Maximum likelihood and maximum a posteriori estimation of Gaussian spectra. Application to attenuation measurements and color Doppler velocimetry

J.-F. Giovannelli[1], *J. Idier*[1], *B. Querleux*[2], *A. Herment*[3] *& G. Demoment*[1]
[1]Laboratoire des Signaux et Systèmes, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France.
email: giova@lss.supelec.fr & idier@lss.supelec.fr
[2]Laboratoires de recherche de L'Oréal, 93601 Aulnay-Sous-Bois Cedex, France
[3]INSERM U.256, Hôpital Broussais, 75674 Paris Cedex, France.

ABSTRACT: We address the problem of tracking slowly varying mean frequency and spectral width from a series of data vectors $\boldsymbol{y}_p$ $(p = 1, 2, \ldots P)$. When the $\boldsymbol{y}_p$ are small-sized usual methods (periodogram, correlation lags...) suffer from strong variability. Moreover, the discrete nature of observed signals results in ambiguity in spectral domain. First in a Gaussian setting both for the shape of the spectra and for the statistics of observed signals, we provide a Maximum Likelihood (ML) technique is provided for spectral moments estimation. Second, we propose a Maximum A Posteriori (MAP) method in order to account also for the slowly varying character of the spectral moments. Noticeable improvement are obtained with respect to usual methods: strong variance reduction in the field of ultrasound Attenuation Measurement (AM), and correct estimation beyond the Shannon limit in the field of Color Flow Mapping (CFM).

KEYWORDS: Color flow mapping, attenuation measurement, Gaussian spectra, maximum likelihood, maximum a posteriori

1. INTRODUCTION AND PROBLEM STATEMENT

Ultrasonic characterization of biological tissues provides a cheap and harmless support for diagnosis and control. This article is both devoted to ultrasound Color Flow Mapping (CFM) and ultrasound Attenuation Measurement (AM). Both problems are closely related to first and second-order spectral moments estimation.

1.1. Color Flow Mapping

Fig. 1 presents the principles of measurements for CFM. Four-pulses Doppler data $\boldsymbol{y}_p$ are simulated at $P = 64$ consecutive depths $(p = 1, 2, \ldots P)$. Fig. 1 presents both the signals and the associated mean frequency sequences.

The aim is to estimate the spectral moments $\boldsymbol{\theta}_n = [m_n, r_n]$ $(n = 1, 2, \ldots N)$. An alternate form of the problem in which each couple of spectral moments is estimated from a collection $\boldsymbol{Y}_n$ of several $\boldsymbol{y}_p$ is also addressed.

Since the velocity is usually distributed according to a Gaussian density, the proposed method approximates the true PSD with a Gaussian parametric model.

The specificity of the problem is two-fold. On the one hand, the very low size of signals $\boldsymbol{y}_p$ (here $L = 4$ samples) results in a reduced amount of information provided by the measurement. On the other hand, the discrete character of the data induces spectral aliasing and ambiguous frequency recovery.

1.2. Attenuation Measurement

Principles of measurements are presented on Fig. 2. An experimental study has been based on 9 sets of B-mode echographic in vivo data from the same context. Fig. 2 shows one of the data sets.

From such a set of data Kuc [6] proposes two methods of estimation for tissue attenuation.

Spectral difference. As a matter of fact differences between log-spectra at variable depths very simply yield the attenuation function, at least in a ideal context.

Spectral shift. In the following we rather focus on an alternate method which relies on two assumptions. First, linear attenuation is assumed within the propagating medium:

$$S_d(f) = S_0(f) \exp -4d\beta f \qquad (1)$$

Second, a Gaussian spectrum is assumed for the incident wavelet:

$$S_0(f) = K_0 \exp -(f - f_0)^2/2r_0 \qquad (2)$$

Then [6] shows that the PSD remains Gaussian during propagation. Furthermore, the spectrum mean linearly decreases while the spectrum variance remains constant. The slope of the mean decrease and the constant level of the variance jointly yield the characteristic attenuation coefficient.

1.3. A common problem

Both CFM and AM applications require that the first two unknown spectral moments $\boldsymbol{\theta}_n = [m_n, r_n]$ of Gaussian spectra be estimated as functions of n ($n = 1, 2, \ldots N$), from a series of quasi-stationary data vectors $\boldsymbol{y}_p$.

From the above definitions and assumptions AM and CFM are based on Gaussian PSD parameters estimation according to the distance covered by the ultrasonic wave in the medium. Tracking Gaussian PSD parameters is in the scope of the following sections.

