

A Bayesian Method for Long AR Spectral Estimation: A Comparative Study

Jean-François Giovannelli, Guy Demoment, and Alain Herment

Abstract—In this paper, we address the problem of smooth power spectral density estimation of zero-mean stationary Gaussian processes when only a short observation set is available for analysis. The spectra are described by a long autoregressive model whose coefficients are estimated in a Bayesian regularized least squares (RLS) framework accounting for the spectral smoothness prior. The critical computation of the trade-off parameters is addressed using both maximum likelihood (ML) and generalized cross-validation (GCV) criteria in order to automatically tune the spectral smoothness. The practical interest of the method is demonstrated by a computed simulation study in the field of Doppler spectral analysis. In a Monte Carlo simulation study with a known spectral shape, investigation of quantitative indexes such as bias and variance, but also quadratic, logarithmic, and Kullback distances shows interesting improvements with respect to the usual least squares method, whatever the window data length and the signal-to-noise ratio (SNR).

I. INTRODUCTION

THE SPECTRAL estimation problem has been extensively studied because of its obvious practical importance as well as its theoretical interest. Literature on the spectral estimation problem is abundant and varied since this problem arises in very different branches of engineering and applied physics: nondestructive testing, attenuation measurements, Doppler velocimetry, etc. This paper deals with the particular situation when only a short set of data is available for spectral estimation and when the spectra are known to be smooth in some sense. This paper's objective is to obtain improvement in terms of resolution of fine details of analyzed structure by processing very short windows.

The well-known periodogram is the most popular spectral estimation approach for two reasons. The first one is the computational efficiency of the fast Fourier transform (FFT) algorithm which simplifies real time applications. The second reason is its good performance when sufficient data are available. Unfortunately, when the window data length decreases, the bias and variance of the periodogram both increase, to which the literature often refers [1]–[3]. Many variations of the periodogram are also described in the literature: windowed, averaged periodograms, etc. The conclusion of the study of these methods is that the bias of this estimator can be

reduced if an increase in variance is accepted, and vice versa, but bias and variance cannot be reduced simultaneously. Hence, such a method is unusable in the particular situation when only a very short set of data is available for spectral estimation.

Parametric spectral estimation methods have gained attention as potentially interesting tools in the last two decades. They allow the improvement of the statistical properties of spectral estimators with respect to the Fourier-based methods. Estimation of the parameters of ARMA and MA models needs the resolution of a set of nonlinear equations, whereas the AR parameters estimates can be calculated by solving a set of linear ones. Moreover, algorithms, such as Levinson's, used to solve this set of equations are computationally efficient. When the AR modeling assumption is valid, spectral estimates are less biased and have lower variability than the Fourier-based ones. For these reasons, the AR method became the most popular approach to parametric spectral estimation [1]–[3]. However, as often noticed in the literature, two facts make this method less statistically reliable for a shorter observation vector. The first fact is the instability of techniques used to determine the model order, especially for short sets of data. The second fact is that the conventional least squares framework leads to a parsimonious parametric model (i.e., small number of parameters with respect to large number of data), which cannot be a good replica of the true power spectral density.

The Bayesian method used within the scope of this paper helps to alleviate this problem. The basic idea of this approach, introduced by Kitagawa and Gersch [4], is to take into account the spectral smoothness assumption in the estimation method. Such a Bayesian method is free from the parsimony principle, and allows the reliable estimation of N parameters from only N observations. Hence, this method enables the estimation of the PSD in a broader class of spectra and the description of various spectral shapes.

The paper is organized as follows. In Section II, the problem of spectral estimation of a zero-mean stationary Gaussian process is summarized, with particular attention paid to its connection to AR modelization and estimation strategy. After a short description of the Bayesian approach to parameter estimation, Section III is devoted to the presentation of the Bayesian spectral analysis method. Special attention is paid to the problem of automatic estimation of the trade-off parameters in Section IV. Section V and VI are respectively devoted to simulation methodology and description of results obtained in

Manuscript received October 18, 1994; revised September 20, 1995.

J.-F. Giovannelli and G. Demoment are with the Laboratoire des Signaux et Systèmes (CNRS-SUPÉLEC-UPS), 91192 Gif-sur-Yvette, Cedex, France (e-mail: giova@Iss.supelec.fr).

A. Herment is with INSERM U66, Imagerie Biomédicale Morphologique et Fonctionnelle, Hôpital Pitié, 75634 Paris, Cedex 13, France.

Publisher Item Identifier S 0885-3010(96)01754-6.

the field of Doppler velocimetry. Finally, conclusions are presented in Section VII as well as points still to be investigated.

II. CONVENTIONAL AR PARAMETER ESTIMATION

A. The Spectral Estimation Problem

The problem of spectral estimation is to determine the spectral content of a random process based on a finite number of observations. Mathematically, the spectral content is the PSD of a discrete-time second-order stationary process and is defined as

$$S_{\mathbf{x}}(f) = \sum_{-\infty}^{+\infty} r(k) \exp(-2i\pi fk),$$

where $r(k)$ are the correlation lags of the process.

Our field of interest is the case of real zero-mean stationary Gaussian N -dimensional observation vector $\mathbf{x} = [x_1, x_2, \dots, x_N]^t$. Hence, the probability density function for such a vector is

$$f_{X|\theta}(\mathbf{x}|\theta) = f_{X|R}(\mathbf{x}|R) = (2\pi)^{-N/2} (\det R)^{-1/2} \exp\left(-\frac{1}{2} \mathbf{x}^t R^{-1} \mathbf{x}\right), \quad (1)$$

where the Toeplitz positive covariance matrix R is parameterized by an N -dimensional parameter vector θ which may be equivalently chosen as N correlation lags, N reflection coefficients, or $N - 1$ autoregressive parameters and the noise power.

The density given by (1), seen as a function of θ , is the likelihood of θ , denoted $L(\theta)$ and gathers all the information provided by the data. Since $L(\theta)$ is a function of the N -dimensional parameter θ , it is important to note that the spectral estimation problem requires estimation of θ i.e., estimation of N coefficients from N observations.

B. Maximum Likelihood (ML) and Least Squares Estimation

A common approach to the parameter estimation problem is the ML approach, but the likelihood given by (1) is nonlinear with respect to θ due to the nonlinear dependence of R with respect to θ and to the presence of the determinant and inverse of R . Several techniques [5], [6] were proposed in order to maximize $L(\theta)$ with respect to θ , but such a nonlinear optimization problem is impractical for real-time applications. In order to avoid this problem, a particular choice for θ and an approximation of the likelihood are achieved, leading to a computationally efficient estimator. Generally, θ is chosen as the AR coefficients $\mathbf{a} = [a_1, a_2, \dots, a_{N-1}]^t$, and as long as the PSD is not sharply peaked, as in our case by hypothesis, an approximate expression for the likelihood is

$$f_{X|A}(\mathbf{x}|\mathbf{a}) = (2\pi\sigma_u^2)^{-N/2} \cdot \exp\left(-\frac{1}{2\sigma_u^2} (\mathbf{x} - X\mathbf{a})^t (\mathbf{x} - X\mathbf{a})\right) \quad (2)$$

(see [2, p. 185]). The data vector \mathbf{x} and the data matrix X (also called observation vector \mathbf{x} and observation matrix X) are designed in the classical manner [2], [3] as follows:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_N \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad X = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ x_1 & 0 & \cdots & 0 \\ x_2 & x_1 & \cdots & 0 \\ x_3 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-1} & x_{N-2} & x_2 & x_1 \\ x_N & x_{N-1} & & x_2 \\ & & \ddots & \\ 0 & \cdots & x_{N-1} & x_{N-2} \\ 0 & \cdots & x_N & x_{N-1} \\ 0 & \cdots & 0 & x_N \end{bmatrix}. \quad (3)$$

Equation (2) clearly shows that the maximization of the AR parameter approximated likelihood is equivalent to the minimization of the usual least squares criterion [2], [3]

$$Q_0(\mathbf{a}) = (\mathbf{x} - X\mathbf{a})^t (\mathbf{x} - X\mathbf{a}). \quad (4)$$

The explicit expression for the minimizer of this criterion, the least squares estimate, is well known

$$\hat{\mathbf{a}}_{LS} = \arg \min Q_0(\mathbf{a}) = (X^t X)^{-1} X^t \mathbf{x}. \quad (5)$$

C. Short Time Case

As pointed out above, the signal is parameterized by N parameters, thus estimation of N autoregressive coefficients from N observations is desirable. In such a situation, least squares estimation techniques lead to a large variance and results in many spurious spectral peaks, hence even if the solution has small bias, it yields to an unacceptable spectra. An alternative is possible to alleviate the conflict between long model and least squares methods: on the one hand, the usual approach gives up the very first idea of estimating N coefficients, on the other hand, Kitagawa and Gersch [4] propose to abandon the usual least squares principle.

As already mentioned, the first solution is often adopted. In order to reduce the estimation variance and avoid spurious peaks, in a least squares framework, the model order is drastically reduced, for instance, to one-third or one-quarter of the observation vector length. Such an approach is efficient when enough data are available, but fails for a short data set because the model is then too poor to describe a wide class of spectra.

