Sample Cumulants of Stationary Processes: Asymptotic Results

José A. R. Fonollosa, *Member, IEEE*

***Abstract*—In this paper, we present the formulas of the covariances of the second-, third-, and fourth-order sample cumulants of stationary processes. These expressions are then used to obtain the analytic performance of FIR system identification methods as a function of the coefficients and the statistics of the input sequence. The lower bound in the variance is also compared for different sets of sample statistics to provide insight about the information carried by each sample statistic. Finally, the effect that the presence of noise has on the accuracy of the estimates is studied analytically. The results are illustrated graphically with plots of the variance of the estimates as a function of the parameters or the signal-to-noise ratio. Monte Carlo simulations are also included to compare their results with the predicted analytic performance.**

I. INTRODUCTION

HIGHER order statistics have been shown to be very useful in applications were non-Gaussian signals are present. It is well known that second-order statistics (i.e., correlation) are phase blind and that only by using, implicitly or explicitly, higher order statistics is it possible to estimate the true phase of a linear process or the parameters of a nonlinear model [1], [2]. Higher order statistics have also been applied in problems were colored Gaussian noise is present since cumulant-based methods can still provide asymptotically unbiased estimates in this case. In addition, it is also important to mention that the use of higher order statistics can drastically improve the performance of methods based only on second-order statistics.

Several cumulant-based system identification methods are now available in the literature. Nevertheless, in almost all the cases, the performance of these developed methods has been evaluated only through Monte Carlo simulations and for a limited number of cases. Clearly, these simulations are insufficient to predict the general behavior of cumulant-based algorithms. The main purpose of this paper is to develop the analytic tools required to perform the asymptotic performance evaluation of parametric methods based on second-, third-, and fourth-order sample cumulants. Only the analysis of parametric methods is considered in this paper. Nonparametric polyspectral methods can be analyzed directly from the asymptotic theory of estimates of higher order spectra [9], [10].

Manuscript received February 20, 1993; revised September 28, 1994. This work was supported by CIRIT, Generalitat de Catalunya, Spain, under grant BE9132. The associate editor coordinating the review of this paper and approving it for publication was Prof. Mysore Raghuveer

The author is with the Universitat Politècnica de Catalunya, ETSE Telecomunicació, Barcelona, Spain.

IEEE Log Number 9409766.

The first and most difficult step encountered in the analytic study of cumulant-based methods is the calculation of the covariances of higher order moment estimates. This task was addressed in part in [3], where the asymptotic covariances of second- and third-order sample moments of stationary linear processes were derived. Here, those results are completed with the covariances of sample fourth-order moments, and these new expressions are then used to obtain both the covariances of third- and fourth-order sample cumulants. Moreover, we present expressions that are valid for any stationary process, for example, linear processes in noise.

Even if the process is known to be zero mean, the sample cumulants are usually computed after subtracting the sample mean from the given samples of the process. This subtraction does not modify the asymptotic covariances of second-order statistics, but it does affect the covariances of the estimated cumulants. Our results show that the covariances of the sample cumulants are simpler than the covariances of the sample moments, i.e., they have a significantly reduced number of terms.

The derivation of the covariances requires the computation of up to eighth-order moments, which involves several hundred terms. Although it is possible to manipulate these terms using a compact notation [7], hand calculation and simplification of explicit expressions is quite discouraging. Symbolic algorithms for *Mathematica* [8] have been used to avoid hand derivations and obtain simplified expressions. With this symbolic math package, it is not difficult to obtain the general expression of moments of any order as a function of the cumulants of the process. Then, we can program the steps followed in [3], as well as additional simplification rules, to obtain expressions of the covariances of sample moments. Using the relation between sample cumulants and sample moments, the analytic simplified expression of the covariances of sample cumulants can also be obtained.

II. MOMENTS AND CUMULANTS

Let $I = \{x_1, x_2, \cdots, x_m\}$ be a set of random variables. The moment of I, i.e., the expectation of the product of the elements in I will be denoted as

$$\mathrm{mom}[I] = \mathrm{mom}[x_1, x_2, \cdots, x_m] = E[x_1 x_2 \cdots x_m].$$

With this notation, cumulants can be defined as a function of

1053-587X/95$04.00

moments with the expression [7], [14]

$$\text{cum}[\boldsymbol{I}] = \sum_{\cup_{p=1}^{q}\boldsymbol{I}_p=\boldsymbol{I}} (-1)^{q-1}(q-1)! \prod_{p=1}^{q} \text{mom}[\boldsymbol{I}_p] \quad (2.1)$$

where the summation extends over all partitions of set $\boldsymbol{I}$, i.e., the unordered collection of nonintersecting nonempty sets $\boldsymbol{I}_p$ such that $\cup_{p=1}^{q}\boldsymbol{I}_p = \boldsymbol{I}$. The moment-to-cumulant (2.1) can be inverted to obtain the cumulant-to-moment formula

$$\text{mom}[\boldsymbol{I}] = \sum_{\cup_{p=1}^{q}\boldsymbol{I}_p=\boldsymbol{I}} \prod_{p=1}^{q} \text{cum}[\boldsymbol{I}_p]. \quad (2.2)$$

A complete example that illustrates the use of (2.1) and (2.2) for the case $m = 4$ can be found in Appendix A of [2]. For that case, the number of partitions is 15, but it increases very quickly with the order and for $m = 8$, the number of partitions is 4140. For zero mean variables, the partitions that have a subset with one element can be discarded. Even in this case, the number of terms is still 715 for $m = 8$. The partitions required by (2.1) and (2.2) can be found grouped in a manageable form in [7] for orders up to $m = 8$.

A. Stationary Processes

The mth-order moment of a strictly stationary random process $x(t)$, which is denoted by $M_{m,x}(i_1, i_2, \cdots, i_{m-1})$, is defined as the joint mth-order moment of the random variables $\{x(t), x(t+i_1), \cdots, x(t+i_{m-1})\}$

$$\begin{aligned} &M_{m,x}(i_1, i_2, \cdots, i_{m-1}) \\ &\quad = \text{mom}[\{x(t), x(t+i_1), x(t+i_2), \cdots, x(t+i_{m-1})\}] \end{aligned} \quad (2.3)$$

The mth-order cumulant is similarly defined as

$$\begin{aligned} &C_{m,x}(i_1, i_2, \cdots, i_{m-1}) \\ &\quad = \text{cum}[\{x(t), x(t+i_1), x(t+i_2), \cdots, x(t+i_{m-1})\}]. \end{aligned} \quad (2.4)$$

Due to the stationarity of the process, the right side of (2.3) and (2.4) are independent of t, i.e., the mth-order moment and cumulant are only a function of the $m-1$ lags $i_1, i_2, \cdots, i_{m-1}$.

B. Linear Processes

Although in the next sections of this paper we present results that are valid for a more general class of stationary process, we pay special attention to linear processes and linear processes in stationary noise due to their importance in signal processing.

Linear processes are defined as the output $y(t)$ of a linear system whose input is a sequence of independent, identically distributed (i.i.d.) random variables.

$$y(t) = \sum_{n=-\infty}^{\infty} h(n)v(t-n). \quad (2.5)$$

We will assume the following:

A1) The impulse response $h(t)$ is exponentially stable, i.e., for some $\alpha > 0, A > 0$,

$$|h(t)| \le Ae^{-t|\alpha|}.$$

A2) The cumulants of orders up to $2m$ of the i.i.d. input $v(t)$ are finite

$$\begin{aligned} &C_{k,v}(i_1, i_2, \cdots, i_{m-1}) \\ &= \begin{cases} \gamma_k & \text{if } i_1 = i_2 = \cdots = i_{m-1} = 0 \\ 0 & \text{otherwise} \end{cases} \quad 1 \le k \le 2m. \end{aligned}$$

Then, the cumulants of the linear stationary process $y(t)$ are related to the impulse response $h(t)$ by the Bartlett–Brillinger–Rosenblatt formula

$$C_{m,y}(i_1, i_2, \cdots, i_{m-1}) = \gamma_m H_m(i_1, i_2, \cdots, i_{m-1}) \quad (2.6)$$

where

$$H_m(i_1, i_2, \cdots, i_{m-1}) = \sum_{n=-\infty}^{\infty} \prod_{k=0}^{m-1} h(n+i_k) \qquad (i_0 = 0). \quad (2.7)$$

It is also straightforward to check that the cumulants $C_{m,y}(i_1, i_2, \cdots, i_{m-1})$ are finite and summable, i.e., infinite summation over any or all of the indices is also finite.

The output may be observed in presence of stationary additive noise $w(t)$. Wee assume the following

A3) The noise $w(t)$ is independent of the input.

A4) The cumulants of the noise $C_{m,w}(i_1, i_2, \cdots, i_{m-1})$ are finite and summable.

Then, the cumulants of the resulting process

$$x(t) = y(t) + w(t)$$

are the sum of the cumulants of each term

$$\begin{aligned} &C_{m,x}(i_1, i_2, \cdots, i_{m-1}) \\ &\quad = \gamma_m H_m(i_1, i_2, \cdots, i_{m-1}) + C_{m,w}(i_1, i_2, \cdots, i_{m-1}), \end{aligned} \quad (2.8)$$

and $C_{m,x}(i_1, i_2, \cdots, i_{m-1})$ is also summable.