2. Usual methods

This issue is classical in several applied domains. Most usual methods are based on the computation of empirical correlation coefficients within each $\boldsymbol{y}_p$ and estimation of N independent mean frequencies.

2.1. Bibliography

The most popular mean frequency estimator is proportional to the phase of the first empirical correlation lag [9]. In [8] this estimator is shown to be equivalently provided by first order autoregression in a least squares framework.

Better accuracy is provided by the use of all the estimated correlation lags [9]. The mean frequency of each $\boldsymbol{y}_p$ is estimated as a linear combination of all the empirical correlations lags. The result is actually the mean frequency of the periodogram in an arbitrary reduced frequency domain of size one such as $[-0.5, 0.5]$. This illuminates the fact that the resulting estimator may presents a strong bias if the true frequency does not strictly range over the assumed interval.

In order to alleviate this problem, [2] proposes an iterative shift of the data up to the center of the frequency domain and [7] proposes to substitute the maximum for the average of the periodogram. The latter choice is equivalent to the maximum likelihood estimate of the frequency of a pure sinusoid embedded in a Gaussian noise (see [5], p.410).

These methods are based on the empirical correlation of each $\boldsymbol{y}_p$ and estimation of N independent mean frequencies. We have implemented the different methods and typical periodogram results are presented here.

2.2. FFT analysis.

Periodogram computation is the basic and most popular method for spectral analysis based on simple Fast Fourier Transforms (FFT).

AM: A set of $L = 16$ samples periodograms has been calculated and averages over the $K = 75$ traces have been performed, one at each of the $N = 20$ depths. Computation of mean frequencies and spectral widths yield the results shown in Fig. 3.

These results suffer from the basic default of the periodogram *i.e.* strong variability of the estimated spectral parameters. On the one hand high variability from an image to the other is observed. On the other hand, each mean frequency or spectral width sequence is too rough to allow reliable conclusions.

CFM: No frequency tracking beyond $f = 0.5$ is achieved so true frequencies beyond $f = 0.5$ induce strong bias as shown in Fig. 3.

3. Maximum Likelihood

3.1. Single channel processing

As mentioned above we focus on Gaussian spectra (case of pure sinusoids can be found in [3]). Gaussian spectrum reads:

$$S_x(f) = r_x \ (2\pi r)^{-1/2} \exp\left(\frac{-1}{2r}(f-m)^2\right), \qquad (3)$$

and inverse Fourier transform yields the continuous time correlation:

$$r_x(\tau) = r_x \ \exp\left(-2\pi^2 r \tau^2\right) \exp\left(-2i\pi m \tau\right) \qquad (4)$$

Discrete samples of the correlation are collected in the covariance matrix:

$$R_x = r_x \Pi_x(m, r) \qquad (5)$$

$$\Pi_x(m, r) = \text{Tœplitz}\left[\pi(0), \pi(1), \pi(2), \ldots, \pi(L-1)\right] \qquad (6)$$

$$\pi(l) = \exp\left(-2\pi^2 r l^2\right) \exp\left(-2i\pi m l\right) \qquad (7)$$

Additive Gaussian zero-mean stationary white observation noise $\boldsymbol{b}$ yields observation vector $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{b}$.

Hence the observation covariance is $R_y = R_x + R_b = r_x \Pi_x + r_b I$.

Stationary Gaussian hypothesis for the statistics of $\boldsymbol{x}$ and $\boldsymbol{b}$ gives a parametric structure to the likelihood functions $f(\boldsymbol{y}_p|\boldsymbol{\theta}_n)$. The opposite of the logarithm yields the anti-log-likelihood (ALL):

$$ALL_y(\boldsymbol{\theta}) \propto \log(\det R_y) + \boldsymbol{y}^\dagger R_y^{-1} \boldsymbol{y} \tag{8}$$

3.2. K channels processing

When a set of K signals $\boldsymbol{y}_p$ (such as Fig. 1 and 2) is available instead of a single measured signal, estimation of a single couple of spectral moments from joint processing is preferable than *ad hoc* averaging of K estimated spectral moments.

