ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Wideband Spectrum Sensing by Model Order Selection

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Mariani, A., Giorgetti, A., Chiani, M. (2015). Wideband Spectrum Sensing by Model Order Selection. IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, 14(12), 6710-6721 [10.1109/TWC.2015.2458320].

Availability:
This version is available at: https://hdl.handle.net/11585/546465 since: 2021-10-15

Published:
DOI: http://doi.org/10.1109/TWC.2015.2458320

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

12 October 2024

This is the final peer-reviewed accepted manuscript of:

A. Mariani, A. Giorgetti and M. Chiani, "Wideband Spectrum Sensing by Model Order Selection," in *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6710-6721, Dec. 2015.

The final published version is available online at:

https://doi.org/10.1109/TWC.2015.2458320

Rights / License:

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

Wideband Spectrum Sensing by Model Order Selection

Andrea Mariani, *Member, IEEE*, Andrea Giorgetti, *Senior Member, IEEE*, and Marco Chiani, *Fellow, IEEE*

***Abstract*—Spectrum sensing is an essential functionality in cognitive radio (CR) systems allowing to discover spectrum opportunities and enabling primary user (PU) protection. Wideband spectrum sensing (WS) improves the awareness of the surrounding radio environment by jointly monitoring multiple frequency bands. In this paper we propose a WS approach based on the observation of a frequency domain representation of the received signal and the adoption of model order selection (MOS) to identify the occupied frequency components. We provide a general formulation of the problem valid for any kind of spectral representation and then focus on the case in which discrete Fourier transform (DFT) is used. This choice is motivated by the fact that DFT blocks are available in many wireless systems, such as OFDM receivers and recently proposed software radio architectures. We provide analytical expressions for the maximum probability of correct selection of the occupied sub-bands valid for MOS approaches encompassed within the generalized information criterion (GIC). We then propose a method for designing the selection algorithm to balance overestimation and underestimation. Numerical results show that the MOS scheme derived for DFT can be successfully applied also when more accurate frequency representations, such as the multitaper (MT) spectrum estimation, are adopted.**

***Index Terms*—Cognitive radio, generalized information criterion, information theoretic criteria, misspecified statistical model, model order selection, wideband spectrum sensing.**

I. Introduction

THE proliferation of wireless services is expected to increase the demand for radio spectrum resources in the near future [1]. To satisfy this request cognitive radio (CR) technologies have been proposed for a more flexible use of frequency bands allowing secondary users (SUs) to exploit portions of spectrum that are temporarily or locally unused [2]. The first functionality of a CR system is to observe the environment to find communication opportunities - often called *white spaces* or *spectrum holes* - through spectrum sensing. Most of the spectrum sensing techniques proposed in literature have been devised to assess the occupancy of a single frequency band [2], [3]. For example, the most common sensing technique is energy detection, in which the received energy is compared to a decision threshold properly set according to different criteria [4]. A better knowledge of the radio environment can be reached by wideband spectrum sensing (WS), that consists in a joint observation of multiple frequency sub-bands and joint decision on the occupancy in each sub-band. Wideband sensing can be also useful for RF monitoring to ensure coexistence in heterogeneous scenarios such as ISM bands and professional mobile radio (PMR) evolution [5].

Wideband sensing is typically based on the adoption of a frequency domain representation of the received signal and the computation of some metrics to evaluate the occupancy state of each sub-band [6]. In CR contexts non parametric spectrum estimation techniques are the most suitable because they do not require any assumption on the received signal. This class of techniques comprehends the classical estimators, based on the periodogram, as well as more advanced approaches, such as the multitaper (MT) method [7]–[9].