Our approach is on the opposite method, and we propose to adopt the second solution: our criterion is a modified version of the least squares, and we go on estimating a long model. Estimation of N parameters from N observations is an ill-posed problem which suffers from a lack of information to infer N AR parameters from N data. Literature about these problems is abundant and varied since they arise in almost every branch of engineering and applied physics [7]. The resolution of ill-posed

problems needs the use of the regularization concept, i.e., the introduction of prior information in the solution. The Bayesian framework provides an attractive and coherent framework to deal with these problems. This standpoint is developed in the next section.

III. BAYESIAN APPROACH TO SPECTRAL ESTIMATION

A. Theoretical Background

In a Bayesian framework, introduction of prior information in the solution involves a change in the estimation criterion: instead of maximizing a likelihood of the parameter, the posterior likelihood is maximized, which contains information provided by the observations (via the likelihood) and by the *a priori* assumption about the expected solution (via the prior law). This compound criterion incorporates both “data-based knowledge” and “prior knowledge” about θ [4], [7].

The first problem to address is the choice of a prior probability density $f_{\Theta}(\theta)$ for the vector θ . This density is supposed to contain the prior knowledge about the expected solution (see Section III-B). The information provided by the data is introduced through the conditional density for the observations, $f_{X|\Theta}(\mathbf{x}|\theta)$. Finally, the Bayes rule combines data and prior in the posterior density for θ

$$f_{\Theta|X}(\theta|\mathbf{x}) = \frac{f_{X|\Theta}(\mathbf{x}|\theta)f_{\Theta}(\theta)}{f_X(\mathbf{x})}. \quad (6)$$

In a strict Bayesian sense, (6) yields the solution to the problem since it gathers all the information about the AR coefficients. However, the need of one spectrum implies the choice of an estimator, and a popular choice is the maximum *a posteriori* (MAP). This punctual estimator is defined as the maximizer of the posterior density

$$\hat{\theta}_{\text{MAP}} = \arg \max_{\theta} f_{\Theta|X}(\theta|\mathbf{x}).$$

The computation of this solution requires values for remaining parameters, called the hyperparameters, (parameters of the prior law, noise variance, etc.) which balance the solution between the data and the prior. The crucial problem of tuning the compromise between the fidelity to the data and the fidelity to the prior knowledge is addressed in Section IV.

B. Spectrum Smoothness Prior

The aim of this section is to design a prior law for the AR coefficients, modeling the PSD smoothness and leading to an easily computable solution.

The choice of a particular class of prior law is first driven by the need for a real-time computable estimation. This constraint leads to the consideration of the class of Gaussian prior

$$f_A(\mathbf{a}) = (2\pi)^{N/2} \det R_a^{-1/2} \exp\left(-\frac{1}{2} \mathbf{a}^t R_a^{-1} \mathbf{a}\right), \quad (7)$$

where the smoothness information about the PSD is introduced via the prior correlation R_a . Hence the objective of the followings is to find such a prior correlation.

The power spectral density for an AR process is [2]

$$S_x(f) = \frac{\sigma_u^2}{\left|1 - \sum_{k=1}^p a_k e^{2j\pi f k}\right|^2}, \quad (8)$$

which can be rewritten

$$S_x(f) = \frac{\sigma_u^2}{|1 - A(f)|^2}, \quad \text{with } A(f) = \sum_{k=1}^p a_k e^{2j\pi f k}. \quad (9)$$

When the PSD is known to be smooth, Kitagawa and Gersch [4] proposed to constrain the PSD variations, i.e., to penalize its high variations. They consider the k th derivative of $A(f)$ defined in (9) in order to measure the PSD variations. Averaging the squared modulus of this derivative over the whole reduced frequency domain, they define the PSD k th smoothness by

$$D_k = \int_0^1 \left| \frac{\partial^k}{\partial f^k} A(f) \right|^2 df. \quad (10)$$

After elementary algebra, Kitagawa and Gersch showed in [4] that

$$D_k \propto \mathbf{a}^t \Delta_k \mathbf{a}, \quad (11)$$

where the Δ_k matrix, called the k th smoothness matrix, is defined by

$$\Delta_k = \begin{bmatrix} 1^{2k} & 0 & 0 & \cdots & 0 \\ 0 & 2^{2k} & 0 & \cdots & 0 \\ 0 & 0 & 3^{2k} & \cdots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \cdots & p^{2k} \end{bmatrix}.$$

A small value of D_k means a small value of the averaged k th derivative of $A(f)$, hence a rather smooth PSD. At the limit, if $D_k = 0$, then $\mathbf{a} = 0$ and $S_x(f) = C^{te}$ over the whole frequency domain, i.e., the PSD estimate is completely flat. On the contrary, a large value of D_k implies strong variations of $A(f)$, hence a peaky PSD.

The covariance matrix R_a is designed from the k th order smoothness matrix Δ_k through the following:

$$R_a^{-1} = \frac{\lambda}{\sigma_u^2} \Delta_k.$$

The Gaussian prior defined in (7) with covariance R_a^{-1} favors smooth over peaked spectra.

C. Posterior Law and MAP Estimate

Section II shows that the approximated likelihood $f_{X|A}(\mathbf{x}|\mathbf{a})$ for the AR coefficients is built up as given in (2). The previous section gives a prior law $f_A(\mathbf{a})$ for the AR coefficients in (7). The posterior probability density function for \mathbf{a} is then derived by applying the Bayes rule (6)

$$f_{A|X}(\mathbf{a}|\mathbf{x}) = \frac{f_{X|A}(\mathbf{x}|\mathbf{a})f_A(\mathbf{a})}{f_X(\mathbf{x})}. \quad (12)$$

The denominator of (12), $f_X(\mathbf{x})$, being independent from \mathbf{a} , contributes nothing more than a normalizing constant K . Elementary algebra leads to the posterior law

$$f_{A|X}(\mathbf{a}|\mathbf{x}) = K \exp -\frac{1}{2\sigma_u^2} Q(\mathbf{a}), \quad (13)$$

where $Q(\mathbf{a})$ is given by

$$Q(\mathbf{a}) = (\mathbf{x} - X\mathbf{a})^t(\mathbf{x} - X\mathbf{a}) + \lambda \mathbf{a}^t \Delta_k^{-1} \mathbf{a} \quad (14)$$

and called the regularized least squares (RLS) criterion.

Since both the prior law $f_A(\mathbf{a})$ and the conditional law $f_{X|A}(\mathbf{x}|\mathbf{a})$ are Gaussian, and the model is linear, the posterior density is also Gaussian. Hence the choice of an estimator is no longer crucial: the MAP, the posterior mean, etc., are strictly equal. The MAP estimator $\hat{\mathbf{a}}_{\text{MAP}}$ is defined as the posterior probability density maximizer, or equivalently as the minimizer of the RLS criterion $Q(\mathbf{a})$ of (14)

$$\begin{aligned} \hat{\mathbf{a}}_{\text{RLS}} &= \hat{\mathbf{a}}_{\text{MAP}} \\ &= \arg \max f_{A|X}(\mathbf{a}|\mathbf{x}) \\ &= \arg \min Q(\mathbf{a}). \end{aligned} \quad (15)$$

Since the problem is linear and Gaussian, we still have an explicit expression for its minimum

$$\hat{\mathbf{a}}_{\text{RLS}} = (X^t X + \lambda \Delta_k)^{-1} X^t \mathbf{x}. \quad (16)$$

The crucial parameter λ , called the regularization parameter, balances between the prior and data-based solutions; the question of its estimation is addressed in Section IV.

D. Quadratic Regularization Interpretation

This section is devoted to the interpretation of this method in terms of quadratic regularization, outside the Bayesian framework.

The regularized criterion of (14) is composite. On the one hand, its first term is the data-based criterion $Q_0(\mathbf{a}) = (\mathbf{x} - X\mathbf{a})^t(\mathbf{x} - X\mathbf{a})$, on the other hand, its second term $Q_\infty(\mathbf{a}) = \mathbf{a}^t \Delta_k \mathbf{a}$ is a prior criterion. Combining these two criteria, the proposed criterion incorporates both the prior and data-based criteria

$$\begin{aligned} Q(\mathbf{a}) &= Q_0(\mathbf{a}) + \lambda Q_\infty(\mathbf{a}) \\ &= (\mathbf{x} - X\mathbf{a})^t(\mathbf{x} - X\mathbf{a}) + \lambda \mathbf{a}^t \Delta_k \mathbf{a}. \end{aligned} \quad (17)$$

As in the usual least squares case, an expression is available for the minimizer of this criterion

$$\hat{\mathbf{a}}_{\text{RLS}} = \arg \min Q(\mathbf{a}) \quad (18)$$

$$= (X^t X + \lambda \Delta_k)^{-1} X^t \mathbf{x}. \quad (19)$$

For λ small enough ($\lambda = 0$ at the limit), the criterion reduces to the usual least squares case, $Q(\mathbf{a}) = Q_0(\mathbf{a})$, and the usual least squares solution is found: $\hat{\mathbf{a}} = \mathbf{a}_0 = (X^t X)^{-1} X^t \mathbf{x}$. For λ large enough ($\lambda = \infty$ at the limit), the criterion becomes the prior one $Q(\mathbf{a}) = Q_\infty(\mathbf{a})$, and the prior solution is found: $\hat{\mathbf{a}} = \mathbf{a}_\infty = 0$ so the PSD estimate is constant over the whole frequency domain.

Between the two extreme ‘‘prior-based’’ and ‘‘data-based’’ solutions, an acceptable solution must still be found. Therefore,

the crucial parameter λ which gives a convenient solution has to be estimated from the data. The following section is devoted to this fundamental problem.

