distribution. $\qquad\square$

The resulting Gaussian scale mixture has a Mackay's Bessel K distribution [4]. Note that Mackay's Bessel K distribution has exponential, log concave tails. Student's t distribution, which has mixing density *inverse* gamma distributed, does not admit skew because it has asymptotically algebraic tails, while skewed NVMMs require exponential tails [4].

3.3. Generalized Inverse Gaussian and Generalized Hyperbolic

The Generalized Inverse Gaussian (GIG) density [4, 8, 9] can be thought of as a combination and generalization of the gamma and inverse gamma distributions. The GIG density has the form,

$$\mathcal{N}^\dagger(\xi; \lambda, \delta^2, \kappa^2) = \frac{(\kappa/\delta)^\lambda}{2\,K_\lambda(\delta\kappa)}\,\xi^{\lambda-1}\exp\left(-\tfrac{1}{2}\left(\delta^2\xi^{-1} + \kappa^2\xi\right)\right) \tag{3}$$

for $\xi > 0$, where K_λ is the Bessel K function, or modified Bessel function of the second kind. The moments of the Generalized Inverse Gaussian are easily found by direct integration, using the fact that (3) integrates to one,

$$E\{\xi^r\} = \left(\frac{\delta}{\kappa}\right)^r \frac{K_{\lambda+r}(\delta\kappa)}{K_\lambda(\delta\kappa)} \tag{4}$$

The parameters λ, δ, and κ are redundant given the scale. As we will estimate the scale of the complex normal and the norm of the drift vector directly, we will eliminate this redundancy by setting $\kappa = 1$ in the mixing density.

The generalized hyperbolic distribution [4], in dimension d, is the Normal Variance Mean Mixture with GIG mixing density. We write the generalized hyperbolic density in the form,

$$\mathcal{GH}(\mathbf{x}; \Sigma, \boldsymbol{\beta}, \delta, \lambda) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \frac{1}{\delta^\lambda K_\lambda(\delta)} \times$$
$$\left(\frac{\delta(\mathbf{x})}{\kappa(\boldsymbol{\beta})}\right)^{\lambda-d/2} K_{\lambda-d/2}\big(\delta(\mathbf{x})\kappa(\boldsymbol{\beta})\big) \exp(\boldsymbol{\beta}^T \Sigma^{-1} \mathbf{x})$$

where $\delta(\mathbf{x}) \triangleq \sqrt{\delta^2 + \|\mathbf{x}\|^2_{\Sigma^{-1}}}$ and $\kappa(\boldsymbol{\beta}) \triangleq \sqrt{1 + \|\boldsymbol{\beta}\|^2_{\Sigma^{-1}}}$.

4. MAXIMUM LIKELIHOOD BY EM

Given the mixing random variable ξ, a NVMM is Normal,

$$p(\mathbf{x}; \Sigma, \boldsymbol{\beta}) = \int_0^\infty \mathcal{N}\big(\mathbf{x}; \xi\boldsymbol{\beta}, \xi\Sigma\big) p(\xi) \, d\xi \qquad (5)$$

Writing out the "complete likelihood" of $(\xi, \mathbf{x})$, we get,

$$p(\xi, \mathbf{x}; \Sigma, \boldsymbol{\beta}, \delta, \lambda) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\xi^{-d/2}p(\xi; \lambda, \delta)$$
$$\times \exp\left(-\tfrac{1}{2}\xi^{-1}\mathbf{x}^T\Sigma^{-1}\mathbf{x} + \boldsymbol{\beta}^T\Sigma^{-1}\mathbf{x} - \tfrac{1}{2}\xi\boldsymbol{\beta}^T\Sigma^{-1}\boldsymbol{\beta}\right) \quad (6)$$

It can be seen that the posterior density of ξ given $\mathbf{x}$ is also GIG. Taking the logarithm, we see that we require two posterior moments, $E\{\xi|\mathbf{x}\}$ and $E\{\xi^{-1}|\mathbf{x}\}$. which are found using (4) to be,

$$E\{\xi^r|\mathbf{x}; \Sigma, \boldsymbol{\beta}, \delta, \lambda\} = \left(\frac{\delta(\mathbf{x})}{\kappa(\boldsymbol{\beta})}\right)^r \frac{K_{\lambda-d/2+r}(\delta(\mathbf{x})\kappa(\boldsymbol{\beta}))}{K_{\lambda-d/2}(\delta(\mathbf{x})\kappa(\boldsymbol{\beta}))}$$

In the case of the complex cross-spectra, the complex random variable can be seen as a two dimensional random vector containing the real and imaginary parts. In this case the variance is just a scalar, $\Sigma = \sigma^2$, giving the cross-spectral amplitude correlation. Let $\mathbf{x}$ be the vector of real and imaginary parts of the cross-periodogram statistic, so that the cross-periodogram is in $\mathbf{R}^2$.