In the proper vectorial setting, conditional independence knowing $\boldsymbol{\theta}_n$ is introduced in order to process several vectors. In other words, no correlation information between the $\boldsymbol{y}_p$ is available, save the fact that they have the same spectral moments. Such an assumption leads to likelihood factorization *i.e.* summation of ALL of (8):

$$ALL_Y(\boldsymbol{\theta}) = \sum_{k=1}^{k=K} ALL_{y_k}(\boldsymbol{\theta}) \propto \log\det R_y + 2\mathrm{Tr}\left\{R_y^{-1} M\right\} \tag{9}$$

$$\text{with} \quad M = \frac{1}{K} \sum_{k=1}^{k=K} \boldsymbol{y}_k^\dagger \boldsymbol{y}_k \tag{10}$$

Let us note the 1-periodicity likelihood with respect to mean frequency. As a consequence, frequency ambiguity is not alleviated by the maximum likelihood approach. As depicted by ALL criteria in Fig. 5, periodicity strictly implies equi-likelihood between f and every other $f + p$.

3.3. N depths processing

Let us consider the global problem of estimating all the spectral moments, namely $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \boldsymbol{\theta}_N]$, on the basis of the whole data set *i.e.* from $\boldsymbol{\mathcal{Y}}$.

As long as we do not incorporate dependence between the different depths given the series of spectral moments, decorrelation conditionally to $\boldsymbol{\Theta}$ is assumed. This assumption leads to summation of ALL at each depth according to:

$$ALL_{\mathcal{Y}}(\boldsymbol{\Theta}) \propto \sum_{n=1}^{n=N} ALL_{Y_n}(\boldsymbol{\theta}_n) \tag{11}$$

Since this criterion is a separable function of the $\boldsymbol{\theta}_n$, maximization over $\boldsymbol{\Theta}$ amounts to N optimization over the $\boldsymbol{\theta}_n$ *i.e.* independent treatment at each depth according to the two previous subsections. Maximization of each likelihood (ML) yields N independent estimates $\hat{\boldsymbol{\theta}}_{ML_n}$.

3.4. Optimization: gradient procedure

A gradient procedure can achieve the maximization and it performs well in numerous practical cases.

Calculation of the gradient criterion involves two derivation formulae found in [4]: $\partial(\log\det R)/\partial R = R^{-1}$ and $\partial(\boldsymbol{y}^\dagger R_y^{-1} \boldsymbol{y})/\partial R = R_y^{-1} \boldsymbol{y}^\dagger \boldsymbol{y} R_y^{-1}$, and it yields:

$$\frac{\partial ALL_{\mathcal{Y}}(\boldsymbol{\Theta})}{\partial m_n} = \mathrm{Tr}\left\{ \left(R_y^{-1} - R_y^{-1} M R_y^{-1}\right) \left(\frac{\partial \Pi_x}{\partial m_n}\right)^\dagger \right\} \tag{12}$$

$$\frac{\partial ALL_{\mathcal{Y}}(\boldsymbol{\Theta})}{\partial r_n} = \mathrm{Tr}\left\{ \left(R_y^{-1} - R_y^{-1} M R_y^{-1}\right) \left(\frac{\partial \Pi_x}{\partial r_n}\right)^\dagger \right\} \tag{13}$$

where

$$\frac{\partial \pi_x(l)}{\partial m_n} = 2j\pi k \pi_x(l), \frac{\partial \pi_x(l)}{\partial r_n} = -2\pi^2 k^2 \pi_x(l) \tag{14}$$

3.5. ML results

Likelihood maximization provides the results given by Fig. 4 in the field of CFM and AM.

AM: On can see a slight reduction in variability compared to the periodograms of Fig. 3. Better repeatability is obtained but the overall result still shows a poor degree of robustness.

CFM: Compared to Fig. 3, lower bias is obtained at the expense of increased variability. Moreover, as an expected consequence of likelihood periodicity, no frequency tracking is achieved beyond $f = 0.5$.

Both periodogram and ML results lack of robustness because estimation is performed independently at each depth. Reduction of variability and extended frequency tracking are obtained in the next section by accounting for the slowly varying character of the parameters in the medium.