IV. HYPERPARAMETER ESTIMATION

The method described above in (16) requires values for three hyperparameters σ_u^2 , λ , and k . The parameter σ_u^2 is a scaling factor while the two other parameters design the spectral shape. The parameter λ is of major importance for the spectral shape, so our study is focused on its estimation. On the contrary, the smoothness order is of lower influence and is usually fixed to $k = 1$ [its influence is nevertheless evaluated in Section VI-C1)].

The problem of hyperparameters estimation is the most delicate problem in regularization approaches, and has been extensively studied [8]–[12]. Numerous techniques have been proposed and compared in these papers and two approaches seem to be of great interest. The first strategy is founded on ML and allows estimation of both σ_u^2 and λ . The second strategy, called generalized cross validation (GCV), provides an alternative to λ estimation. The aim of this section is to give a brief overview of the different possible methods, and especially ML and GCV which are both implemented in the simulation study in Section VI.

A. Maximum Likelihood

We first investigate methods directly derived from the Bayesian framework. One of the interests of this framework is to provide coherent techniques to estimate the hyperparameters.

The parameters of interest are the noise variance and AR parameters, but the hyperparameter λ may be considered as a nuisance parameter. In a strict Bayesian framework, λ 's integration out of the estimation problem may be desirable. Such a calculus may be driven with a Jeffreys prior [13], $f_\Lambda(\lambda) = 1/\lambda$ for instance, but the calculus is out of the scope of this article. However, it may be mentioned that this approach leads to a nonlinear criterion on \mathbf{a} , and consequently to an untractable method as far as real-time applications are concerned.

A second approach consists of estimating both the hyperparameters and AR parameters in the same pattern: the maximization of the joint probability density function, called the generalized likelihood

$$GL(\mathbf{a}, \lambda, \sigma_u^2) = f_{X,A}(\mathbf{x}, \mathbf{a}|\lambda, \sigma_u^2),$$

simultaneously over all the AR parameters \mathbf{a} , λ , and σ_u^2 . It can be shown that for this kind of problem (linear and Gaussian), the solution is degenerated and leads to $\lambda = 0$ or $\sigma_u^2 = 0$.

The most commonly employed technique consists of maximizing the marginal likelihood obtained by integrating the AR parameters out of the problem

$$f_X(\mathbf{x}|\lambda, \sigma_u^2) = \int_{\mathbf{a}} f_{X,A}(\mathbf{x}, \mathbf{a}) d\mathbf{a} \quad (20)$$

$$= \int_{\mathbf{a}} f_{X|A}(\mathbf{x}|\mathbf{a}) f_A(\mathbf{a}) d\mathbf{a}. \quad (21)$$

In [4], Kitagawa and Gersch showed that the hyperparameter likelihood is

$$\begin{aligned} L(\lambda, \sigma_u^2) &= f_X(\mathbf{x}|\lambda, \sigma_u^2) \\ &= (2\pi\sigma_u^2)^{-N/2} (\det \Delta_k)^{1/2} \lambda^{N/2} \\ &\quad \cdot (\det (M_1))^{1/2} \exp -\frac{1}{2\sigma_u^2} \mathbf{x}^t (I - M_2) \mathbf{x}, \end{aligned} \quad (22)$$

with M_1 and M_2 given by

$$M_1 = X^t X + \lambda \Delta_k \quad (23)$$

$$M_2 = X^t (X^t X + \lambda \Delta_k)^{-1} X = X^t M_1^{-1} X. \quad (24)$$

The hyperparameters λ and σ_u^2 are finally chosen as the maximizers $\hat{\lambda}$ and $\hat{\sigma}_u^2$ of the likelihood with respect to λ and σ_u^2

$$(\hat{\lambda}, \hat{\sigma}_u^2) = \arg \max f_X(\mathbf{x}|\lambda, \sigma_u^2),$$

or equivalently, the minimizer of the opposite of the logarithm of the likelihood, namely the antilog-likelihood (ALL)

$$(\hat{\lambda}, \hat{\sigma}_u^2) = \arg \min -\log f_X(\mathbf{x}|\lambda, \sigma_u^2).$$

Such a maximization is a 2-D optimization problem, but can be explicitly optimized with respect to $\hat{\sigma}_u^2$ as shown below.

1) σ_u^2 Estimation: The ML estimate for σ_u^2 is derived from (22) by nullifying its derivative with respect to σ_u^2 , and gives the usual empirical estimate for the noise variance

$$\hat{\sigma}_u^2 = \frac{1}{N} \mathbf{x}^t (I - M_2) \mathbf{x}. \quad (25)$$

It should be stressed that this expression is an explicit λ function through the M_2 dependence with λ .

2) λ Estimation: Replacing (21) into the likelihood obtained in (22), one can easily obtain the following for the ALL:

$$\text{ALL}(\lambda) = -\log(\det(M_1)) + N \log(\lambda) + \log(\mathbf{x}^t (I - M_2) \mathbf{x}).$$

The hyperparameter λ is then chosen as the ALL minimizer

$$\hat{\lambda} = \lambda_{ML} = \arg \min \text{ALL}(\lambda).$$

B. Generalized Cross Validation (GCV)

The previous methods are nondeterministic by nature since they are derived from the Bayesian framework. This section is devoted to a deterministic method, namely the GCV, derived from the quadratic regularization interpretation presented in Section III-D. In the class of deterministic solutions for hyperparameters estimation, the very first idea, called λ_χ or λ_{RESID} (see [11]) is based on the probability density function of the residual sum of the squares $Q_0(\mathbf{a})$. Since the signal is Gaussian and the model is linear, the residual sum of the squares follows a χ_N^2 distribution

$$Q_0(\mathbf{a}) \sim \chi_N^2 \sigma_u^2.$$

This motivates the choice of λ_χ as the solution of

$$Q_0(\mathbf{a}) = N \sigma_u^2$$

since $N \sigma_u^2$ is the expected value of the residual distribution. This method has been popular throughout the history of

regularization techniques, but in practical cases when \mathbf{a} is replaced by $\hat{\mathbf{a}}$, the residual sum of the squares does not exactly follow a χ_N^2 distribution.

In order to overcome this obstacle, Thompson *et al.* [11] proposed a better approximation of the distribution of the residual sum of the squares. They introduced an equivalent degree of freedom (EDF) for the residual $N' = N - \text{Tr}\{M_2\}$, and proposed a λ estimation called λ_{EDF} as the solution of

$$Q_0(\mathbf{a}) = \text{Tr}\{I - M_2\} \sigma_u^2.$$

The properties of these estimators have been examined and their performances compared in [8], [10], and [11]. It has been reported that these two techniques, especially the first one, substantially overregularize the solution. Moreover, they absolutely require the knowledge of the noise variance, and have been reported as interesting only if the exact noise variance (or a very good estimate of it) is available.

The cross validation (CV) criterion [9] is an estimate of the MSE, calculated from the data only. The basic principle is very simple and consists of removing one observation x_i from the data, and predicting it on the basis of the regularized solution obtained from the remaining data. The difference between the true and predicted data yields an error, and averaging the prediction errors over all the removed data leads to an approximate MSE. This error is a function of λ , called the cross-validation criterion

$$\text{CV}(\lambda) = \|M(I - M_2)\mathbf{x}\|,$$

where M is a diagonal matrix with the i th diagonal element $1/(1 - a_{ii})$, a_{ii} being the ii th entry of M_2 . The minimum of this criterion should give a good value of the hyperparameters. In fact, we shall not minimize CV, but a modified version of it called GCV. This criterion does not differ greatly from the CV, presents more pleasant properties [9], and takes the form

$$\text{GCV}(\lambda) = \frac{\|(I - M_2)\mathbf{x}\|}{\text{Tr}\{I - M_2\}}.$$

λ is chosen as λ_{GCV} , the minimizer of this criterion with respect to λ

$$\hat{\lambda} = \lambda_{\text{GCV}} = \arg \min \text{GCV}(\lambda).$$

Since the GCV criterion is a function of λ only, this strategy represents an alternative to λ estimation, but does not allow simultaneous estimation of σ_u^2 . The noise power estimate remains the usual estimate given (25) for ML strategy.

V. METHODOLOGY OF THE SIMULATION STUDY

A theoretical comparative study between the performances of the proposed and usual least squares methods is strongly desirable, but is a very difficult task because of the lack of an explicit expression for $\hat{\lambda}$. Theoretical results are available only for small $\hat{\lambda}$ values and in the GCV case [9], [11], but such small values are out of the domain of interest in our practical case. Therefore, a comparative simulated study is required. By Monte Carlo experiments [14], statistical results have been obtained in the following way: a large amount of signals have been simulated, and for each of them the

PSD has been estimated from the different methods and in different conditions. Then, assuming that the large number law is applicable, by averaging these estimates, several estimation characteristics were obtained.