III. Sample Moments and Cumulants

Most of the methods for estimating the parameters of a given process $x(t)$ are based on sample moments. These can be defined, in the first-, second-, third-, and fourth-order cases as

$$m_1 = \frac{1}{N}\sum_{t=1}^{N} x(t) \quad (3.1a)$$

$$m_2(i) = \frac{1}{N}\sum_{t=1}^{N} x(t)x(t+i) \quad (3.1b)$$

$$m_3(i,j) = \frac{1}{N}\sum_{t=1}^{N} x(t)x(t+i)x(t+j) \quad (3.1c)$$

$$m_4(i,j,k) = \frac{1}{N}\sum_{t=1}^{N} x(t)x(t+i)x(t+j)x(t+k) \quad (3.1d)$$

These definitions give unbiased estimates of the moments. Other definitions are possible, depending on the type of data windowing employed [15]. Asymptotically, all the definitions are equivalent, i.e., they all converge with probability one to the true moments with the same asymptotic covariance.

Sample cumulants are then computed using (2.4) and the moment to cumulant (2.1) with sample statistics replacing true statistics. In practice, the sample mean m_1 is subtracted from $x(t)$, and then, the simpler but equivalent formulas that follow are used to compute the sample cumulants of the process:

$$x_o(t) = x(t) - m_1 \tag{3.2a}$$

$$c_2(i) = \frac{1}{N}\sum_{t=1}^{N} x_o(t)x_o(t+i) \tag{3.2b}$$

$$c_3(i,j) = \frac{1}{N}\sum_{t=1}^{N} x_o(t)x_o(t+i)x_o(t+j) \tag{3.2c}$$

$$\begin{aligned} c_4(i,j,k) = \frac{1}{N}\sum_{t=1}^{N} & x_o(t)x_o(t+i)x_o(t+j) \\ & \cdot x_o(t+k) - c_2(i)c_2(k-j) \\ & - c_2(j)c_2(k-i) - c_2(k)c_2(j-i) \end{aligned} \tag{3.2d}$$

The derivation of the covariances of sample cumulants begins with the computation of the covariances of sample moments. Then, the above relation between sample moments and sample cumulants is used to obtain the covariances of these statistics. We observe that even if the process is known to be zero mean, we consider that the sample mean is subtracted from the given samples $x(t)$ using (3.2a). This step is usually recommended in the literature [1]. We will show that this subtraction affects the covariances of the sample cumulants reducing the number of terms in their formulas.

Let $\boldsymbol{m}_N = (m_i(\cdots), mj(\cdots), \cdots)^t$ be a vector of sample moments, and let $\boldsymbol{M} = (M_{i,x}(\cdots), M_{j,x}(\cdots), \cdots)^t$ be the vector with the same-indexed true moments. Assuming $C_{m,x}(\cdot)$ is absolutely summable, the sample moments are known to have the following properties [3], [9]:

i) They converge almost surely to the moments of the process, i.e.

$$\lim_{N\to\infty} \boldsymbol{m}_N = \boldsymbol{M} \qquad \text{a.s.}$$

ii) Their covariance is $O(N^{-1})$, i.e.

$$\lim_{N\to\infty} NE[(\boldsymbol{m}_N - \boldsymbol{M})(\boldsymbol{m}_N - \boldsymbol{M})^t] = \boldsymbol{\Sigma}(\boldsymbol{M}) < \infty.$$

iii) Their third- and fourth-order central moments are $o(N^{-1})$, i.e.

$$\lim_{N\to\infty} NE[(\boldsymbol{m}_N - \boldsymbol{M})_i(\boldsymbol{m}_N - \boldsymbol{M})_j(\boldsymbol{m}_N - \boldsymbol{M})_k] = 0$$

$$\begin{aligned}\lim_{N\to\infty} NE[&(\boldsymbol{m}_N - \boldsymbol{M})_i(\boldsymbol{m}_N - \boldsymbol{M})_j \\ &\cdot (\boldsymbol{m}_N - \boldsymbol{M})_k(\boldsymbol{m}_N - \boldsymbol{M})_l] = 0.\end{aligned}$$

Let $\boldsymbol{g}(\boldsymbol{m}_N)$ be a vector-valued continuous function with continuous and bounded partial derivatives of first and second orders in some open neighborhood of $\boldsymbol{M}$; then, $\boldsymbol{s}_N = g(\boldsymbol{m}_N)$ also has the properties i)–iii). In particular, $\boldsymbol{s}_N$ converges almost surely to $\boldsymbol{S} = g(\boldsymbol{M})$, i.e.

$$\lim_{N\to\infty} \boldsymbol{s}_N = \boldsymbol{S} \qquad \text{a.s.}$$

and its asymptotic covariance is given by the following theorem:

Theorem 1 [3]: Under the above assumptions

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{M}) &= \lim_{N\to\infty} NE[(\boldsymbol{s}_N - \boldsymbol{S})(\boldsymbol{s}_N - \boldsymbol{S})^t] \\ &= \boldsymbol{G}(\boldsymbol{M})\boldsymbol{\Sigma}(\boldsymbol{M})\boldsymbol{G}^t(\boldsymbol{M})\end{aligned}$$

where $\boldsymbol{G}(\boldsymbol{M})$ is the Jacobian matrix of $\boldsymbol{g}(\boldsymbol{m})$, evaluated at $\boldsymbol{m} = \boldsymbol{M}$.

This theorem allows one to obtain the asymptotic performance of any parametric method based on higher order statistics if we are able to compute the asymptotic covariance $\boldsymbol{\Sigma}(\boldsymbol{M})$ of the sample moments and the Jacobian $\boldsymbol{G}(\boldsymbol{M})$ of the method with respect to these statistics.

Since $\boldsymbol{s}_N$ has the same properties as $\boldsymbol{m}_N$, Theorem 1 can be applied recursively to a function of $\boldsymbol{s}_N$. In fact, since most of the methods based on higher order statistics deal directly with cumulants instead of moments, it will be easier to work with the covariances of the sample cumulants and the Jacobian of the method with respect to these statistics. The computation of the asymptotic covariances of the sample moments and cumulants is discussed in the following section. Then, as an example of their application, we analyze the performance of cumulant-based methods in different applications.

IV. Asymptotic Covariances of the Sample Moments and Cumulants of Stationary Processes

A. Sample Moments

The asymptotic cross covariance of two sample moments $m_i(a_1, \cdots, a_{i-1})$ and $m_j(b_1, \cdots, b_{j-1})$ is given by the summation [3]

$$\begin{aligned}\lim_{N\to\infty} & N\mathrm{cov}\{m_i(a_1,\cdots,a_{i-1}), m_j(b_1,\cdots,b_{j-1})\} \\ &= \sum_{t=-\infty}^{\infty} \mathrm{cov}\{x(0)x(a_1)\cdots x(a_{i-1}), \\ &\quad x(t)x(t+b_1)\cdots x(t+b_{j-1})\} \\ &= \sum_{t=-\infty}^{\infty} (M_{i+j,x}(a_1,\cdots,a_{i-1},t,t+b_1,\cdots,t+b_{j-1}) \\ &\quad - M_{i,x}(a_1,\cdots,a_{i-1})M_{j,x}(b_1,\cdots,b_{j-1})).\end{aligned} \tag{4.1}$$

In the following, we will assume this summation is finite. For linear stationary processes, this is always true under assumptions A1) and A2) of Section II.

The evaluation of the above formula requires the computation of moments whose order is equal to the sum of the orders of each sample moment. Hence, if we are interested in the covariances of sample fourth-order moments, we have to be able to express the eighth-order moments of a process as a function of its parameters.

For linear systems, the computation of these moments as a function of the impulse response is performed through the

Bartlett–Brillinger–Rosenblatt and moment-to-cumulant formulas. Although the catalog of complementary set partitions provided by McCullagh in [7] may simplify this task, symbolic packages as *Mathematica* [8] seem to be the best tools to deal with the huge number of cumulants required to compute sixth-, seventh-, and eighth-order moments of linear processes.

The simplest example of the use of (4.1) corresponds to the covariance of the sample mean. Equation (4.1) gives, in this case

$$\lim_{N\to\infty} N\operatorname{cov}\{m_1, m_1\} = \sum_{t=-\infty}^{\infty} \operatorname{cov}\{x(0), x(t)\} = \sum_{t=-\infty}^{\infty} C_{2,x}(t). \tag{4.2}$$

Explicit expressions for the covariances of sample moments of order three or less were derived in [3]. Here, we complete that work presenting the expressions of the covariances of sample fourth-order moments and the cross covariances of these moments with others of lower order. In addition, we deal with expressions valid for any stationary process. This allows us to study nonlinear processes or the effect of additive noise.

Even if the process is symmetrically distributed, i.e., its odd-order cumulants are zero, the resulting formulas in the fourth-order case have hundreds of terms and will be omitted in this paper. Since most of the methods based on higher order statistics deal directly with cumulants instead of moments, the covariances of sample moments are interesting only as an intermediate step in the calculation of the covariances of sample cumulants.