4. Maximum A Posteriori

4.1. Prior choice

The slowly varying character of the spectral moments implies positive autocorrelation of the variables within

$\boldsymbol{m} = [m_1, m_2, \ldots, m_N]$ and $\boldsymbol{r} = [r_1, r_2, \ldots, r_N]$. In order to introduce proper correlation in a Bayesian framework, let us consider the family of Gauss prior densities $f(\boldsymbol{\Theta})$.

$$f(\boldsymbol{m}_f) = G(0, R_m) \quad \text{and} \quad f(\boldsymbol{r}_f) = G(0, R_r) \quad (15)$$

Let the inverse of the covariance matrix under the following factored form:

$$R^{-1} = r\Delta^t\Delta. \quad (16)$$

In AM and CFM we found that two types of Gauss-Markov prior could be appropriate, respectively when Δ is a first or second finite difference matrix:

- first order difference

$$\Delta_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & \ldots & \\ 0 & 1 & -1 & 0 & \ldots & \\ \vdots & & & \ddots & \ddots & \\ 0 & 0 & 0 & & 1 & -1 \end{bmatrix}, \quad (17)$$

- second order difference

$$\Delta_2 = \begin{bmatrix} 1 & -2 & 1 & & \ldots & \\ 0 & 1 & -2 & & \ldots & \\ \vdots & & & \ddots & & \\ 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}. \quad (18)$$

In the field of CFM, information of continuity is properly taken into account using first order regularization:

$$R^{-1}_{\boldsymbol{m}_f} = r_m\Delta_1^t\Delta_1 \quad \text{and} \quad R^{-1}_{\boldsymbol{r}_f} = r_r\Delta_1^t\Delta_1 \quad (19)$$

In the case of AM, linear variation is expected for the mean frequency, so the second order difference model appears a more adequate prior for the mean frequency series. On the other hand, mean variance is assumed constant so a first order difference model is chosen.

$$R^{-1}_{\boldsymbol{m}_f} = r_m\Delta_2^t\Delta_2 \quad \text{and} \quad R^{-1}_{\boldsymbol{r}_f} = r_r\Delta_1^t\Delta_1 \quad (20)$$

Exhaustive definition of the prior law involves the joint density of $\boldsymbol{m}_f$ and $\boldsymbol{r}_f$, *i.e.*, not only the marginal densities. Since no information is known about statistical dependence between the respective variations of $\boldsymbol{m}_f$ and $\boldsymbol{r}_f$, separability is naturally assumed for the joint density as the product of its marginals: $f(\boldsymbol{m}_f, \boldsymbol{r}_f) = f(\boldsymbol{m}_f)f(\boldsymbol{r}_f)$.

4.2. Posterior density

Fusion of prior and data information is achieved by the Bayes rule, which provides the posterior density of Θ:

$$f(\boldsymbol{m}_f, \boldsymbol{r}_f|\boldsymbol{\mathcal{Y}}) \propto f(\boldsymbol{\mathcal{Y}}|\boldsymbol{m}_f, \boldsymbol{r})\, f(\boldsymbol{r}_f)\, f(\boldsymbol{m}_f). \quad (21)$$

Equivalently the Anti-Log-Posterior-Likelihood function (ALPL) reads:

$$ALPL_{\mathcal{Y}}(\boldsymbol{\Theta}) = ALL_{\mathcal{Y}}(\boldsymbol{\Theta}) + ALP(\boldsymbol{\Theta}) \quad (22)$$

where $ALL_{\mathcal{Y}}(\boldsymbol{\Theta})$ is defined by (11) and the Anti-Log-Prior function (ALP) is introduced according to:

$$ALP(\boldsymbol{\Theta}) = \boldsymbol{m}_f^t R_m^{-1}\boldsymbol{m}_f + \boldsymbol{r}_f^t R_r^{-1}\boldsymbol{r}_f. \quad (23)$$

Our point estimate is the maximum *a posteriori* $\hat{\boldsymbol{\theta}}_{MAP}$, *i.e.* , the minimizer of ALPL:

$$\hat{\boldsymbol{\Theta}}_{MAP} = \arg\min_{\boldsymbol{\Theta}} \{ALPL_{\mathcal{Y}}(\boldsymbol{\Theta})\} \quad (24)$$

Simple visual interpretation of regularization is provided by Fig. 5. Whereas ALL was strictly periodic, ALPL favors one solution over the others according to prior probabilities, so that frequency ambiguity is overcome provided that effective computation of the unique global minimum can be achieved.