The application of WS is primarily related to hardware front-end requirements, such as the linearity of analog components and analog to digital converters (ADCs) characteristics [10]–[12]. To get around such constraints some approaches are based on sequential sensing on multiple bands, frequency sweeping or filter-banks [13]–[16]. To reduce hardware complexity reception schemes that use sampling frequencies lower than the Nyquist rate have been proposed [17]. In this context, two relevant classes of WS approaches are based on compressive sensing [18], and multichannel sub-Nyquist sampling [19], [20]. These strategies are effective in situations in which the primary user (PU) signal occupancy is sparse in the frequency domain [18], [21]. In [22], [23] WS has been formalized as a generalized likelihood ratio (GLR) detector assuming the presence of a given amount of unoccupied spectrum. Standard information theoretic criteria (ITC) have been adopted in [24], [25] where an energy detector (ED) in each sub-band is used.

In this paper we propose a WS approach based on the observation of a frequency domain representation of the received signal and the adoption of model order selection (MOS) to identify the occupied frequency components. We provide, in particular, the following contributions.

- We formulate WS as a MOS problem in which the order of the model is the number of occupied frequency components. We provide a general formulation valid for any kind of spectral representation and then focus on the case in which discrete Fourier transform (DFT) is used. This choice is motivated by the simplicity of the approach and the fact that often DFT blocks are already implemented in common systems such as orthogonal frequency-division

This work was supported in part by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement CONCERTO n. 288502, and in part by the Italian Ministry of Education, Universities and Research (MIUR) under Research Projects of Significant National Interest PRIN 2011 GRETA. This paper was presented, in part, at the Int. Conf. on Cognitive Radio and Advanced Spectrum Management (COGART 2011), Barcelona, Spain, Oct. 2011, and at the IEEE Int. Conf. on Commun. (ICC 2014), Sydney, Australia, Jun. 2014.

The authors are with the CIRI-ICT/DEI, University of Bologna, Via Venezia 52, 47521 Cesena, ITALY (e-mail: {a.mariani, andrea.giorgetti, marco.chiani}@unibo.it).

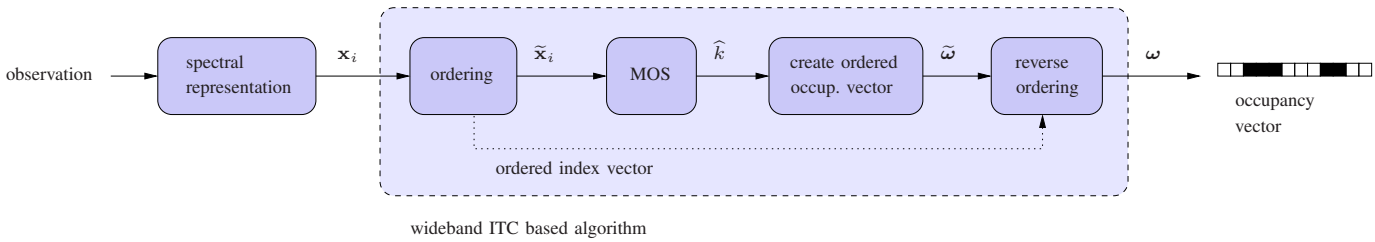


Fig. 1. Block diagram of the WS scheme proposed.

multiplexing (OFDM) receivers and recently proposed software defined radio (SDR) architectures [6], [26]–[28].

- We propose a design strategy for the generalized information criterion (GIC) which aims at minimizing underestimation while keeping overestimation probability below a target level.
- We provide tight analytical approximations of the maximum probability of overestimation, which are used for implementing the proposed design strategy.
- We prove that the MOS scheme derived for the DFT exhibits excellent performance also when weighted overlapped segment averaging (WOSA) and MT spectral estimates are adopted. This ensures a wide applicability of the proposed method.
- The WS scheme provides an estimate of the noise power floor which can be useful for flexible radio operations [6].