B. Sample Cumulants

Let $\boldsymbol{s}_N$ be a vector of sample cumulants as given by (3.2). From the asymptotic covariances of sample moments (4.1) and applying Theorem 1, we obtain the asymptotic covariances of sample cumulants. In this section, we present the explicit expressions of the covariances of the second-, third-, and fourth-order sample cumulants of stationary processes derived using this procedure.

The covariances of the sample second-order cumulants are well known and simple to obtain. They are included here for completeness. The third-order cases were first studied in [3] for sample moments of linear processes. For zero-mean processes, it is known that the covariances of sample second-order moments are equal to those of sample cumulants, i.e., the subtraction of the sample mean does not affect the asymptotic covariance of the autocorrelation. This is not the case when third-order statistics are considered. The following equations show a reduced number of terms compared with those presented in [3] for third-order moments of zero-mean processes. In addition, the presented expressions are also valid for nonzero mean and nonlinear processes.

Let us denote

$$G_4(i,j) = \sum_{t=-\infty}^{\infty} C_{4,x}(i,t,t+j) \tag{4.3}$$

$$G_{22}(i) = \sum_{t=-\infty}^{\infty} C_{2,x}(t)C_{2,x}(t+i) \tag{4.4}$$

$$G_5(i,j,k) = \sum_{t=-\infty}^{\infty} C_{5,x}(i,t,t+j,t+k) \tag{4.5}$$

$$G_{23}(i,j) = \sum_{t=-\infty}^{\infty} C_{2,x}(t)C_{3,x}(t+i,t+j) \tag{4.6}$$

$$G_6(i,j,l,m) = \sum_{t=-\infty}^{\infty} C_{6,x}(i,j,t,t+l,t+m) \tag{4.7}$$

$$G_{6b}(i,l,m,n) = \sum_{t=-\infty}^{\infty} C_{6,x}(i,t,t+l,t+m,t+n) \tag{4.8}$$

$$G_{222}(i,j) = \sum_{t=-\infty}^{\infty} C_{2,x}(t)C_{2,x}(t+i)C_{2,x}(t+j) \tag{4.9}$$

$$G_{24a}(c,i,j) = \sum_{t=-\infty}^{\infty} C_{2,x}(t)C_{4,x}(c,t+i,t+j) \tag{4.10}$$

$$G_{24b}(i,j,k) = \sum_{t=-\infty}^{\infty} C_{2,x}(t)C_{4,x}(t+i,t+j,t+k) \tag{4.11}$$

$$G_{33}(c,i,j) = \sum_{t=-\infty}^{\infty} C_{3,x}(c,t)C_{3,x}(t+i,t+j) \tag{4.12}$$

$$G_8(i,j,k,l,m,n) = \sum_{t=-\infty}^{\infty} C_{8,x}(i,j,k,t,t+l,t+m,t+n) \tag{4.13}$$

$$G_{2222}(i,j,k) = \sum_{t=-\infty}^{\infty} C_{2,x}(t)C_{2,x}(t+i)C_{2,x}(t+j)C_{2,x}(t+k) \tag{4.14}$$

$$G_{224}(c,i,j,k) = \sum_{t=-\infty}^{\infty} C_{4,x}(c,t,t+i)C_{2,x}(t+j)C_{2,x}(t+k) \tag{4.15}$$

$$G_{26}(c,d,i,j,k) = \sum_{t=-\infty}^{\infty} C_{2,x}(t)C_{6,x}(c,d,t+i,t+j,t+k) \tag{4.16}$$

$$G_{44a}(c,d,i,j,k) = \sum_{t=-\infty}^{\infty} C_{4,x}(c,t,t+i)C_{4,x}(d,t+j,t+k) \tag{4.17}$$

$$G_{44b}(c,d,i,j,k) = \sum_{t=-\infty}^{\infty} C_{4,x}(c,d,t)C_{4,x}(t+i,t+j,t+k). \tag{4.18}$$

The asymptotic expression for the covariances of the sample second-order cumulants of stationary processes is

$$\lim_{N\to\infty} N\operatorname{cov}\{c_2(i), c_2(l)\} = G_4(i,l) + G_{22}(l-i) + G_{22}(l+i). \tag{4.19}$$

TABLE I
TWENTY FOUR TERMS CORRESPONDING TO $G_{2222}(-i+l,-j+m,-k+n)\{24\}$

G_{2222}(-i+l,-j+m,-k+n) +	G_{2222}(-i+l,-k+m,-j+n) +	G_{2222}(i+l,i-j+m,i-k+n) +	G_{2222}(i+l,i-k+m,i-j+n) +
G_{2222}(-j+l,-i+m,-k+n) +	G_{2222}(-j+l,-k+m,-i+n) +	G_{2222}(i-j+l,i+m,i-k+n) +	G_{2222}(i-j+l,i-k+m,i+n) +
G_{2222}(j+l,-i+j+m,j-k+n) +	G_{2222}(j+l,j-k+m,-i+j+n) +	G_{2222}(-i+j+l,j+m,j-k+n) +	G_{2222}(-i+j+l,j-k+m,j+n) +
G_{2222}(-k+l,-i+m,-j+n) +	G_{2222}(-k+l,-j+m,-i+n) +	G_{2222}(i-k+l,i+m,i-j+n) +	G_{2222}(i-k+l,i-j+m,i+n) +
G_{2222}(j-k+l,j+m,-i+j+n) +	G_{2222}(j-k+l,-i+j+m,j+n) +	G_{2222}(k+l,-i+k+m,-j+k+n) +	G_{2222}(k+l,-j+k+m,-i+k+n) +
G_{2222}(-i+k+l,k+m,-j+k+n) +	G_{2222}(-i+k+l,-j+k+m,k+n) +	G_{2222}(-j+k+l,k+m,-i+k+n) +	G_{2222}(-j+k+l,-i+k+m,k+n)

The asymptotic expression for the cross covariances of the sample third-order cumulants and the sample second-order cumulants of stationary processes is

$$\begin{aligned}\lim_{N\to\infty} & N\mathrm{cov}\{c_2(i), c_3(l,m)\}\\ &= G_5(i,l,m) + G_{23}(l-i, m-i)\\ &\quad + G_{23}(-l-i, m-l-i)\\ &\quad + G_{23}(-m-i, l-m-i)\\ &\quad + G_{23}(l+i, m+i)\\ &\quad + G_{23}(-l+i, m-l+i)\\ &\quad + G_{23}(-m+i, l-m+i)\end{aligned} \tag{4.20}$$

and the asymptotic expression for the covariances of the sample third-order cumulants of stationary processes is

$$\begin{aligned}\lim_{N\to\infty} & N\,\mathrm{cov}\{c_3(i,j), c_3(l,m)\}\\ &= G_6(i,j,l,m) + G_{222}(l-i, m-j)\\ &\quad + G_{222}(m-i, l-j) + G_{222}(l+i, m+i-j)\\ &\quad + G_{222}(l+j, m+j-i) + G_{222}(m+i, l+i-j)\\ &\quad + G_{222}(m+j, l+j-i)\\ &\quad + G_{33}(i, l-j, m-j)\\ &\quad + G_{33}(i, -j-l, m-j-l)\\ &\quad + G_{33}(i, -j-m, l-j-m)\\ &\quad + G_{33}(j, l-i, m-i)\\ &\quad + G_{33}(j, -i-l, m-i-l)\\ &\quad + G_{33}(j, -i-m, l-i-m)\\ &\quad + G_{33}(j-i, i-m, l+i-m)\\ &\quad + G_{33}(j-i, i-l, m+i-l)\\ &\quad + G_{33}(j-i, i+l, i+m)\\ &\quad + G_{24a}(j-i, l-i, m-i)\\ &\quad + G_{24a}(j-i, -l-i, m-l-i)\\ &\quad + G_{24a}(j-i, -m-i, l-m-i)\\ &\quad + G_{24a}(j, l+i, m+i)\\ &\quad + G_{24a}(j, -l+i, m-l+i)\\ &\quad + G_{24a}(j, -m+i, l-m+i)\\ &\quad + G_{24a}(i, l+j, m+j)\\ &\quad + G_{24a}(i, -l+j, m-l+j)\\ &\quad + G_{24a}(i, -m+j, l-m+j).\end{aligned} \tag{4.21}$$

The above expressions are valid for any stationary process. In the fourth-order case, we need to restrict ourselves to a less general class of processes to avoid formulas with thousands of terms or complicated index notations.

Since fourth-order sample cumulants are especially interesting in processes with null odd-order cumulants, in the following formulas, we will consider only these processes. The resulting expressions of the covariances are quite simpler than in the general case, but they still present up to 147 G terms.