4.3. Optimization

The most natural approach is to design an extension of the previous gradient procedure. ALPL gradient is simply yielded by summation between the previous ALL gradient and ALP gradient:

$$\frac{\partial ALPL_{\mathcal{Y}}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} = \frac{\partial ALL_{\mathcal{Y}}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} + \frac{\partial ALP}{\partial \boldsymbol{\Theta}} \quad (25)$$

On the one hand, since the prior is Gaussian, ALP gradient is the derivative of a quadratic form, hence it is merely linear with respect to $\boldsymbol{m}_f$ and $\boldsymbol{r}_f$:

$$\frac{\partial ALP(\boldsymbol{\Theta})}{\partial \boldsymbol{m}_f} = R_m^{-1}\boldsymbol{m}_f \;\; ; \;\; \frac{\partial ALP(\boldsymbol{\Theta})}{\partial \boldsymbol{r}_f} = R_r^{-1}\boldsymbol{r}_f \quad (26)$$

On the other hand, ALL gradient is given by (12) and (13).

Gradient procedures perform well in many practical cases, especially in AM. However convergence of such

a local method towards the global maximum is not guaranteed, especially in CFM where it is clear (for instance, see Fig. 5) that we have been able to get rid of ambiguous periodicity at the expense of creating local minima.

In order to cope with global optimization, we implemented a dynamic programming procedure. It is based on the hidden Markov chain interpretation available for the Gauss-Markov prior, provided that the possible values of the parameters are properly quantized on an arbitrary grid. Then efficient global maximization is performed by the well-known Viterbi algorithm [1]. In our case, it is furthermore possible to show that the required likelihood functions are computable at low cost using FFT by-products.

In order to get rid of the quantization effect, true global optimization can be reached afterwards using the dynamic programming solution as a nearly optimal starting point for gradient optimization.

4.4. MAP Results

Fig. 6 has been computed using the MAP method.

AM: The Bayesian technique provides more reliable results in two senses. On the one hand, each trace shows better regularity since smoothness is introduced as a prior feature. On the other hand variability has also been reduced between images. Since all images are known to proceed from the same context, this is a clearly positive result even if no theoretical reference is available for sake of absolute comparison in such a real data situation.

CFM: In this case, the Bayesian approach yields again more regular and smooth results than classical methods. Moreover the estimated mean frequency sequence remains close to the theoretical profile, even beyond usual Shannon limits.

5. Conclusion

The problem of spectral moments estimation has been addressed, as it arises in the fields of ultrasound Color Flow Mapping (CFM) and ultrasound Attenuation Measurements (AM).

Under Gaussian statistics assumption for observed signals, and provided that the Power Spectral Density (PSD) of signals is a Gaussian distribution function, we proposed a solution based upon Gaussian PSD parameters tracking.

Parameter estimation was coped in a Bayesian framework, which allowed us to introduce the slowly varying character of the spectral moments as a basic prior feature within Gauss-Markov modeling.

In order to compute the Maximum A Posteriori (MAP) solution, we showed that simple gradient procedures could be trapped in local maxima. Instead we proposed a computationally efficient two-step combination between dynamic programming and a gradient procedure.

Noticeable improvements are obtained with respect to usual methods: strong variance reduction in the field of ultrasound Attenuation Measurement (AM), as well as in Color Flow Mapping (CFM). Moreover it has been clearly assessed that the new Bayesian method provides tracking capability beyond usual Shannon limits, as a specific consequence of the incorporated Markov prior. To our knowledge, such a property is an original contribution to the field of frequency tracking.

6. Acknowledgments

Partial support for this work was provided by L'Oréal under contract number 770566, 1993.

7. References

[1] G.D. Forney, "The Viterbi algorithm", *Proceedings of the IEEE*, vol.61, pp.268-277, 1973.

[2] A. Herment, G. Demoment , P. Dumée, J.P. Guglielmi & A. Delouche, "A new adaptive mean frequency estimator: Application to constant variance color flow mapping", *IEEE Trans. on Ultrason. Ferroelec. Freq. Contr.*, vol.40, pp.796-804, 1993.

[3] J. F. Giovannelli, J. Idier, A. Herment & G. Demoment, "Mean frequency tracking using Markov model", Submitted to *ICASSP 94*.

[4] G.C. Goodwin & R.L. Payne, *Dynamic system identification. Experiment design and data analysis.* Academic Press, 1977.

[5] S.M. Kay, *Modern spectral estimation, theory and applications*, Prentice Hall, 1988.

[6] R. Kuc, "Employing spectral estimation for characterizing diffuse liver disease", *Tissue characterization with ultrasound*, vol.2, pp.147-167, 1986.

[7] K.W. Li, N. Held, J.C. Curlander & C. Wu, "Doppler parameter estimation for spaceborne synthesis aperture radars", *IEEE Trans. on Geosc. and Remote Sensing.*, vol.23, pp.47-56, 1985.