A. Measures of Error and Assessment of Performance

1) *PSD Estimation Performance Measurement*: We now present different classes of comparison measures investigated to compare estimation methods.

a) *Bias, variance, and mean square error*: The first statistical characteristic is the bias $B(f)$ which characterizes the mean difference between the estimation expectation $E(f)$ and the true value $S_x(f)$. The second characteristic is the standard deviation $SD(f)$, i.e., the square root of the estimation variance $V(f)$. It quantifies the mean variability of the estimate around its expected value $E(f)$. Finally, the MSE, $MSE(f) = B(f)^2 + V(f)$, integrating both bias and variance was also used.

b) *Integrated performance index*: As these quantities are functions of frequency, a qualitative and visual criterion is the only possibility for comparing the different estimation methods over the whole frequency domain. In order to avoid this difficulty, as proposed in [15], three indexes integrating the frequency dependence have been used: the integrated bias (IB), the integrated variance (IV), and the integrated mean square error (IMSE). They are calculated by integrating respectively the bias, variance, and MSE over the whole frequency domain.

c) *Logarithmic distance*: The second class of index consists of a measure of the mean dissimilarity between the true and estimated log-spectrum. We first define the log-distance (LD) between the true and estimated PSD

$$LD(f) = [\log S_x(f) - \log \hat{S}_x(f)]^2.$$

Since it is also a function of frequency, it is helpful to build an integrated index

$$ILD = \int_{-1/2}^{1/2} LD(f) df,$$

which is a distance between the true and estimated log-PSD. A global statistical index considers the mean ILD (MILD) under the true probability distribution and takes the following form:

$$MILD = E\{ILD\}.$$

d) *Kullback distance*: The third comparison criterion is entropic. The Kullback dissimilarity measure is the entropy of the true probability distribution with respect to the estimated, and is defined by

$$KD(f_t, f_e) = E_{f_t} \left\{ \frac{f_e(x)}{f_t(x)} \right\},$$

where f_t and f_e are, respectively, the true and estimated probability distribution of the process. In our study, since the processes are Gaussian, the Kullback distance is

$$KD(f_t, f_e) = -\frac{1}{2} (N + \log \det R_e^{-1} R_t - \text{Tr} \{R_e^{-1} R_t\}).$$

Then, the error measure is the mean Kullback distance (MKD)

$$MKD = E_{f_t} \{K(f_t, f_e)\}.$$

2) *λ Estimation Performances Measurement*: In order to assess and compare the different methods of estimation for the hyperparameters, the first step is to determine a reference, and this is done in the following way. As the true PSD is known in our simulation study, the “true” λ can be evaluated as λ_{IMSE} , the minimizer of the IMSE defined in the previous section. Such a λ_{IMSE} is not an actual estimate since it requires the knowledge the true PSD and a large amount of simulated signal. Nevertheless, it is an interesting reference in order to assess the two actually practicable methods i.e., ML and GCV (which estimate the hyperparameters from a single signal realization and, of course, without knowing the true PSD). This λ_{IMSE} will be considered as the true λ value in order to evaluate bias B_λ , variance V_λ , and MSE MSE_λ .

B. Theoretical Spectrum and Simulated Signals

1) *Choice of a Field of Application*: On the one hand, the studied method concerns several fields of application, such as nondestructive testing, attenuation measurements, Doppler imaging, etc., and on the other hand, a simulation study requires the choice of a spectral shape and simulation model related to the considered application. Among the applications, we have made the choice of ultrasonic Doppler velocimetry, because it is one of the most difficult problems for several reasons:

- 1) Ultrasonic Doppler signals are known to be very complex and difficult to analyze since they are the results of a highly complex nonlinear backscattering phenomenon [16].
- 2) In ultrasound Doppler velocimetry, important variations of the signal-to-noise ratio (SNR) during the cardiac cycle are observed (the SNR falls down during diastole due to the clutter rejection filter).
- 3) The reduction of window data length is crucial in order to perform a more time-resolvent analysis of rapidly varying nonstationary flows, as pointed out by several authors, e.g., [17]–[19].

In Doppler spectral analysis, a seminal paper by Vaitkus *et al.* [20] compares numerous spectral estimation methods, parametric or not parametric. Comparison of several indexes, including IMSE for these methods, has indicated a slight superiority for the least squares AR method, especially for low SNR, while keeping an efficient computational cost. Our study follows that of Vaitkus *et al.*, and its objective is to show that the RLS AR methods achieve a new improvement with respect to the least squares AR methods. Particular attention will be paid to improvement in time resolution of fine details of the rapidly varying flow by processing very short windows.

2) *Choice of Simulation Model and Spectral Shape*: In order to perform a simulation study, two other choices are required: the first deals with spectral shape and the second concerns simulations of signals with the chosen PSD.

a) *Spectral shape*: A currently used spectral shape (see Fig. 1) is proposed in [21] and closely approximates the PSD seen around the peak systole under normal flow conditions in a carotid artery. This shape is used by several authors and seems to be recognized as a typical spectral shape [15], [22].

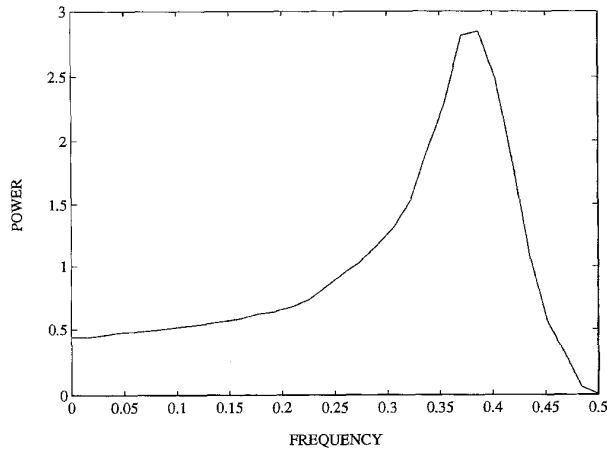


Fig. 1. Theoretical power spectral density approximating that typically seen around peak systole, used for simulated signals.

b) Simulation model: In a detailed study of the backscattered ultrasound from blood insonated by a continuous monochromatic wave, Mo and Cobbold in [21] showed that, under acceptable assumptions, the Doppler signal $x(t)$ is a zero-mean stationary Gaussian process.

Mo and Cobbold proposed the following simulation model [23] derived from their previously mentioned [21] physical study, and which can be assumed to faithfully reproduce the complex physical nature of the real signals. The chosen PSD function $S_x(f)$ shown in Fig. 1 is sampled on a fine frequency grid and each sample is multiplied by a χ^2 random variable. The square root of each obtained variable is multiplied by a uniformly distributed complex phase term, and finally an inverse Fourier transform yields to the simulated signal $x(n)$. It can be shown that $x(n)$ is a zero-mean stationary Gaussian process having the given PSD $S_x(f)$.

c) Simulated signals: The simulation model and the DSP described above have been used to generate 500 signals of 256 samples from the theoretical PSD shown in Fig. 1 in the same manner as [15] and [22]. The following study was made from 16–256 of these samples in order to compare the methods for different window data lengths. The effect of an additive Gaussian white noise on the estimation has also been investigated for an SNR from –30 to 30 dB.

VI. RESULTS AND DISCUSSION

In this simulation study, particular attention is paid to the case in which 16 samples are observed with an SNR of 20 dB. In this situation, both λ (Section VI-A) and PSD estimation properties (Section VI-B) are investigated. Complementary results are given in Section VI-C1), VI-C2), and VI-C3) which investigate the influence of smoothness order, window data length, and SNR.

A. λ Estimation Properties

In this section, a study of λ estimation properties is presented in the case of 16 samples with an SNR of 20 dB and first smoothness order. We first find out the best λ value in

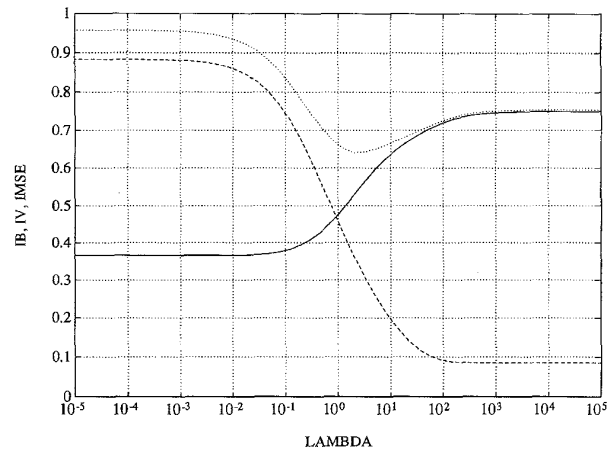


Fig. 2. Integrated bias (solid line), integrated variance (dashed line), and IMSE (dotted line) as a function of λ .

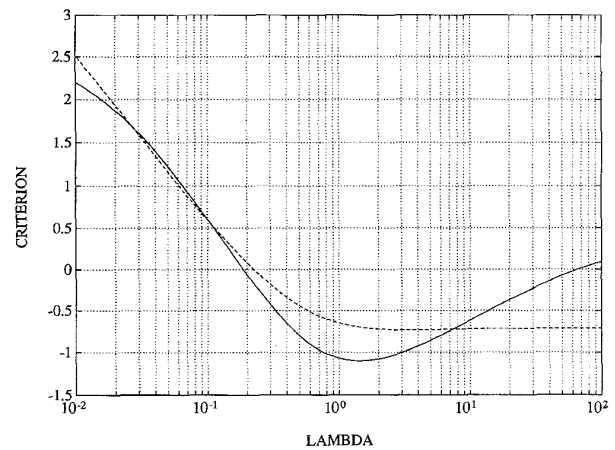


Fig. 3. The two λ estimation criteria obtained from Sig1: GCV (solid line) and ALL (dashed line).

Section VI-A1), then we compare the two estimation methods within each other and with the best λ value in Section VI-A2).