The maximization steps can be computed in closed form. Define $\xi_t \triangleq E\{\xi|\mathbf{x}_t; \Sigma, \boldsymbol{\beta}, \lambda, \delta\}$ and $\gamma_t \triangleq E\{\xi^{-1}|\mathbf{x}_t; \Sigma, \boldsymbol{\beta}, \lambda, \delta\}$, in dimension $d = 2$, and define the quantities,

$$a = \tfrac{1}{N}\sum_{t=1}^N \gamma_t, \quad b = \tfrac{1}{N}\sum_{t=1}^N \xi_t$$

Also let $\mathbf{c} = \tfrac{1}{N}\sum_{t=1}^N \mathbf{x}_t$. Then the updates are given by,

$$\boldsymbol{\beta} = \mathbf{c}/b \qquad (7)$$
$$\sigma^2 = \tfrac{1}{2}a - \mathbf{c}^T\boldsymbol{\beta} + \tfrac{1}{2}b\|\boldsymbol{\beta}\|^2 \qquad (8)$$

This algorithm monotonically increases the likelihood with no step size parameters.

We may also fit the parameter δ by gradient ascent on the complete log likelihood as follows,

$$\Delta\delta \propto \frac{K_{\lambda+1}(\delta)}{K_\lambda(\delta)} - \frac{2\lambda}{\delta} - a\delta$$
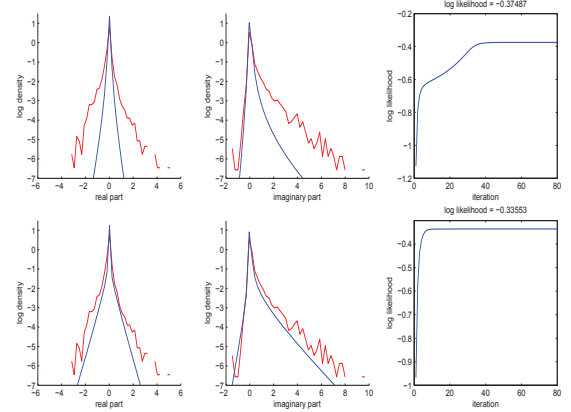

Fig. 3. Fits using the EM algorithm. Top row: one model, $\lambda = 0$. Bottom row: two models $\lambda_1 = 0, \lambda_2 = 1$. The δ parameters were optimized with the σ and $\boldsymbol{\beta}$ parameters. Last panel demonstrates monotonicity of the closed form algorithm.

5. DISCUSSION AND CONCLUSION

The probabilistic EM framework allows straightforward extension to a mixture of complex NVMM densities. In Figure 3, we see in the top row that while the Maximum Likelihood fit of a single generalized hyperbolic is log convex and more likely than the single Wishart model, it nevertheless poorly captures the coherence tail. The bottom row of Figure 3 shows that the ML estimate of a mixture densities, a non-coherent (circular) complex GSM and a coherent NVMM (non-circular) density, yields a good fit. This is not surprising since examination of the data shows that the signals are not oscillating at alpha continuously, but only over certain periods. The mixture model then gives an estimate of the relative proportions of the coherent and non-coherent periods, as well as a simple Bayesian detection for the (blindly learned) coherence model vs. the non-coherent model. The EM algorithm provides monotonic convergence with closed form updates.

6. REFERENCES

[1] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, D. Touretzky, M. Mozer, and M. Hasselmo, Eds. 1996, pp. 145–151, MIT Press, Cambridge, MA.

[2] M. B. Priestley, *Spectral Analysis and Time Series*, vol. 2, Academic Press: London, 1981.

[3] D. R. Brillinger, *Time Series: Data Analysis and Theory*, SIAM: Philadelphia, 2001, Reproduction of original publication by Holden Day, Inc., San Francisco, 1981.

[4] O. Barndorff-Nielsen, J. Kent, and M. Sørensen, "Normal variance-mean mixtures and *z* distributions," *International Statistical Review*, vol. 50, pp. 145–159, 1982.

[5] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. Roy. Statist. Soc. Ser. B*, vol. 36, pp. 99–102, 1974.

[6] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152–177, 1963.

[7] E. J. Hannan, *Mutltiple Time Series*, John Wiley & Sons, Inc., New York, 1970.

[8] T. Eltoft, T. Kim, and T.-W. Lee, "Multivariate scale mixture of Gaussians modeling," in *Proceedings of the 6th International Conference on Independent Component Analysis*, J. Rosca et al., Ed. 2006, Lecture Notes in Computer Science, pp. 799–806, Springer-Verlag.

[9] J. A. Palmer, K. Kreutz-Delgado, B. D. Rao, and S. Makeig, "Modeling and estimation of dependent subspaces with non-radially symmetric and skewed densities," in *Proceedings of the 7th International Conference on Independent Component Analysis*, M. E. Davies, C. J. James, S. A. Abdallah, and M. D Plumbley, Eds. 2007, Lecture Notes in Computer Science, Springer-Verlag.