The asymptotic expression for the cross covariances of the sample fourth-order cumulants and the sample second-order cumulants of stationary processes is

$$\begin{aligned}\lim_{N\to\infty} & N\mathrm{cov}\{c_4(i,j,k), c_2(l)\}\\ &= G_{6b}(l,i,j,k) + G_{24b}(-i-l, -i+j-l, -i+k-l)\\ &\quad + G_{24b}(i-l, j-l, k-l)\\ &\quad + G_{24b}(-j-l, i-j-l, -j+k-l)\\ &\quad + G_{24b}(-k-l, i-k-l, j-k-l)\\ &\quad + G_{24b}(-i+l, -i+j+l, -i+k+l)\\ &\quad + G_{24b}(i+l, j+l, k+l)\\ &\quad + G_{24b}(-j+l, i-j+l, -j+k+l)\\ &\quad + G_{24b}(-k+l, i-k+l, j-k+l)\end{aligned} \tag{4.22}$$

and the asymptotic expression for the covariances of the sample fourth-order cumulants of stationary processes is

$$\begin{aligned}\lim_{N\to\infty} & N\mathrm{cov}\{c_4(i,j,k), c_4(l,m,n)\}\\ &= G_8(i,j,k,l,m,n)\\ &\quad + G_{2222}(-i+l, -j+m, -k+n)\{24\}\\ &\quad + G_{224}(i, l, -j+m, -k+n)\{72\}\\ &\quad + G_{26}(i, j, k-l, k-l+m, k-l+n)\{16\}\\ &\quad + G_{44a}(i, -j+k, l, -j+m, -j+n)\{18\}\\ &\quad + G_{44b}(i, j, -k-l, -k-l+m, -k-l+n)\{16\}\end{aligned} \tag{4.23}$$

where the numbers in brackets indicate the number of similar G terms. The complete list of $G_{2222}, G_{224}, G_{26}, G_{44a}$, and G_{44b} terms can be found in the Tables I–V.

As we have mentioned before, these expressions of the covariances of sample cumulants are also valid for nonzero mean stationary processes. Furthermore, it is also interesting to observe they do not depend on the mean of the process.

C. Stationary Linear Processes

For linear processes or linear processes in noise, (2.6)–(2.8) are used to obtain the final expressions of the covariances as a function of the impulse response and the cumulants of the noise.

TABLE II
SEVENTY TWO TERMS CORRESPONDING TO $G_{224}(i,l,-j+m,-k+n)\{72\}$

$G_{224}(i,l,-j+m,-k+n)$ +	$G_{224}(i,l,-k+m,-j+n)$ +	$G_{224}(i,m,-j+l,-k+n)$ +	$G_{224}(i,m,-k+l,-j+n)$ +
$G_{224}(i,-l+m,-j-l,-k-l+n)$ +	$G_{224}(i,-l+m,-k-l,-j-l+n)$ +	$G_{224}(i,n,-j+l,-k+m)$ +	$G_{224}(i,n,-k+l,-j+m)$ +
$G_{224}(i,-l+n,-j-l,-k-l+m)$ +	$G_{224}(i,-l+n,-k-l,-j-l+m)$ +	$G_{224}(i,-m+n,-j-m,-k+l-m)$ +	$G_{224}(i,-m+n,-k-m,-j+l-m)$ +
$G_{224}(j,l,-i+m,-k+n)$ +	$G_{224}(j,l,-k+m,-i+n)$ +	$G_{224}(j,m,-i+l,-k+n)$ +	$G_{224}(j,m,-k+l,-i+n)$ +
$G_{224}(j,-l+m,-i-l,-k-l+n)$ +	$G_{224}(j,-l+m,-k-l,-i-l+n)$ +	$G_{224}(j,n,-i+l,-k+m)$ +	$G_{224}(j,n,-k+l,-i+m)$ +
$G_{224}(j,-l+n,-i-l,-k-l+m)$ +	$G_{224}(j,-l+n,-k-l,-i-l+m)$ +	$G_{224}(j,-m+n,-i-m,-k+l-m)$ +	$G_{224}(j,-m+n,-k-m,-i+l-m)$ +
$G_{224}(-i+j,l,i+m,i-k+n)$ +	$G_{224}(-i+j,l,i-k+m,i+n)$ +	$G_{224}(-i+j,m,i+l,i-k+n)$ +	$G_{224}(-i+j,m,i-k+l,i+n)$ +
$G_{224}(-i+j,-l+m,i-l,i-k-l+n)$ +	$G_{224}(-i+j,-l+m,i-k-l,i-l+n)$ +	$G_{224}(-i+j,n,i+l,i-k+m)$ +	$G_{224}(-i+j,n,i-k+l,i+m)$ +
$G_{224}(-i+j,-l+n,i-l,i-k-l+m)$ +	$G_{224}(-i+j,-l+n,i-k-l,i-l+m)$ +	$G_{224}(-i+j,-m+n,i-m,i-k+l-m)$ +	$G_{224}(-i+j,-m+n,i-k-m,i+l-m)$ +
$G_{224}(k,l,-i+m,-j+n)$ +	$G_{224}(k,l,-j+m,-i+n)$ +	$G_{224}(k,m,-i+l,-j+n)$ +	$G_{224}(k,m,-j+l,-i+n)$ +
$G_{224}(k,-l+m,-i-l,-j-l+n)$ +	$G_{224}(k,-l+m,-j-l,-i-l+n)$ +	$G_{224}(k,n,-i+l,-j+m)$ +	$G_{224}(k,n,-j+l,-i+m)$ +
$G_{224}(k,-l+n,-i-l,-j-l+m)$ +	$G_{224}(k,-l+n,-j-l,-i-l+m)$ +	$G_{224}(k,-m+n,-i-m,-j+l-m)$ +	$G_{224}(k,-m+n,-j-m,-i+l-m)$ +
$G_{224}(-i+k,l,i+m,i-j+n)$ +	$G_{224}(-i+k,l,i-j+m,i+n)$ +	$G_{224}(-i+k,m,i+l,i-j+n)$ +	$G_{224}(-i+k,m,i-j+l,i+n)$ +
$G_{224}(-i+k,-l+m,i-l,i-j-l+n)$ +	$G_{224}(-i+k,-l+m,i-j-l,i-l+n)$ +	$G_{224}(-i+k,n,i+l,i-j+m)$ +	$G_{224}(-i+k,n,i-j+l,i+m)$ +
$G_{224}(-i+k,-l+n,i-l,i-j-l+m)$ +	$G_{224}(-i+k,-l+n,i-j-l,i-l+m)$ +	$G_{224}(-i+k,-m+n,i-m,i-j+l-m)$ +	$G_{224}(-i+k,-m+n,i-j-m,i+l-m)$ +
$G_{224}(-j+k,l,j+m,-i+j+n)$ +	$G_{224}(-j+k,l,-i+j+m,j+n)$ +	$G_{224}(-j+k,m,j+l,-i+j+n)$ +	$G_{224}(-j+k,m,-i+j+l,j+n)$ +
$G_{224}(-j+k,-l+m,j-l,-i+j-l+n)$ +	$G_{224}(-j+k,-l+m,-i+j-l,j-l+n)$ +	$G_{224}(-j+k,n,j+l,-i+j+m)$ +	$G_{224}(-j+k,n,-i+j+l,j+m)$ +
$G_{224}(-j+k,-l+n,j-l,-i+j-l+m)$ +	$G_{224}(-j+k,-l+n,-i+j-l,j-l+m)$ +	$G_{224}(-j+k,-m+n,j-m,-i+j+l-m)$ +	$G_{224}(-j+k,-m+n,-i+j-m,j+l-m)$

TABLE III
SIXTEEN TERMS CORRESPONDING TO $G_{26}(i,j,k-l,k-l+m,k-l+n)\{16\}$

$G_{26}(i,j,k-l,k-l+m,k-l+n)$ +	$G_{26}(i,j,k+l,k+m,k+n)$ +	$G_{26}(i,j,k-m,k+l-m,k-m+n)$ +	$G_{26}(i,j,k-n,k+l-n,k+m-n)$ +
$G_{26}(i,k,j-l,j-l+m,j-l+n)$ +	$G_{26}(i,k,j+l,j+m,j+n)$ +	$G_{26}(i,k,j-m,j+l-m,j-m+n)$ +	$G_{26}(i,k,j-n,j+l-n,j+m-n)$ +
$G_{26}(j,k,i-l,i-l+m,i-l+n)$ +	$G_{26}(j,k,i+l,i+m,i+n)$ +	$G_{26}(j,k,i-m,i+l-m,i-m+n)$ +	$G_{26}(j,k,i-n,i+l-n,i+m-n)$ +
$G_{26}(-i+j,-i+k,-i-l,-i-l+m,-i-l+n)$ +	$G_{26}(-i+j,-i+k,-i+l,-i+m,-i+n)$ +	$G_{26}(-i+j,-i+k,-i-m,-i+l-m,-i-m+n)$ +	$G_{26}(-i+j,-i+k,-i-n,-i+l-n,-i+m-n)$

TABLE IV
EIGHTEEN TERMS CORRESPONDING TO $G_{44a}(i,-j+k,l,-j+m,-j+n)\{18\}$

$G_{44a}(i,-j+k,l,-j+m,-j+n)$ +	$G_{44a}(i,-j+k,m,-j+l,-j+n)$ +	$G_{44a}(i,-j+k,-l+m,-j-l,-j-l+n)$ +	$G_{44a}(i,-j+k,n,-j+l,-j+m)$ +
$G_{44a}(i,-j+k,-l+n,-j-l,-j-l+m)$ +	$G_{44a}(i,-j+k,-m+n,-j-m,-j+l-m)$ +	$G_{44a}(j,-i+k,l,-i+m,-i+n)$ +	$G_{44a}(j,-i+k,m,-i+l,-i+n)$ +
$G_{44a}(j,-i+k,-l+m,-i-l,-i-l+n)$ +	$G_{44a}(j,-i+k,n,-i+l,-i+m)$ +	$G_{44a}(j,-i+k,-l+n,-i-l,-i-l+m)$ +	$G_{44a}(j,-i+k,-m+n,-i-m,-i+l-m)$ +
$G_{44a}(-i+j,k,l,i+m,i+n)$ +	$G_{44a}(-i+j,k,m,i+l,i+n)$ +	$G_{44a}(-i+j,k,-l+m,i-l,i-l+n)$ +	$G_{44a}(-i+j,k,n,i+l,i+m)$ +
$G_{44a}(-i+j,k,-l+n,i-l,i-l+m)$ +	$G_{44a}(-i+j,k,-m+n,i-m,i+l-m)$		