1) Determination of λ_{IMSE} : Simulations were made on a logarithmic grid of 100 values from $\lambda = 10^{-5}$ to $\lambda = 10^5$, and the indexes IB, IV, and IMSE are presented in Fig. 2.

For λ small enough ($\lambda < 10^{-3}$), the solution is not regularized, i.e., no prior knowledge is introduced. It is equivalent to say that the least squares solution is found, i.e., the solution is entirely based on data. Therefore, as expected, the statistical properties are still those pointed out in Section II: even if the estimation is low biased, the strong variance leads to unreliable results.

At the other extreme, for λ high enough ($\lambda > 10^3$), the solution is infinitely regularized, i.e., almost no data are taken into account in the estimation. It is equivalent to say that the solution is entirely prior based, i.e., the estimated spectrum is completely smooth and constant over the whole frequency domain. Therefore, as expected, the statistical properties of the estimation are opposite: the estimation has a very low variance, but the very strong bias leads to unusable results. Between these two extreme situations, IMSE shows a mini-

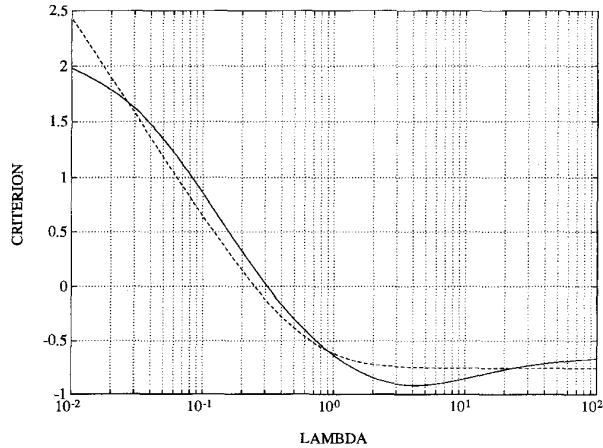


Fig. 4. The two λ estimation criteria obtained from Sig2: GCV (solid line) and ALL (dashed line).

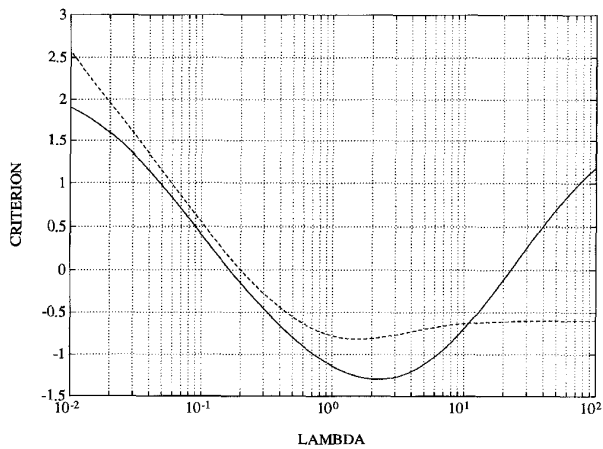


Fig. 5. The two λ estimation criteria obtained from Sig3: GCV (solid line) and ALL (dashed line).

imum $IMSE = 0.64$ for $\lambda_{IMSE} = 2.02$. This value $\lambda_{IMSE} = 2.02$ is assumed to be the true value of λ in the following comparisons.

2) *Comparison of the λ Estimates:* As mentioned above, in practical cases, λ must be estimated from each signal, and of course without knowing the actual PSD. Such a value is now sought by computing the two criteria (ALL and GCV) on a logarithmic grid of 100 values between 10^{-2} and 10^2 . In order to understand the characteristics of the λ estimation methods presented in Section IV, we first describe results obtained from three particular signals, and in a second step, we provide statistical characteristics.

a) *Estimation from typical signal realizations:* First, as an example, Figs. 3, 4, and 5 show typical forms of the two estimation criteria: ALL [ALL(λ)] and GCV [GCV(λ)] for three particular signals, referred to as Sig1, Sig2, and Sig3. For Sig1 and Sig3, one can observe a minimum for each criterion (see Table I). On the contrary, for Sig2, one can see that the ALL has no minimum or, equivalently, the minimum is rejected to $\lambda = \infty$ ($\lambda = 100$ in our simulation).

TABLE I
 λ ESTIMATION FROM THE GIVEN SIGNAL SIG1, SIG2, AND SIG3, USING THE TWO ESTIMATION METHODS: ML AND GCV

	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{GCV}$
Sig1	3.51	1.39
Sig2	100	4.23
Sig3	1.52	2.21

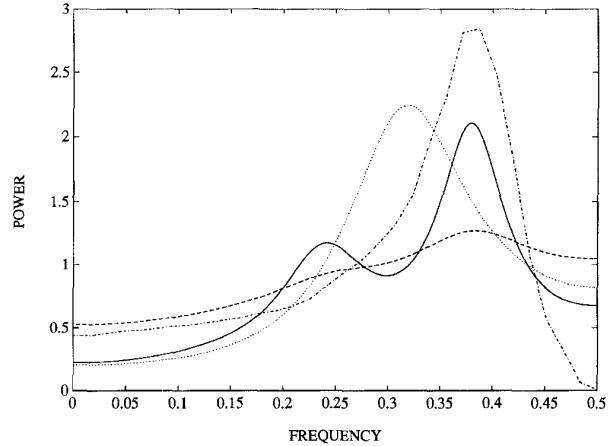


Fig. 6. Comparison of the two practicable regularized methods [$\hat{\lambda} = \lambda_{GCV}$ (solid line) and $\hat{\lambda} = \lambda_{ML}$ (dashed line)] and the usual least squares method (dotted line) from Sig1. The theoretical power spectral density is also plotted (dash-dot line).

The corresponding estimated spectra are given in Figs. 6, 7, and 8. As expected, for high $\hat{\lambda}$ values, the PSD estimate is smooth, at the limit for infinite $\hat{\lambda}$ values ($\hat{\lambda} = 100$ is large enough in our simulations) the PSD is entirely smooth and constant over the whole frequency domain. On the contrary, for smaller $\hat{\lambda}$ values, the PSD estimate is not smooth.

For some signal realizations (Sig2, for instance), the estimated regularization parameter is $\lambda = \infty$, so the regularized PSD estimate is constant over the whole frequency domain. It is important to note that at the same time, for these particular signal realizations, the least squares solution is also unreliable. When information is almost absent from the measured signal (for Sig2, for instance), the least squares solution yields a PSD estimate, which is completely different from the true PSD and presents almost any shape. In such cases, as there is so little information in the measured signal, the regularized solution leads to the prior solution, i.e., a flat spectrum. It is a noticeable advantage of the regularized method: when there is not enough information in the signal, instead of giving anything as a PSD estimate, which is the usual approach, a fixed ultrasmooth solution (regularized) is given. In this sense, it is an argument in favor of the regularized method that it can discriminate between informative and uninformative data.

b) *Statistical characteristics:* Averaged results for the 500 simulated signals are now presented. For each simulated signal, the criteria have been computed and the two λ

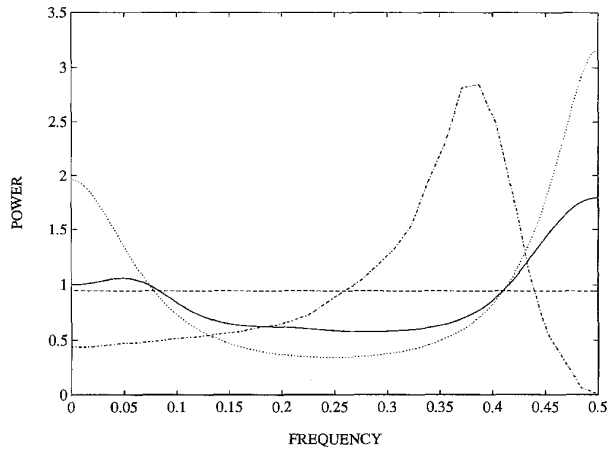


Fig. 7. Comparison of the two practicable regularized methods [$\hat{\lambda} = \lambda_{GCV}$ (solid line) and $\hat{\lambda} = \lambda_{ML}$ (dashed line)] and the usual least squares method (dotted line) from Sig2. The theoretical power spectral density is also plotted (dash-dot line).

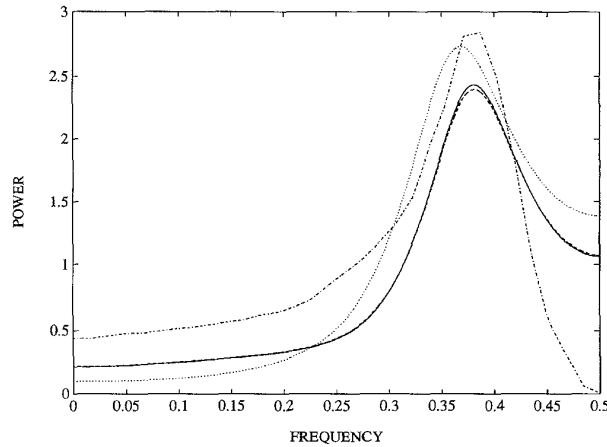


Fig. 8. Comparison of the two practicable regularized methods [$\hat{\lambda} = \lambda_{GCV}$ (solid line) and $\hat{\lambda} = \lambda_{ML}$ (dashed line)] and the usual least squares method (dotted line) from Sig3. The theoretical power spectral density is also plotted (dash-dot line).

estimates ($\hat{\lambda}_{ML}, \hat{\lambda}_{GCV}$) have been calculated. Fig. 9 shows the $\hat{\lambda}$ repartition for the two estimation methods.