Most of the sensing algorithms proposed in literature are based on the comparison of a detection metric to a threshold, which setting is a difficult task in practice due to the dependence on unknown parameters. In particular, considering energy based techniques, including frequency domain analysis, the threshold setting depends on the noise power, which must be properly estimated [4]. The WS technique proposed, instead, is blind since it does not require neither noise power knowledge nor any a-priori information about the number and the characteristics of the signals present. With respect to the previously proposed WS schemes based on ITC [24], [25], we propose a general framework that can be adopted with any frequency domain representation and, considering DFT, we adopt GIC and propose a penalty design strategy.

The performance of the proposed scheme are compared with existing WS techniques based on the ED analyzing the probability of correct detection of the number of occupied frequency sub-bands and the detection probability on each sub-band. Numerical results reveal that the proposed algorithm remarkably outperforms ED based techniques.

The paper is organized as follows. In Section II we formulate WS as a MOS problem, providing the general scheme as well as DFT-based case. ITC are reviewed in Section III. In Section IV we propose a design strategy for the penalty associated with the GIC and in Section V some numerical results are presented. Throughout the paper, boldface letters denote matrices and vectors, $X \sim \chi^2_m$ is a central chi squared distributed random variable (r.v.) with m degrees of freedom (d.o.f.) and $X \sim \mathcal{G}(\kappa, \theta)$ is a gamma distributed r.v. with shape parameter κ and scale parameter θ. We denote the cumulative distribution function (CDF) of the r.v. X with $F_X(x)$. Moreover, $|\mathbf{A}|$ and $\mathrm{tr}\{\mathbf{A}\}$ stand, respectively, for the determinant and the trace of the matrix $\mathbf{A}$; $\mathrm{diag}\{\mathbf{A}\}$ stands for the vector containing the diagonal elements of $\mathbf{A}$; $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ stand, respectively, for simple and Hermitian transposition.

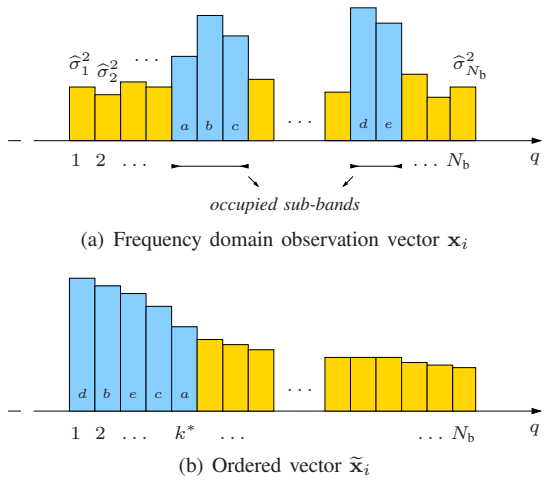
II. Wideband Sensing by Model Order Selection

The proposed WS technique is based on the observation of N independent frequency domain vectors $\mathbf{x}_i = (x_{1,i}, x_{2,i}, \dots x_{N_\mathrm{b},i})^{\mathrm{T}}$, each with N_b frequency components, where $i \in \{1, 2, \dots, N\}$ denotes the observation at the ith time instant. We assume that frequency occupancy does not change during the N observations. The vector $\mathbf{x}_i$ can be any kind of frequency domain representation, such as a power spectral density (PSD) estimate, the output of a filter bank, a compressive sampling reconstruction of the spectrum or, simply, the output of a N_b-points DFT. We generally refer to the elements of $\mathbf{x}_i$ as frequency bins. If PU signals are present in the observed frequency band we consider that they occupy k^* bins, while the remaining $N_\mathrm{b} - k^*$ contain only noise. The objective of WS is to identify the k^* occupied bins. To accomplish this goal we formulate WS as a MOS problem in which k^* is the order of the model. The proposed scheme estimates k^* and also identifies the occupied bins, following the approach described in the following.