TABLE V
SIXTEEN TERMS CORRESPONDING TO $G_{44b}(i,j,-k-l,-k-l+m,-k-l+n)\{16\}$

$G_{44b}(i,j,-k-l,-k-l+m,-k-l+n)$ +	$G_{44b}(i,j,-k+l,-k+m,-k+n)$ +	$G_{44b}(i,j,-k-m,-k+l-m,-k-m+n)$ +	$G_{44b}(i,j,-k-n,-k+l-n,-k+m-n)$ +
$G_{44b}(i,k,-j-l,-j-l+m,-j-l+n)$ +	$G_{44b}(i,k,-j+l,-j+m,-j+n)$ +	$G_{44b}(i,k,-j-m,-j+l-m,-j-m+n)$ +	$G_{44b}(i,k,-j-n,-j+l-n,-j+m-n)$ +
$G_{44b}(j,k,-i-l,-i-l+m,-i-l+n)$ +	$G_{44b}(j,k,-i+l,-i+m,-i+n)$ +	$G_{44b}(j,k,-i-m,-i+l-m,-i-m+n)$ +	$G_{44b}(j,k,-i-n,-i+l-n,-i+m-n)$ +
$G_{44b}(-i+j,-i+k,i-l,i-l+m,i-l+n)$ +	$G_{44b}(-i+j,-i+k,i+l,i+m,i+n)$ +	$G_{44b}(-i+j,-i+k,i-m,i+l-m,i-m+n)$ +	$G_{44b}(-i+j,-i+k,i-n,i+l-n,i+m-n)$

For example, for a linear process in white Gaussian noise with variance σ^2, we obtain

$$\begin{aligned} G_{22}(i) &= \sum_{t=-\infty}^{\infty} (\gamma_2 H_2(t) + \sigma^2\delta(t))(\gamma_2 H_2(t+i) \\ &\quad + \sigma^2\delta(t+i)) \\ &= \gamma_2^2 \sum_{t=-\infty}^{\infty} H_2(t)H_2(t+i) + 2\sigma^2\gamma_2 H_2(i) \\ &\quad + \sigma^4\delta(i). \end{aligned} \tag{4.24}$$

For the MA process without noise or in presence of noise with a finite number of nonzero cumulants, the number of nonzero terms in (2.7), (2.8), and (4.3)–(4.18) is finite, and both the H and G terms are computed directly as the summation indicated in those formulas.

For ARMA processes, the exact or symbolic computation of the H and G terms is quite more complicated. The procedure followed in [3] for H_1, H_2, and H_3 can be extended easily for H_4, although the computational cost increases exponentially and may be prohibitive for high-order systems. The exact computation of $G_{22}, G_{23}, G_{222}, G_{24}$, and G_{33} is described in [3] for linear systems without noise. Similar expressions can be derived for the $G_{2222}, G_{224}, G_{26}, G_{44a}$, and G_{44b} terms, but the complexity and computational cost for high-order systems limit their practical interest.

Under assumption A1), the products of H terms are also exponentially stable. Hence, in practice, approximate results may be obtained considering only a finite number of terms in the G summations. With this truncation, we can easily include the effect of noise with known cumulants.

Additionally, for linear process without noise or in additive Gaussian noise, the summations G_4, G_5, G_6, G_{6b}, and G_8

can be computed with the following simple expressions (see Appendix A):

$$G_4(i,l) = \gamma_4 H_2(i) H_2(l) \quad (4.25)$$

$$G_5(i,l,m) = \gamma_5 H_2(i) H_3(l,m) \quad (4.26)$$

$$G_6(i,j,l,m) = \gamma_6 H_3(i,j) H_3(l,m) \quad (4.27)$$

$$G_{6b}(i,l,m,n) = \gamma_6 H_2(i) H_4(l,m,n) \quad (4.28)$$

$$G_8(i,j,k,l,m,n) = \gamma_8 H_4(i,j,k) H_4(l,m,n). \quad (4.29)$$

V. Performance Analysis: Examples

To date, higher order statistics are widely used in signal processing applications. However, when addressing specific problems or applications, it is not easy to compare and choose a good solution. Although a great quantity of simulation results can be found in the literature, many questions of theoretical or practical interest still remain unanswered. Simulations are usually very limited in scope and do not allow one to predict the general behavior of the derived algorithms.

The expressions of the covariances provide us with an important tool to study the properties of cumulant-based methods and find an answer to most of the open questions. In this section, we present several applications of these expressions to cumlant-based system identification problems with the aim of obtaining the first answers to some of these questions. In the context of FIR system identification problems, we study the information carried by different statistics of different or the same order, the performance of different linear methods, and the effect of noise in the performance of the algorithms. Monte Carlo simulations are also included to corroborate the analytic results.

Theorem 1 and the discussion following it allows one to compute the covariance of the estimates given by any cumulant-based method from the covariances of the sample cumulants and the Jacobian matrix of the estimates with respect to the sample cumulants. Again, with the help of a symbolic algorithm, we obtained the performance of FIR system identification methods as a function of the coefficients and the statistics of the noise. Using these expressions, the variance in the estimation of the MA parameters is analyzed and plotted in different conditions and for different methods. The Jacobian matrices of the methods considered in the examples can be found in Appendix B.

In [3], Porat and Friedlander obtained the lower bound on the asymptotic covariances of all estimates based on a specific set of statistics, and in [3] and [13], they proposed a method with this optimum performance. The high complexity of this algorithm limits its practical interest, but the lower bound itself is of great importance as a reference. It is also a clear choice to compare the information carried by different sets, and we use it here for this purpose.

In the following examples, we consider that the statistics of the input are unknown to the estimation method. The finite impulse response or coefficients corresponding to a MA(q) processes are denoted as $b_n, 0 \leq n \leq q, (b_0 = 1)$.

Example 1: In this example, we compare the information carried by second- and third-order sample cumulants of an MA(1) process as a function of the coefficient b_1. The input to the linear system is a zero-mean i.i.d. exponentially distributed random sequence. The cumulants γ_m of this input sequence are

$$\gamma_1 = 0 \qquad \gamma_2 = 1 \qquad \gamma_3 = 2$$
$$\gamma_4 = 6 \qquad \gamma_5 = 24 \qquad \gamma_6 = 120.$$

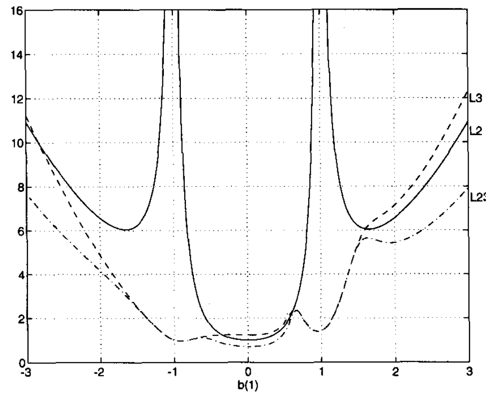

Fig. 1. Asymptotic standard deviation in the estimation of the coefficient b_1 of an MA(1) process as a function of its value. Lower bound using only autocorrelation (L2), using only third-order cumulants (L3), and using both (L23).

To compare the three sets of statistics, we consider the lower bound in the variance of any estimate of the coefficient b_1 based on each set of sample statistics [3]. The three curves of Fig. 1 show the asymptotic standard deviation in the estimation of b_1 for each of these sets. L2 corresponds to the lower bound using the sample autocorrelations ($c_2(0)$ and $c_2(0)$), L3 to the lower bound using the sample third-order cumulants ($c_3(0,0), c_3(0,1)$, and $c_3(1,1)$), and L23 to the lower bound using both sets of sample statistics. From this figure, it is clear that second-order statistics, apart from being phase blind, do not provide good estimates when the coefficient b_1 is close to 1 or to -1.

In general, correlation-based MA or ARMA system identification methods do not provide good estimates when the zeros are close to the unit circle. In these important cases, cumulant-based methods provide clearly better results for non-Gaussian processes.

Example 2: This example is an analytic study of the simulation results obtained by Mendel and Wang in [11] for the following nonminimum phase MA(2) model:

$$H(z) = 1 + b_1 z^{-1} + b_2 z^{-2} = 1 - 2.333 z^{-1} + 0.667 z^{-2}$$

with the same type of i.i.d. input of Example 1.

In [11], the authors applied the estimated cumulants $c_3(0,0)$, $c_3(0,1)$, $c_3(1,1)$, $c_3(0,2)$, $c_3(1,2)$, $c_3(2,2)$, $c_2(0)$, $c_2(1)$, and $c_2(2)$ to a structured network training algorithm to determine the MA parameters. They observed that accurate parameters were obtained by using higher order cumulants only, and correlation information did not seem to speed convergence. They also study the effect of different orderings of the training patterns. The results seem to indicate that the first cumulants $c_3(0,0), c_3(0,1)$, and $c_3(1,1)$ provide more information about the parameters than the other sample statistics.