As mentioned above for some signal realizations, the criteria may have no minimum (or equivalently, the minimum is rejected to $\lambda = \infty$). One can see that from the 500 simulated signals, the proportion of criteria without minimum is 30% when using the ML estimation criterion and 25% when using the GCV criterion. The GCV estimation procedure seems to be more robust, and in this sense is the best one.

In order to give a more accurate analysis, we eliminate the case of criterion without minimum from the results given here. The indexes of interest given in Table II can then be calculated. Table II gives the $\hat{\lambda}$ expected value, bias, standard deviation, and mean square error assuming that λ_{IMSE} is the true value.

These results first show that both the ML and GCV methods overestimate the parameter λ several times and, as a direct consequence, oversmooth the estimated spectra. This point needs

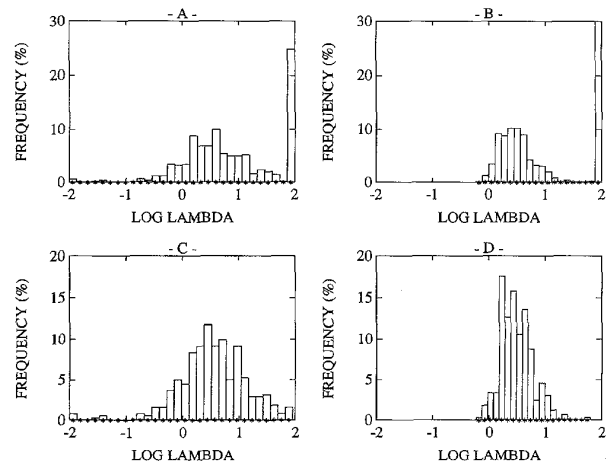


Fig. 9. Histogram of $\hat{\lambda}$. (a) GCV criterion, (b) ML, as obtained from the computations, (c) GCV criterion, and (d) ML, after elimination of the infinite values.

TABLE II
 λ ESTIMATION STATISTICAL PROPERTIES FOR THE TWO METHODS: ML AND GCV

	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{GCV}$
Expectation	3.93	8.12
Bias	1.91	6.10
Standard Deviation	2.12	3.66
Mean Square Error	2.85	7.11

to be made precise. Although the regularization parameter is overestimated, it remains in the correct range, i.e., around the minimum of the IMSE presented in Fig. 2.

Moreover, since the IMSE is slowly varying around its minimum, the error is rather indifferent to the variations of λ as long as it remains in the correct decade (here between 1 and 10). The slowly varying character of the error with respect to λ is a strong argument for the robustness of the estimation method. Finally, one of the interests of the estimation method is to automatically find the range of the IMSE minimum without knowing the true spectrum, using one signal only.

Now comparing ML and GCV, the results of Table II also show that the ML method overestimates λ slightly less than the GCV method, hence there is less oversmoothing of the estimated spectra. From this point of view, one can say that the ML estimation procedure seems to be the best one.

A definitive comparison of the two λ estimation methods is given in terms of recovering the spectrum, by comparing the PSD estimation performances (see Section VI-B).

B. PSD Estimation Performances

In order to give an assessment of an improvement with respect to the usual least squares solution, we will first find out the AR order giving the least error in Section VI-B1), and then we will show that the regularized method yields better

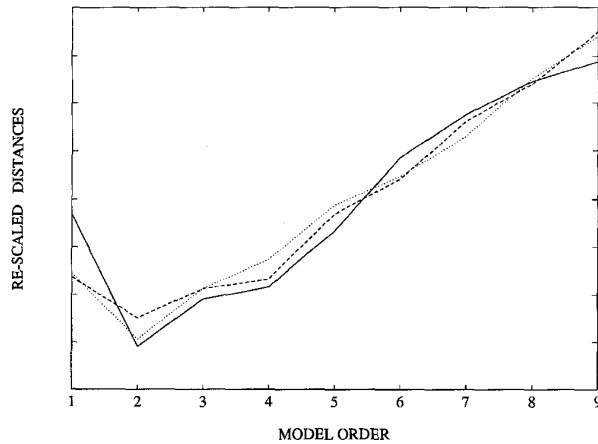


Fig. 10. Kullback distance (solid line), logarithmic distance (dashed line), and IMSE (dotted line) as a function of AR order.

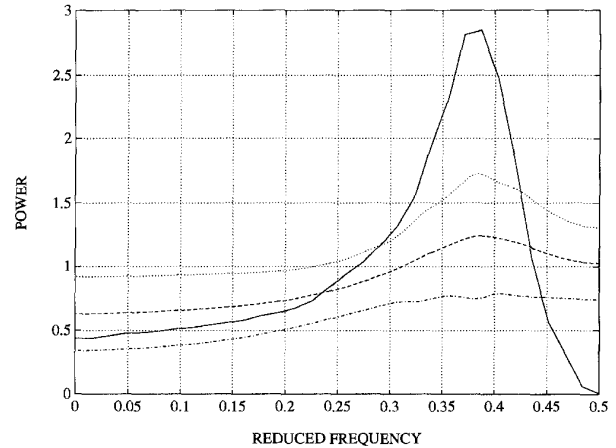


Fig. 12. Regularized least squares solution with λ_{ML} . Actual DSP (solid line), estimator expectation $E(f)$ (dashed line), $E(f) + SD(f)$ (dotted line), and $E(f) - SD(f)$ (dash-dotted line).

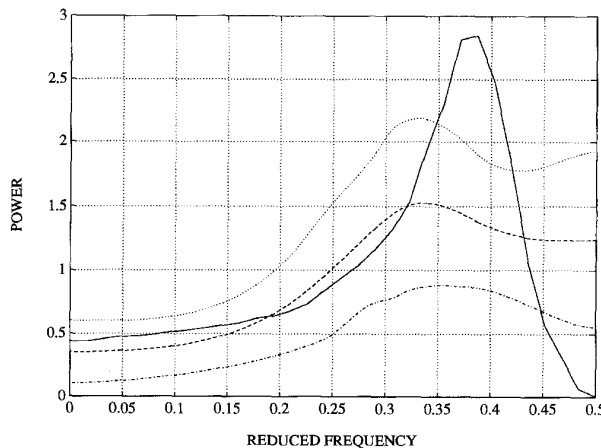


Fig. 11. Best selected usual least squares solution. Actual DSP (solid line), estimator expectation $E(f)$ (dashed line), $E(f) + SD(f)$ (dotted line), and $E(f) - SD(f)$ (dash-dotted line).

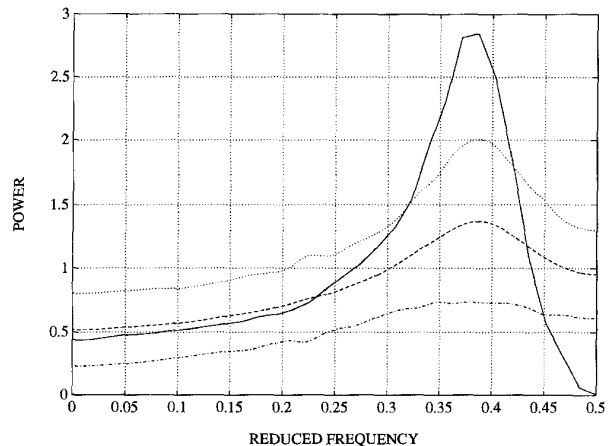


Fig. 13. Regularized least squares solution with λ_{GCV} . Actual DSP (solid line), estimator expectation $E(f)$ (dashed line), $E(f) + SD(f)$ (dotted line), and $E(f) - SD(f)$ (dash-dotted line).

results in Section VI-B2). The main point of the study is to make a comparison between the usual least squares solution (knowing the best order) and the regularized solution (without knowing the best regularization parameter).

1) *Best Least Squares Model Order Selection:* Since the problem of automatic choice for the model order is a difficult one, especially for a short data set, an exhaustive study has been performed and the best order has selected in the following manner.

For each possible order ($p = 1$ to 16), the three indexes IMSE, MKD, and MILD have been calculated and results are given in Fig. 10. Among all the possible orders p , the order $p_0 = 2$ minimizes the three indexes. Hence, the usual least squares solution with any model order selection criterion cannot give better results than $p = 2$. In this sense, the second-order AR model is the best for least squares estimation methods.

2) *Assessment of an Improvement of Performance with Respect to the Best Least Squares Solution:* We now present a comparison between the best usual least squares PSD esti-

TABLE III
IMSE, MILD, MKD FOR THE THREE METHODS: LEAST SQUARES (LS), RLS WITH MAXIMUM LIKELIHOOD (ML), AND GCV FOR λ ESTIMATION

	LS	RLS (ML)	RLS (GCV)
IMSE	0.75	0.69	0.68
MILD	1.01	0.86	0.87
MKD	3.12	2.25	2.23

mation method ($p = 2$) and the regularized method. In the same way as previously, $\hat{S}_x(f)$ has been calculated using the regularized method with the two practicable λ estimation criteria (ML and GCV), and for the first smoothness order. As an example with particular signals, Sig1, Sig2, and Sig3, Figs. 6, 7, and 8 show the estimated PSD.