Let us collect the N column vectors $\mathbf{x}_i$ in the observation matrix

$$\mathbf{X} = (\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_N) \tag{1}$$

and let us sort all vectors $\mathbf{x}_i$ according to the estimated power of each bin, $\widehat{\sigma}^2_q = (1/N) \sum_{i=1}^{N} |x_{q,i}|^2$, so that the elements of the vector $\left(\widehat{\sigma}^2_1, \widehat{\sigma}^2_2, \dots, \widehat{\sigma}^2_{N_\mathrm{b}}\right)$ are in descending order. We denote with $\widetilde{\mathbf{X}}$ the ordered matrix and with $\widetilde{\mathbf{x}}_i$ the corresponding columns. After sorting, MOS provides an estimate of k^*, $\widehat{k}$, which allows to construct the vector $\widetilde{\boldsymbol{\omega}}$ composed by the concatenation of $\widehat{k}$ ones and $N_\mathrm{b} - \widehat{k}$ zeros. Lastly, by means of reverse ordering, we obtain the occupancy vector $\boldsymbol{\omega}$. The qth frequency bin is declared occupied if the corresponding element in $\boldsymbol{\omega}$ equals 1. This WS method is represented by the block diagram in Fig. 1, with the ordering phase depicted in detail in Fig. 2. In some CR scenarios, the SUs may need to detect the presence of signals with fixed channelization. In these situations, if each channel corresponds



(a) Frequency domain observation vector $\mathbf{x}_i$

(b) Ordered vector $\widetilde{\mathbf{x}}_i$

Fig. 2. Ordering phase of the MOS based WS. Ordering is based on the average power of the elements of $\mathbf{x}_i$ (a), providing the vector $\widetilde{\mathbf{x}}_i$ (b).

to a group of consecutive bins, the WS strategy proposed can be modified by grouping sub-bands.

The WS strategy described above can be applied using any frequency representation of data and any MOS approach. For solving the selection problem in this paper we adopt ITC, which are described in the next section. Considering the frequency representation to be adopted, we focus on the case in which DFT of the received signal is used. In Section (V-C2) we also show some numerical results using WOSA and MT spectral estimates.

A. DFT-based wideband sensing

In this subsection we describe the signal model corresponding to the DFT case and derive the expression of the likelihood of the received observation matrix $\widetilde{\mathbf{X}}$, which will be used for detecting the occupied bins.

The output of the DFT at the ith observation is denoted as the N_b length vector

$$\mathbf{x}_i = \mathbf{s}_i + \mathbf{n}_i \tag{2}$$

where $\mathbf{n}_i$ represents the additive white Gaussian noise (AWGN) and $\mathbf{s}_i$ is the aggregation of the PUs signals including the channel effects. The signal-to-noise ratio (SNR) is defined as $\mathsf{SNR} = \mathbb{E}\left\{\mathbf{s}_i{}^\mathrm{H}\mathbf{s}_i\right\} / \left(N_\mathrm{b}\,\sigma^2\right)$, where σ^2 is the unknown noise power at each frequency bin. For many communication signals, such as widely adopted OFDM, the received samples in time domain can be modeled as zero mean complex Gaussian r.v.s [29]. Thus, due to the properties of the DFT operation, $\mathbf{x}_i$ and $\widetilde{\mathbf{x}}_i$ are vectors of zero mean complex Gaussian r.v.s with covariance matrices $\mathbf{\Sigma}_\mathbf{x} = \mathbb{E}\left\{\mathbf{x}_i\mathbf{x}_i{}^\mathrm{H}\right\}$ and $\mathbf{\Sigma}_{\widetilde{\mathbf{x}}} = \mathbb{E}\left\{\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i^\mathrm{H}\right\}$, respectively. Considering ordering performed in the presence of perfect bin power knowledge, if the number of frequency bins containing signals is k, $\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}$ is a block diagonal matrix with two blocks, where the first one, $\mathbf{\Sigma}_{(k)}$, is given by the leading principal minor of order k of $\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}$, and the second is $\sigma^2\mathbf{I}_{N_\mathrm{b}-k}$, where $\mathbf{I}_p$ is a $p\times p$ identity matrix. We decompose the vector $\widetilde{\mathbf{x}}_i$ as $\widetilde{\mathbf{x}}_i = \left(\widetilde{\mathbf{x}}_{(k),i}^\mathrm{T}, \widetilde{\mathbf{n}}_{(k),i}^\mathrm{T}\right)^\mathrm{T}$, where $\widetilde{\mathbf{x}}_{(k),i}$ contains the first k elements of $\widetilde{\mathbf{x}}_i$, so that the first principal minor of $\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}$ can be written as $\mathbf{\Sigma}_{(k)} = \mathbb{E}\left\{\widetilde{\mathbf{x}}_{(k),i}\,\widetilde{\mathbf{x}}_{(k),i}^\mathrm{H}\right\}$. Then the likelihood function of observed frequency bins can be written as