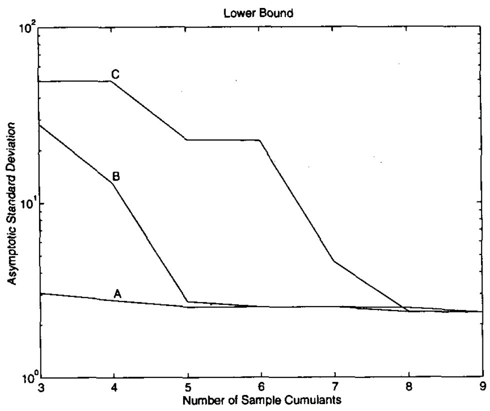

Fig. 2. Asymptotic standard deviation in the estimation of the coefficient b_2 of the MA(2) model of Example 2. Lower bound using different sets of second- and third-order cumulants.

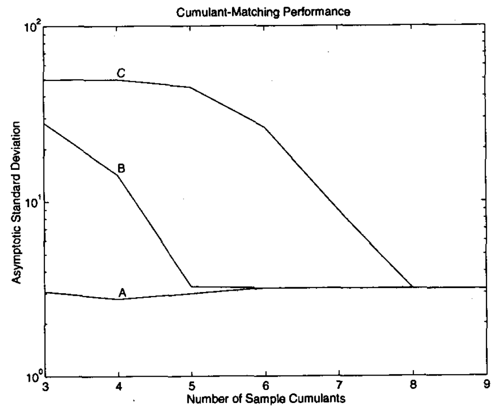

Fig. 3. Asymptotic standard deviation in the estimation of the coefficient b_2 of the MA(2) model of Example 2. Cumulant-matching method using different sets of second- and third-order cumulants.

Our analytic performance evaluations corroborate those simulation results. We compared the information carried by different sets of statistics, and we found that the sample cumulants $c_3(0,0), c_3(0,1)$, and $c_3(1,1)$ were clearly the most important, i.e., the lower bound in the variances of the estimates was not reduced significantly when more cumulants were considered. These results are illustrated graphically in Fig. 2. Let us define the following ordered lists of sample cumulants:

$$\begin{aligned} s_A &= \{c_3(0,0), c_3(0,1), c_3(1,1), c_3(0,2), c_3(1,2), \\ &\quad c_3(2,2), c_2(0), c_2(1), c_2(2)\} \\ s_B &= \{c_3(2,2), c_3(1,2), c_3(0,2), c_3(1,1), c_3(0,1), \\ &\quad c_3(0,0), c_2(2), c_2(1), c_2(0)\} \\ s_C &= \{c_2(0), c_2(1), c_2(2), c_3(2,2), c_3(1,2), \\ &\quad c_3(0,2), c_3(1,1), c_3(0,1), c_3(0,0)\}. \end{aligned}$$

Fig. 2 shows the lower bound in the standard deviation of the estimate of b_2 as a function of the number of statistics considered n and for each list of sample cumulants. For example, the plot corresponding to s_A gives at $n = 5$ the lower bound in the standard deviation of any method using the first five statistics of this list, namely, $c_3(0,0), c_3(0,1), c_3(1,1), c_3(0,2)$, and $c_3(1,2)$. For $n = 3$, s_A clearly provides the estimate with the lower standard deviation. Of course, as n increases, the three curves converge to the same point since the all the lists have the same nine sample cumulants.

Fig. 3 illustrates the performance of the cumulant matching method [12] (or MA Optimization-1 [2]), instead of the lower bound. This method performs a minimization of the sum of the squared differences between the sample cumulants and the cumulants of the proposed model, and it is closely related to the structured network approach of [11]. As in Fig. 2, the standard deviation of the estimate of b_2 is depicted in Fig. 3 as a function of the set of statistics used by the cumulant matching method. Results are similar to those presented in Fig. 2, but an important difference is observed: In some cases, the variance of the estimate increases when more sample cumulants are considered in the minimization.

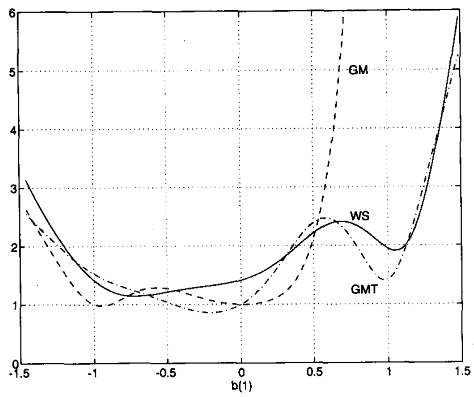

Fig. 4. Asymptotic standard deviation in the estimation of the coefficient b_1 of an MA(1) process as a function of its value. GM, GMT, and WS methods.

Example 3: We study here the identification of a MA(1) process using linear methods based on second- and/or third-order cumulants. As in Example 1, the input to the linear system is an i.i.d. exponentially distributed random sequence. Fig. 4 shows the performance (asymptotic standard deviation in the estimation of b_1) of the three linear methods. GM is the method proposed by Giannakis and Mendel in [4], and GMT is the *modification to reformulated GM algorithm* described in [5]. These two algorithms use second- and third-order cumulants, whereas the WS method developed in [6] uses only third-order cumulants.

The relative performance of the different three different methods depend on the value of b_1. The analysis of the GM method reveals the consistency problems of this method when b_1 approaches 1. The performance of the GMT and WS methods is similar in the range of values shown, although the WS method only uses third-order cumulants.

Example 4: Table I shows the almost perfect agreement between the predicted analytic performance and the simulation results for both the GMT and WS method and for three different values of b_1. The columns corresponding to the analytic performance evaluation show the asymptotic (normalized)

TABLE VI
NORMALIZED STANDARD DEVIATION OF THE ESTIMATE FOR THE GMT AND WS METHODS. ANALYTIC AND SIMULATION RESULTS OF 1000 MONTE CARLO RUNS WITH $N = 1000$

True Parameter	GMT Analytic	GMT Simulation	WS Analytic	WS Simulation
$b_1 = -0.80$	1.2969	1.3025	1.1604	1.1611
$b_1 = -1.25$	2.0264	2.0523	2.1834	2.1762
$b_1 = 1.25$	3.0512	3.0562	2.8146	2.8395

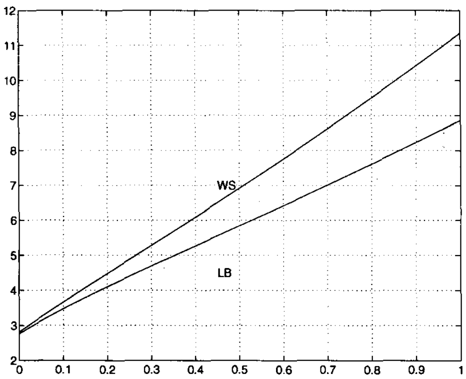

Fig. 5. Asymptotic standard deviation in the estimation of b_1 with the WS method as a function of the ratio between the power of the white Gaussian noise and the power of the output signal (MA(1) process with $b_1 = 1.25$). The lower line corresponds to the lower bound of any method using the sample third-order cumulants $c_3(0,0), c_3(0,1)$, and $c_3(1,1)$.

standard deviations as in previous examples. For the Monte Carlo simulations, the number of samples of each record was $N = 1000$, and in this case, the value of the normalized standard deviation shown in the tables was computed from the results as $\sqrt{N\mathrm{var}(\hat{b}_1)}$. Since we were interested in a accurate estimation of the variance, we performed 1000 Monte Carlo runs for each value of b_1.

Example 5: Fig. 5 shows the performance of WS method [6] as a function of the noise power for an MA(1) process with $b_1 = -1.25$. The lower bound of the asymptotic standard deviation is also included. The additive noise is white and Gaussian. In these two figures, the vertical axis represents the values of the asymptotic standard deviation, whereas the horizontal axis represents the value of the ratio between the power of the noise and the power of the output signal (MA process).

Although, in Fig. 5, the power of the signal is always greater then the power of the noise, in general, we have found that for signal-to-noise ratios below 0 dB, the performance of any cumulant-based method degrades very quickly.

Example 6: In this last example, a BPSK signal is used instead of the exponentially distributed input considered in previous examples. A BPSK signal is a sequence of statistically independent and equiprobable symbols with value 1 or -1. The cumulants γ_m of this sequence are

$$\begin{array}{llll} \gamma_1 = 0 & \gamma_2 = 1 & \gamma_3 = 0 & \gamma_4 = -2 \\ \gamma_5 = 0 & \gamma_6 = 16 & \gamma_7 = 0 & \gamma_8 = -272. \end{array}$$

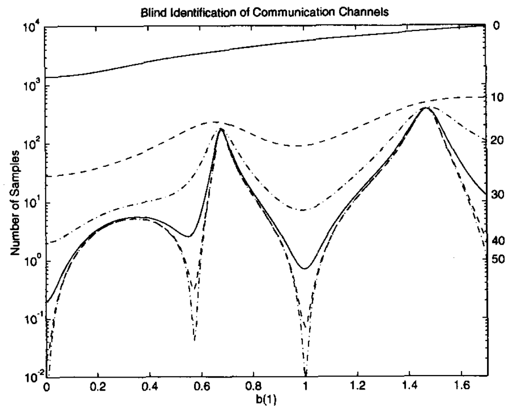

Fig. 6. Number of samples required to estimate the coefficient b_1 with with a standard deviation equal to 0.1 as a function of its value and the SNR. BPSK signal transmitted through an MA(1) channel with additive white Gaussian noise at the output. Lower bound using sample fourth-order cumulants $c_4(0,0,0), c_4(1,1,1)$ and $c_4(-1,-1,-1)$.