Mean estimate $E(f)$ and variability $E(f) \pm SD(f)$ are shown in Figs. 11, 12, and 13, respectively, for the best usual

TABLE IV
IB, IV, AND IMSE USING THE REGULARIZED SOLUTION FOR THREE SMOOTHNESS ORDERS (ZERO, ONE, AND TWO) AND WITH THE TWO λ ESTIMATION METHODS

	$k = 0$		$k = 1$		$k = 2$	
	ML	GCV	ML	GCV	ML	GCV
IB	0.68	0.59	0.61	0.56	0.31	0.55
IV	0.24	0.38	0.30	0.38	0.66	0.43
IMSE	0.71	0.70	0.69	0.68	0.72	0.70

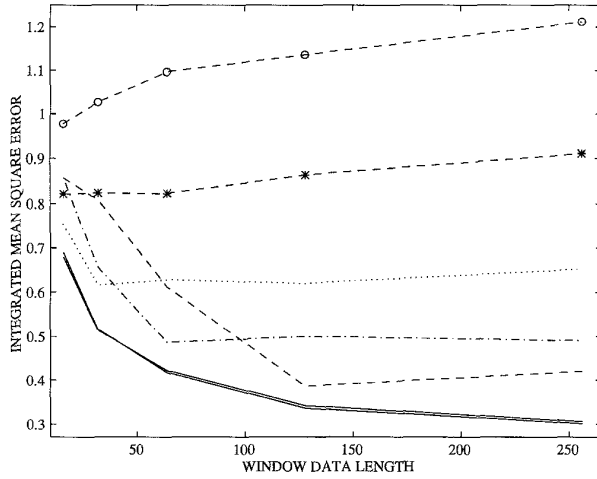


Fig. 14. IMSE for different model orders as a function of window data length. Least squares solution with $P = N/2$ (circle line), $P = N/4$ (star line), $P = N/8$ (dotted line), $P = N/16$ (dash-dotted line), and $P = N/32$ (dashed line). RLS solution with $\hat{\lambda} = \hat{\lambda}_{ML}$ and $\hat{\lambda} = \hat{\lambda}_{GCV}$ (solid lines almost coincident; see Table V for details).

least squares criterion and the regularized criterion with the two practicable λ estimation methods. The integrated indexes are reported in Table III for better readability.

The same indexes obtained from the regularized methods with the two λ estimation methods are shown to be smaller than those obtained for the best least squares solution defined in Section VI-B1). Since advantage has been given to the least squares method in the preliminary study giving $p = 2$ as the best model order, it appears that the proposed regularized method gives better statistical results than the usual least squares method.

Comparing GCV and ML estimation methods, one can see from Table III and from Figs. 12 and 13 that the two methods for hyperparameter estimation behave almost identically with a slight advantage to the GCV method. More accurate conclusions about the comparison between ML and GCV methods are difficult to draw on the basis of this result.

C. Influence of the Analysis Parameters

In the previous study, the smoothness order was fixed at one, the window data length at 16 samples, and the SNR at 20 dB. From this situation, we are now successively varying

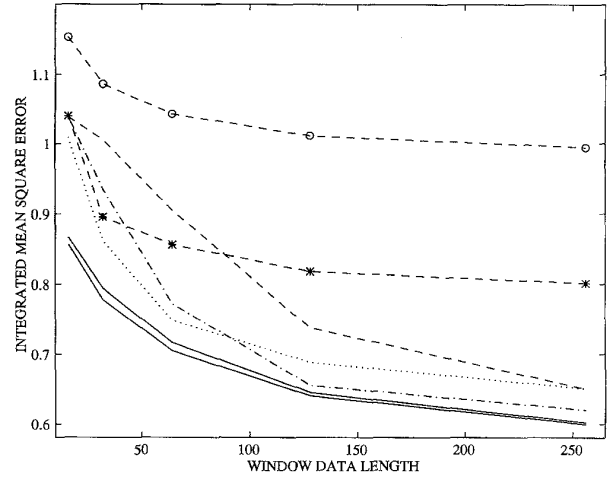


Fig. 15. MILD for different model order as a function of window data length. Least squares solution with $P = N/2$ (circle line), $P = N/4$ (star line), $P = N/8$ (dotted line), $P = N/16$ (dash-dotted line), and $P = N/32$ (dashed line). RLS solution with $\hat{\lambda} = \hat{\lambda}_{ML}$ and $\hat{\lambda} = \hat{\lambda}_{GCV}$ (solid lines almost coincident; see Table VI for details).

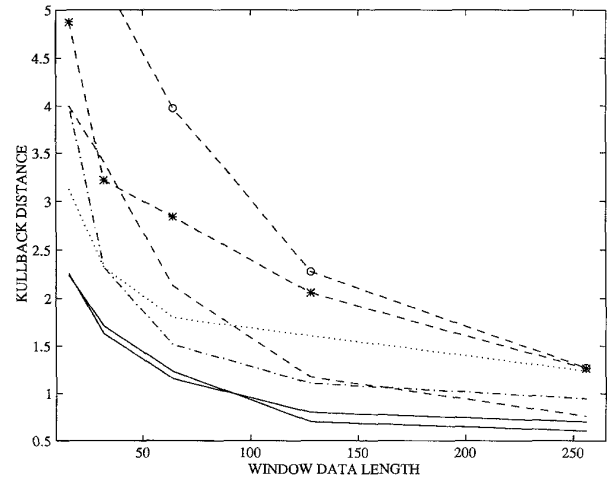


Fig. 16. MKD for different model order as a function of window data length. Least squares solution with $P = N/2$ (circle line), $P = N/4$ (star line), $P = N/8$ (dotted line), $P = N/16$ (dash-dotted line), and $P = N/32$ (dashed line). RLS solution with $\hat{\lambda} = \hat{\lambda}_{ML}$ and $\hat{\lambda} = \hat{\lambda}_{GCV}$ (solid lines almost coincident; see Table VII for details).

one of the three parameters (smoothness order, window data length, and SNR) while keeping the two others constant.

1) *Smoothness Order*: Influence of the smoothness order has been investigated and Table IV gives results for zero, one, and two smoothness orders. This table shows that the first smoothness prior gives very slightly better performances in terms of IMSE. On the other hand, results show low sensitivity to this parameter and the error index remains smaller than the best least squares index.

2) *Data Window Length*: In a first step, for several window data lengths ($N = 16, 32, \dots, 256$), the estimations $\hat{S}_x(f)$ have been calculated using the usual least squares solution with several AR orders p ($p = N/2, N/4, \dots, N/16$), and the three indexes IMSE, MILD, and MKD have been computed. Figs. 14, 15, and 16 show the curves corresponding to the

TABLE V
IMSE AS A FUNCTION OF THE WINDOW DATA LENGTH. NUMERICAL RESULTS OF FIG. 14

N	RLS		LS				
	λ_{GCV}	λ_{ML}	$p = N/2$	$p = N/4$	$p = N/8$	$p = N/16$	$p = N/32$
16	0.68	0.69	0.98	0.82	0.75	0.86	0.86
32	0.51	0.52	1.03	0.82	0.62	0.66	0.81
64	0.42	0.42	1.10	0.82	0.63	0.49	0.61
128	0.34	0.34	1.14	0.86	0.62	0.50	0.39
256	0.31	0.30	1.21	0.91	0.65	0.49	0.42

TABLE VI
MILD AS A FUNCTION OF THE WINDOW DATA LENGTH. NUMERICAL RESULTS OF FIG. 15

N	RLS		LS				
	λ_{GCV}	λ_{ML}	$p = N/2$	$p = N/4$	$p = N/8$	$p = N/16$	$p = N/32$
16	0.86	0.87	1.15	1.04	1.01	1.04	1.04
32	0.78	0.79	1.09	0.90	0.86	0.93	1.01
64	0.70	0.72	1.04	0.86	0.75	0.77	0.90
128	0.64	0.65	1.01	0.82	0.69	0.66	0.74
256	0.60	0.60	0.99	0.80	0.65	0.62	0.65

TABLE VII
MKD AS A FUNCTION OF THE WINDOW DATA LENGTH. NUMERICAL RESULTS OF FIG. 16

N	RLS		LS				
	λ_{GCV}	λ_{ML}	$p = N/2$	$p = N/4$	$p = N/8$	$p = N/16$	$p = N/32$
16	2.25	2.23	7.90	4.87	3.12	4.00	4.00
32	1.63	1.71	5.31	3.21	2.32	2.32	3.41
64	1.15	1.23	3.98	2.84	1.80	1.51	2.13
128	0.80	0.70	2.28	2.06	1.61	1.11	1.17
256	0.70	0.60	1.27	1.26	1.23	0.95	0.76

different AR orders as a function of the window data length, while Tables V, VI, and VII show the corresponding numerical results.

In a second step, the same indexes have been computed from the same simulated signals and window data lengths, but using the regularized solution with $\hat{\lambda} = \hat{\lambda}_{GCV}$ and $\hat{\lambda} = \hat{\lambda}_{ML}$. Results are also presented in Figs. 14, 15, and 16, and in numerical form in Tables V, VI, and VII.

From the results presented in the figures and tables, one can observe that in any case, the RLS solution gives the least error with a slight advantage over the GCV method.

From another standpoint, for given performances, e.g., $IMSE = 0.4$ (respectively, 0.5), the usual method requires $N = 128$ (respectively, 64) data, while the regularized method can achieve the same performance from only $N = 64$ (respectively, 32) data. Hence the proposed method can achieve more resolution along the time axis of rapidly varying nonstationary flow.