[8] T. Loupas & W.N. McDicken, "Low-order complex AR models for mean and maximum frequency estimation in the context of color flow mapping", *IEEE Trans. on Ultrason. Ferroelec. Freq. Contr.*, vol.37, pp.590-601, 1990.

[9] R.F. Woodman, "Spectral moments estimation in MST radars", *Radio Science*, vol.20, pp.1185-1195, 1985.

8. Illustrations

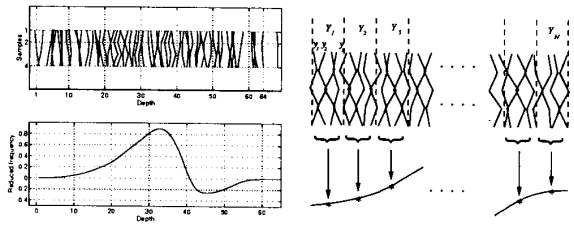

Figure 1: CFM: simulated data (left) and vectorial processing form (right).

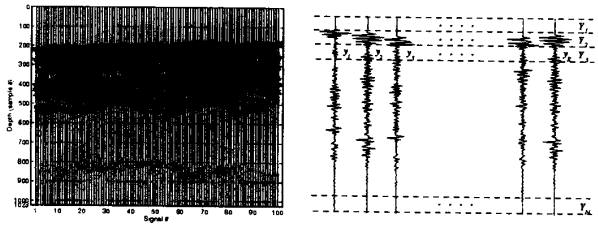

Figure 2: AM: one set of measured data (left) and vectorial processing form (right).

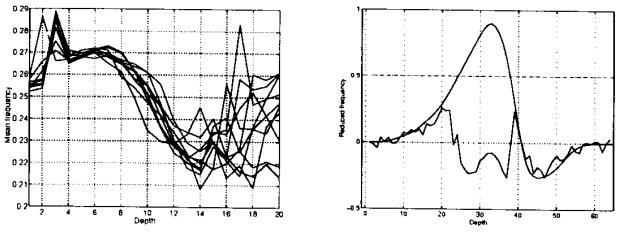

Figure 3: Periodogram results: AM (left) and CFM (right).

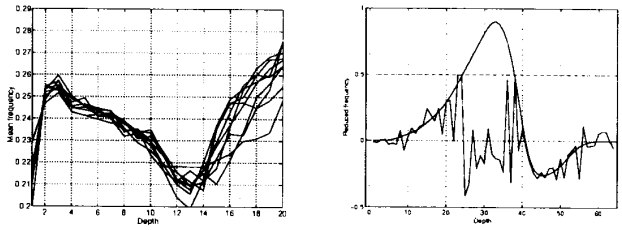

Figure 4: Maximum likelihood results: AM (left) and CFM (right).

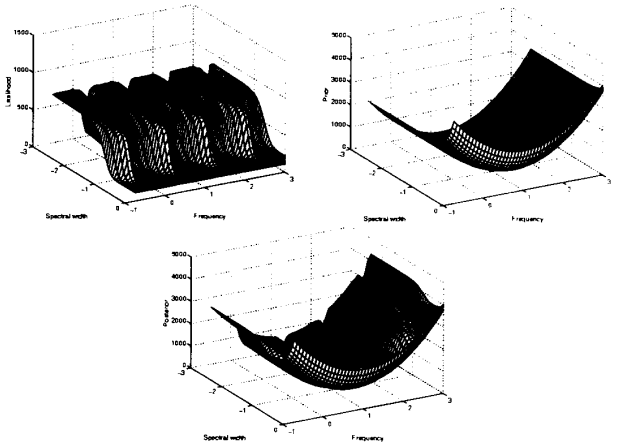

Figure 5: Criteria: likelihood (top-left), prior (top-right) and posterior (bottom)

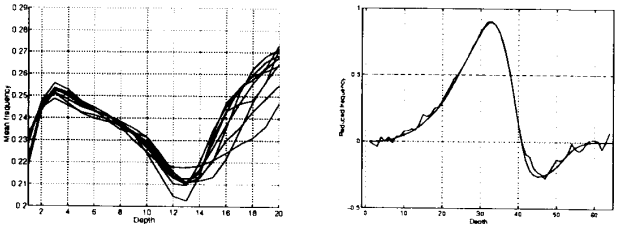

Figure 6: Maximum a posteriori results: AM (left) and CFM (right).