A BPSK signal is transmitted through an MA(1) channel, and we want to estimate b_1 from the sample fourth-order cumulants $c_4(0,0,0), c_4(1,1,1)$, and $c_4(-1,-1,-1)$ of the received signal. Fig. 6 depicts the lower bound in number of samples required to obtain a relative standard deviation of 0.1 in the estimation of b_1 as a function of the value of b_1 and for different signal-to-noise ratios. The relative standard deviation σ_r is defined respect to the total energy of the impulse response, i.e.

$$\sigma_r = \sqrt{\frac{\mathrm{var}(\hat{b}_1)}{1 + b_1^2}} = 0.1 \tag{5.1}$$

Hence, the lower bound in the number of samples N is computed from the lower bound in the asymptotic variance $LB(b_1)$ as

$$N = \frac{LB(b_1)}{\sigma_r^2(1 + b_1^2)} \tag{5.2}$$

The value of the signal-to-noise ratio is defined as

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\gamma_2(1 + b_1^2)}{\sigma_w^2}\right) \tag{5.3}$$

where σ_w^2 is the power of the white Gaussian noise added to the MA process.

From the results, we can observe a fast increase in the required number of samples when the signal-to-noise ratio is below 10 dB.

VI. CONCLUSION

We have presented the expressions of the covariances of the second-, third-, and fourth-order sample cumulants of stationary processes, placing emphasis on linear processes and linear processes in noise. These formulas are of great importance in the analysis of the increasing number of parametric methods based on cumulants and to clarify the interest of higher order statistics in signal processing applications.

Several experiments have been conducted to obtain a first answer to different questions related to the behavior of cumulant-based methods. From the results of examples, it is clear that higher order statistics in many cases, apart from providing phase information, carry useful information to reduce the variance of the estimates. For example, methods using only second-order statistics or methods based on a first estimation of a spectrally equivalent model using second-order statistics do not provide a good estimate of the model when a zero is close to the unit circle. On the contrary, zeros on the unit circle do not represent a problem to cumulant-based methods.

The lower bound on the asymptotic covariance of the estimate was used to study the information provided by different sets of sample cumulants. Cumulants close to the origin seem to be more important than cumulants with higher indexes. The effect of noise on the estimates has also been addressed in Examples 5 and 6. We have found that for signal-to-noise ratios below 0 dB, the performance degrades very quickly.

Apart from the analytic study of the algorithms, we also performed Monte Carlo simulations. The results showed a close agreement with the variance predicted by the asymptotic analysis.

Appendix A
Formulas of G_4, G_5, G_6, G_{6b} and G_8 for Linear Processes

For linear processes without noise or in additive Gaussian noise, the summation of cumulants required to compute G_4, G_5, G_6, G_{6b}, and G_8 can be avoided. In this Appendix, we include the complete derivation of the simplified expression of G_4. The expressions of G_5, G_6 \ and G_{6b} can be easily obtained with a similar approach.

From the definition of G_4, we have that for linear processes in Gaussian noise

$$\begin{aligned} G_4(i,j) &= \sum_{t=-\infty}^{\infty} \gamma_4 H_4(i,t,t+j) \\ &= \gamma_4 \sum_{t=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} h(n)h(n+i)h(n+t)h(n+t+j) \\ &= \gamma_4 \sum_{n=-\infty}^{\infty} h(n)h(n+i) \sum_{t=-\infty}^{\infty} h(n+t)h(n+t+j) \\ &= \gamma_4 \sum_{n=-\infty}^{\infty} h(n)h(n+i)H_2(j) \\ &= \gamma_4 H_2(i)H_2(j). \end{aligned} \quad \text{(B.1)}$$

Appendix B
Covariances of the Estimates

In this Appendix, we present the formulas of the covariances of the estimates given by the cumulant-based methods considered in the examples of Section V.

Let $\theta_N = \boldsymbol{g}(\boldsymbol{c})$ be the estimated parameters and $\theta = \boldsymbol{g}(\boldsymbol{C})$ be the true parameters of a process, where $\boldsymbol{c}$ are the estimated cumulants and $\boldsymbol{C} = \boldsymbol{f}(\theta)$ the true cumulants. From Theorem 1 and the discussion following it, we have that the asymptotic covariance of the estimated parameters is

$$\begin{aligned} (\boldsymbol{P}(\theta) &= \lim_{N\to\infty} NE[(\theta_N - \theta)(\theta_N - \theta)^t] \\ &= \boldsymbol{G}(\theta)\boldsymbol{\Sigma}(\theta)\boldsymbol{G}(\theta)^t \end{aligned} \quad \text{(C.1)}$$

where $\boldsymbol{G}(\theta)$ is the Jacobian matrix of $\boldsymbol{g}(\cdot)$ evaluated at $\boldsymbol{C} = \boldsymbol{f}(\theta)$, and $\boldsymbol{\Sigma}(\theta)$ is the asymptotic covariance of the sample cumulants, i.e.

$$\boldsymbol{\Sigma}(\theta) = \lim_{N\to\infty} NE[(\boldsymbol{c} - \boldsymbol{C})(\boldsymbol{c} - \boldsymbol{C})^t]. \quad \text{(C.2)}$$

The following formulas give the asymptotic covariances of the estimates for different methods. For the sake of completeness, we include the lower bound [3], which is also the asymptotic performance of the weighted cumulant matching method [13] (or MA Optimization-2 [2]).

Lower Bound: Let us denote

$$\frac{\partial f}{\partial \boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}). \quad \text{(C.3)}$$

The lower bound on the asymptotic covariances of all estimates $\boldsymbol{g}(\boldsymbol{c})$ is given by

$$\boldsymbol{B}(\theta) = [\boldsymbol{F}^t(\theta)\boldsymbol{\Sigma}(\theta)\boldsymbol{F}(\theta)]^{-1}. \quad \text{(C.4)}$$

Cumulant Matching Method: The asymptotic covariance of estimate θ_N obtained via the global minimization of

$$V(\boldsymbol{x}) = \tfrac{1}{2}(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{c})^t(\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{c}) \quad \text{(C.5)}$$

which is given by

$$\begin{aligned} \boldsymbol{P}^{(cm)}(\theta) &= [\boldsymbol{F}^t(\theta)\boldsymbol{F}(\theta)]^{-1}\boldsymbol{F}^t(\theta) \\ &\quad \cdot \boldsymbol{\Sigma}(\theta)\boldsymbol{F}(\theta)[\boldsymbol{F}(\theta)\boldsymbol{F}^t(\theta)]^{-1}. \end{aligned} \quad \text{(C.6)}$$

Proof: It is straightforward to show, by standard differential analysis, that the Jacobian matrix of the above estimate is given by

$$\boldsymbol{G}(\theta) = -\left[\frac{\partial V}{\partial x^2}\right]^{-1}\left[\frac{\partial V}{\partial x \partial c}\right]\Bigg|_{\substack{\boldsymbol{x}=\theta \\ \boldsymbol{c}=\boldsymbol{C}}} \quad \text{(C.7)}$$

where the terms of the right-hand side can be easily computed from (C.4)

$$\frac{\partial V}{\partial \boldsymbol{x}^2}\Bigg|_{\substack{\boldsymbol{x}=\theta \\ \boldsymbol{c}=\boldsymbol{C}}} = \boldsymbol{F}^t(\theta)\boldsymbol{F}(\theta) \quad \text{(C.8)}$$

$$\frac{\partial V}{\partial \boldsymbol{x}\partial \boldsymbol{c}}\Bigg|_{\substack{\boldsymbol{x}=\theta \\ \boldsymbol{c}=\boldsymbol{C}}} = -\boldsymbol{F}^t(\theta). \quad \text{(C.9)}$$

Substituting (C.8) and (C.9) into (C.7) and then in (C.1), we obtain (C.6).

Least Squares Estimates: The GM [4] and GMT [5] system identification methods are based on a linear relation

$$\boldsymbol{A}(\boldsymbol{C})\theta = \boldsymbol{b}(\boldsymbol{C}). \quad \text{(C.10)}$$

When both the matrix $\boldsymbol{A}$ and the vector $\boldsymbol{b}$ are computed from the estimated cumulants $\boldsymbol{c}$, we have an overdetermined linear system. Although there are other possibilities, the least squares solution is usually chosen as the estimate.