3) *Signal-to-Noise Ratio*: For several SNR's from -30 to 30 dB from 16 samples of the 500 simulated signals, estimations of $S_x(f)$ have been calculated using the best usual least squares solution and using the two RLS solutions (with $\hat{\lambda}_{ML}$

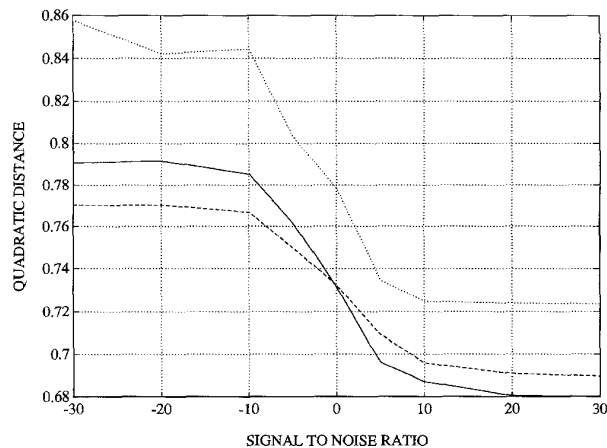


Fig. 17. IMSE as a function of SNR. Usual least squares solution (dotted line), regularized solution and λ_{GCV} (solid line), and regularized solution and λ_{ML} (dashed line).

and $\hat{\lambda}_{GCV}$). The IMSE indexes are shown in Fig. 17. From the simulations results, it can be seen that for each SNR, the regularized solution yields better estimation qualities in terms of IMSE. Moreover, slightly better performance for the ML method for a negative SNR, and better performances for the GCV method for a positive SNR can be seen.

From another point of view, Fig. 17 shows that the performance of the regularized method for any positive SNR is better than or equivalent to the best performance of the least squares method (obtained with SNR = 30 dB), indicating an important gain in terms of SNR.

For given performances (e.g., IMSE = 0.76), the regularized method enables the exploitation of signals corrupted down to -5 dB, while the usual method cannot accept an SNR lower than +3 dB. Hence a gain on noise power (here, 8 dB) is achieved and allows a deeper flow analysis.

VII. CONCLUSIONS

We have addressed the problem of spectral estimation of a zero-mean stationary Gaussian process when only a short span of data is available for analysis (down to 16 observations). In such a situation, usual AR estimation strategies, such as ML or least squares, enforce the estimation of a parsimonious model which precludes the description of a large class of PSD. The Bayesian approach presented here and initially proposed by Kitagawa and Gersch [4] alleviates this limitation since it admits the robust estimation of long AR parameter vectors (typically 16 parameters from 16 observations).

We have performed a large simulation study in order to compare performances with respect to those of the usual method. We have compared two estimation methods of the fundamental regularization parameter: ML and GCV methods in terms of recovering a known PSD. The conclusion of this study is that the GCV method performs slightly better than the ML method. Nevertheless, this result should be taken with caution since the difference observed is very slender.

A statistical comparison of the presented Bayesian method versus the usual method has been achieved in different sit-

uations, varying the data span length and the SNR. The performances have been measured using various indexes: quadratic, logarithmic, and Kullback distances. The simulations confirm and extend the results of Kitagawa and Gersch in different situations and using new performance measures. The result is that the Bayesian method with automatic tuning of the hyperparameters yields better indexes than the usual least squares method, whatever the model order. The conclusion is then in favor of the Bayesian approach, at least for the class of signals within the scope of this paper. Nevertheless, we believe that as long as a spectral smoothness information is available for estimation, the presented method is able to provide better results than the nonregularized method.

ACKNOWLEDGMENT

The authors wish to thank J. Idier for numerous and helpful discussions. They are also grateful to Mr. Grün and Mrs. Groen for their expert editorial assistance.

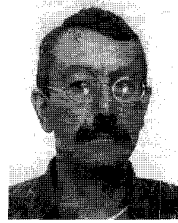
REFERENCES

- [1] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380–1419, 1981.
- [2] S. M. Kay, *Modern Spectral Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [3] S. L. Marple, *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [4] G. Kitagawa and W. Gersch, "A smoothness priors long AR model method for spectral estimation," *IEEE Trans. Automat. Contr.*, vol. AC-30, pp. 57–65, 1985.
- [5] S. M. Kay, "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-21, pp. 56–65, 1983.
- [6] D. T. Pham, "Maximum likelihood estimation of the autoregressive model by relaxation on the reflection coefficients," *IEEE Trans. Signal Processing*, vol. SP-36, pp. 1363–1367, 1988.
- [7] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structure and problems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 2024–2036, 1989.
- [8] N. Fortier, G. Demoment, and Y. Goussard, "Comparison of GCV and ML methods of determining parameters in image restoration by regularisation," *J. Visual Commun. Image Representation*, vol. 4, pp. 157–170, 1993.
- [9] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, pp. 215–223, 1979.
- [10] P. Hall and D. M. Titterton, "Common structure of techniques for choosing smoothing parameter in regression problems," *J. Roy. Statist. Soc. B*, vol. 49, pp. 184–198, 1987.
- [11] A. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterton, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 326–339, 1991.
- [12] D. M. Titterton, "Common structure of smoothing techniques in statistics," *Int. Statist. Rev.*, vol. 53, pp. 141–170, 1985.
- [13] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst. Sci. Cybernetics*, vol. SSC-4, pp. 227–241, 1968.
- [14] Y. A. Shreider, *The Monte Carlo Method*, vol. 87. New York: Pergamon, 1966.
- [15] P. J. Vaitkus and R. S. C. Cobbold, "A comparative study and assessment of Doppler ultrasound spectral estimation techniques. Part II: Method and results," *Ultrasound Med. Biol.*, vol. 14, pp. 673–688, 1988.
- [16] B. A. J. Angelsen, "A theoretical study of the scattering of ultrasound from blood," *IEEE Trans. Biomed. Eng.*, vol. BME-27, pp. 61–67, 1980.
- [17] R. I. Kitney and D. P. Giddens, "Analysis of blood flow velocity waveforms by phase shift averaging and autoregressive spectral estimation," *J. Biomed. Eng.*, vol. 105, pp. 398–401, 1983.
- [18] H. E. Talhami and R. I. Kitney, "Maximum likelihood frequency tracking of the audio pulsed Doppler ultrasound signal using a Kalman filter," *Ultrasound Med. Biol.*, vol. 14, pp. 599–609, 1988.

- [19] Z. Guo, J.-G. Durand, and H. C. Lee, "Comparison of time-frequency distribution techniques for analysis of simulated Doppler ultrasound signals of the femoral artery," *IEEE Trans. Biomed. Eng.*, vol. 41, pp. 332-342, 1994.
- [20] P. J. Vaitkus and R. S. C. Cobbold, "A comparative study and assessment of Doppler ultrasound spectral estimation techniques. Part I: Estimation methods," *Ultrasound Med. Biol.*, vol. 14, pp. 661-672, 1988.
- [21] L. Y. L. Mo and R. S. C. Cobbold, "Speckle in continuous wave Doppler ultrasound spectra: A simulation study," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. UFFC-33, pp. 747-753, 1986.
- [22] T. Loupas and W. N. McDicken, "Low-order complex AR models for mean and maximum frequency estimation in the context of Doppler color flow mapping," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 37, pp. 590-601, 1990.
- [23] L. Y. L. Mo and R. S. C. Cobbold, "A stochastic model of the backscattered Doppler ultrasound from blood," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. UFFC-33, pp. 20-27, 1986.



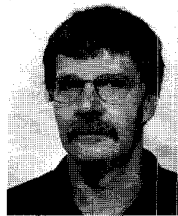
Jean-François Giovannelli was born in Béziers, France, in 1966. He graduated from the École Nationale Supérieure de l'Électronique et de ses Applications, in 1990. He received the Doctorat degree in physics at the Laboratoire des Signaux et Systèmes, Université de Paris-Sud, Orsay, in 1995. He is presently with the Laboratoire des Signaux et Systèmes as a Postdoctoral Student. He is interested in the Bayesian method for inverse problems in signal and image processing, mainly in the field of spectral estimation.



Guy Demoment was born in France, in 1948. He graduated from the École Supérieure d'Électricité, in 1970, and received the Doctorat d'État degree in physics from the Université de Paris-Sud, Orsay, in 1977.

From 1971 to 1977 he was with the Service des Mesures de l'École Supérieure d'Électricité. From 1977 to 1988 he was with the Centre National de la Recherche Scientifique, assigned to the Laboratoire des Signaux et Systèmes. He is presently a Professor in the Department of Physics at the Université de Paris-Sud, Orsay. After some works on biological system modeling, his interests shifted toward inverse problems in signal and image processing.

Dr. Demoment is a member of AF5.5 and SRV.



Alain Herment was born in Paris, France, in 1948. He graduated from ISEP Engineering School of Paris, in 1971. He received the Doctorat d'État degree in physics, in 1984.

Initially, he worked as an Engineer at the Centre National de la Recherche Scientifique. In 1977 he commenced a career as a Researcher at the Institut National pour la Santé et la Recherche Médicale (INSERM). He is currently in charge of the Department of Cardiovascular Imaging at the INSERM Unit 66, Hôpital Pitié, Paris, France. He is interested in signal and image processing for extracting morphological and functional information from image sequences, mainly in the fields of ultrasound investigations, X-ray computed tomography, and digital angiography.