The asymptotic covariance of the least squares solution θ_N

$$\theta_N = \boldsymbol{g}^{(ls)}(\boldsymbol{c}) = [\boldsymbol{A}(\boldsymbol{c})^t\boldsymbol{A}(\boldsymbol{c})]^{-1}\boldsymbol{A}(\boldsymbol{c})^t\boldsymbol{b}(\boldsymbol{c}) \quad \text{(C.11)}$$

is given by [3]

$$\begin{aligned} \boldsymbol{P}^{(ls)}(\theta) &= [\boldsymbol{A}(\boldsymbol{C})^t\boldsymbol{A}(\boldsymbol{C})]^{-1}\boldsymbol{A}(\boldsymbol{C})^t\boldsymbol{D}(\boldsymbol{C}) \\ &\quad \cdot \boldsymbol{\Sigma}\boldsymbol{D}(\boldsymbol{C})^t\boldsymbol{A}(\boldsymbol{C})[\boldsymbol{A}(\boldsymbol{C})^t\boldsymbol{A}(\boldsymbol{C})]^{-1} \end{aligned} \quad \text{(C.12)}$$

where $\boldsymbol{D}(\boldsymbol{C})$ is a matrix whose ith column is

$$\boldsymbol{D}(\boldsymbol{C})_i = \left(\frac{\partial \boldsymbol{b}(\boldsymbol{c})}{\partial \boldsymbol{c}_i} - \frac{\partial \boldsymbol{A}(\boldsymbol{c})}{\partial \boldsymbol{c}_i} \theta \right) \Bigg|_{\boldsymbol{c}=\boldsymbol{C}}. \tag{C.13}$$

WS Method: The WS method [6] requires a separate analysis since a minimum norm solution is computed instead of a least squares solution.

Let $\boldsymbol{I}_1 = (0, \cdots, 0, 1)^t$. The asymptotic covariance of the WS estimate θ_N

$$\theta_N = \boldsymbol{g}^{(ws)}(\boldsymbol{c}) = \boldsymbol{S}_d \boldsymbol{S}_u^t [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \boldsymbol{I}_1 \tag{C.14}$$

is given by

$$\boldsymbol{P}^{(ws)}(\theta) = \boldsymbol{G} \boldsymbol{\Sigma} \boldsymbol{G}^t \tag{C.15}$$

where

$$\boldsymbol{G} = \boldsymbol{D}_1 + \boldsymbol{S}_d[(\boldsymbol{I} - \boldsymbol{S}_u^{\#} \boldsymbol{S}_u) \boldsymbol{D}_2 - \boldsymbol{S}_u^{\#} \boldsymbol{D}_3] \tag{C.16}$$

and the ith column of the matrices $\boldsymbol{D}_1$, $\boldsymbol{D}_2$ and $\boldsymbol{D}_3$ is given by

$$\boldsymbol{D}_{1,i} = \frac{\partial \boldsymbol{S}_d}{\partial \boldsymbol{c}_i} \boldsymbol{w} \tag{C.17}$$

$$\boldsymbol{D}_{2,i} = \frac{\partial \boldsymbol{S}_u^t}{\partial \boldsymbol{c}_i} [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \boldsymbol{I}_1 \tag{C.18}$$

$$\boldsymbol{D}_{3,i} = \frac{\partial \boldsymbol{S}_u}{\partial \boldsymbol{c}_i} \boldsymbol{w}. \tag{C.19}$$

All arguments on the right-hand side of (C.15) are understood to be evaluated at the point $\boldsymbol{C} = \boldsymbol{f}(\theta)$.

Proof: Using the following notation

$$\boldsymbol{S}_u^{\#} = \boldsymbol{S}_u^t [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \tag{C.20}$$

$$\boldsymbol{w} = \boldsymbol{S}_u^{\#} \boldsymbol{I}_1 \tag{C.21}$$

$$\boldsymbol{g}^{(ws)}(\boldsymbol{c}) = \boldsymbol{S}_d \boldsymbol{w} \tag{C.22}$$

we have that

$$\frac{\partial \boldsymbol{g}^{(ws)}(\boldsymbol{c})}{\partial \boldsymbol{c}_i} = \frac{\partial \boldsymbol{S}_d}{\partial c_i} \boldsymbol{w} + \boldsymbol{S}_d \frac{\partial \boldsymbol{w}}{\partial \boldsymbol{c}_i} = \boldsymbol{D}_{1,i} + \boldsymbol{S}_d \frac{\partial \boldsymbol{w}}{\partial \boldsymbol{c}_i} \tag{C.23}$$

$$\begin{aligned} \frac{\partial \boldsymbol{w}}{\partial \boldsymbol{c}_i} &= \frac{\partial \boldsymbol{S}_u^{\#}}{\partial \boldsymbol{c}_i} \boldsymbol{I}_1 \\ &= \frac{\partial \boldsymbol{S}_u^t}{\partial \boldsymbol{c}_i} [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \boldsymbol{I}_1 - \boldsymbol{S}_u^t [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \\ &\quad \cdot \frac{\partial \boldsymbol{S}_u}{\partial \boldsymbol{c}_i} \boldsymbol{S}_u^t [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} - \boldsymbol{I}_1 - \boldsymbol{S}_u^t [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \\ &\quad \cdot \boldsymbol{S}_u \frac{\partial \boldsymbol{S}_u^t}{\partial \boldsymbol{c}_i} [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \boldsymbol{I}_1 \\ &= \frac{\partial \boldsymbol{S}_u^t}{\partial \boldsymbol{c}_i} [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \boldsymbol{I}_1 - \boldsymbol{S}_u^{\#} \frac{\partial \boldsymbol{S}_u}{\partial \boldsymbol{c}_i} \\ &\quad \cdot \boldsymbol{w} - \boldsymbol{S}_u^{\#} \boldsymbol{S}_u \frac{\partial \boldsymbol{S}_u^t}{\partial \boldsymbol{c}_i} [\boldsymbol{S}_u \boldsymbol{S}_u^t]^{-1} \boldsymbol{I}_1 \\ &= \boldsymbol{D}_{2,i} - \boldsymbol{S}_u^{\#} \boldsymbol{D}_{3,i} - \boldsymbol{S}_u^{\#} \boldsymbol{S}_u \boldsymbol{D}_{2,i}. \end{aligned} \tag{C.24}$$

The final expression of the Jacobian matrix (C.16) is obtained by substituting (C.24) into (C.23).

SOFTWARE

The software used to obtain the covariances of sample cumulants and the presented results is available via anonymous FTP (dttix0.upc.es) in the directory pub/cov.

ACKNOWLEDGMENT

This work was performed at the Signal and Image Processing Institute, University of Southern California. The author would like to thank the suggestions of Dr. Mendel and the rest of researchers of the Institute.

REFERENCES

[1] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, pp. 869–891, July 1987.

[2] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.

[3] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *Int. J. Adaptive Contr. Signal Processing*, vol. 3, pp. 191–229, 1989.

[4] G. B. Giannakis and J. M. Mendel, "Identification of non-minimum phase systems using higher-order statistics," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 360–377, 1989.

[5] J. K. Tugnait, "New results on FIR system identification using higher order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, no. 10, pp. 2216–2221, Oct. 1991.

[6] J. A. R. Fonollosa, J. Vidal, and A. Moreno, "FIR system identification using a linear combination of cumulants," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Mar. 1992, pp. 473–476, vol. V.

[7] P. McCullagh, *Tensor Methods in Statistics, Monographs on Statistics and Applied Probability*. New York: Chapman and Hall, 1987.

[8] S. Wolfram, *Mathematica: A System for Doing Mathematics by Computer,*. Reading, MA: Addison-Wesley, 1991.

[9] M. Rosenblatt and J. W. Van Ness, "Estimation of the bispectrum," *Ann. Math. Statist.*, pp. 1120–1136, 1965.

[10] D. R. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of kth order spectra," in *Spectral Analysis of Time Series* (B. Harris, Ed.). New York: Wiley, 1967, pp. 153–188.

[11] J. M. Mendel and L. Wang, "Identification of moving average systems using higher-order statistics and learning," in *Neural Networks: A Dynamical Approach to Machine Intelligence* (B. Kosko, Ed.) Englewood Cliffs, NJ: Prentice Hall, 1990; see also,"Cumulant-based parameter estimation using structured networks," *IEEE Trans. Neural Networks*, vol. 2, pp. 73–83, Jan. 1991.

[12] K.-S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes," *Annals Statist.*, vol. 10, pp. 1195–1208, 1982.

[13] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from higher-order moments," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 485–495, 1989.

[14] D. R. Brillinger, "An introduction to polyspectra," *Ann. Math. Statist.*, vol. 36, pp. 1351–1374, 1965.

[15] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 1991, 2nd ed.

José A. R. Fonollosa was born in Barcelona, Spain, on April 19, 1963. He received the M.S. and Ph.D. degrees in electrical engineering from the Polytechnic University of Catalonia, UPC, Spain, in 1986 and 1989, respectively.

Since 1986, he is with the Department of Signal Theory and Communications of the Polytechnic University of Catalonia, where he is now an Associate Professor. From August 1991 to July 1992, he was visiting the Signal and Image Processing Institute, University of Southern California. His current research interests are in the area of statistical signal processing and its application in communication systems and speech processing algorithms.

Dr. Fonollosa received the 1992 Marconi Young Scientist Award. He is member of the European Speech Communication Association and the Acoustical Society of America. Currently, he is serving as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.