$$\begin{aligned} f\left(\widetilde{\mathbf{X}}\right) &= \prod_{i=1}^{N} f(\widetilde{\mathbf{x}}_i) = \prod_{i=1}^{N} \frac{1}{\pi^{N_\mathrm{b}}\left|\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}\right|} \exp\left(-\widetilde{\mathbf{x}}_i^\mathrm{H}\mathbf{\Sigma}_{\widetilde{\mathbf{x}}}{}^{-1}\widetilde{\mathbf{x}}_i\right) \\ &= \prod_{i=1}^{N} \frac{\exp\left(-\widetilde{\mathbf{x}}_{(k),i}^\mathrm{H}\mathbf{\Sigma}_{(k)}{}^{-1}\widetilde{\mathbf{x}}_{(k),i}\right)\exp\left(-\frac{1}{\sigma^2}\widetilde{\mathbf{n}}_{(k),i}^\mathrm{H}\widetilde{\mathbf{n}}_{(k),i}\right)}{\pi^{N_\mathrm{b}}\left|\mathbf{\Sigma}_{(k)}\right|\left(\sigma^2\right)^{N_\mathrm{b}-k}} \end{aligned}$$

and thus the log-likelihood can be expressed as [30]

$$\begin{aligned} \ln f\left(\widetilde{\mathbf{X}}\right) = &- N_\mathrm{b}N\ln\pi - N\ln\left|\mathbf{\Sigma}_{(k)}\right| - N\left(N_\mathrm{b}-k\right)\ln\sigma^2 \\ &- N\mathrm{tr}\left\{\mathbf{\Sigma}_{(k)}{}^{-1}\,\mathbf{S}_{(k)}\right\} - \frac{N}{\sigma^2}\mathrm{tr}\left\{\mathbf{N}_{(k)}\right\} \end{aligned} \tag{3}$$

where $\mathbf{S}_{(k)} = (1/N)\sum_{i=1}^{N}\widetilde{\mathbf{x}}_{(k),i}\,\widetilde{\mathbf{x}}_{(k),i}^\mathrm{H}$ and $\mathbf{N}_{(k)} = (1/N)\sum_{i=1}^{N}\widetilde{\mathbf{n}}_{(k),i}\,\widetilde{\mathbf{n}}_{(k),i}^\mathrm{H}$.

Considering the most general case, $\mathbf{\Sigma}_{(k)}$ is a non diagonal matrix in which the off-diagonal elements take into account the correlation among the frequency bins. However, it has been observed that a non negligible frequency correlation is experienced only in specific situations, such as in presence of observations with very short duration [31]. This is also confirmed by the fact that increasing the observation time the correlation between the DFT outputs tends to zero [32, Section 3.7].[1] Therefore, considering uncorrelated frequency bins, $\mathbf{\Sigma}_{(k)}$ is diagonal and the log-likelihood (3) reduces to

$$\begin{aligned} \ln f\left(\widetilde{\mathbf{X}};\mathbf{\Theta}^{(k)}\right) = &- N_\mathrm{b}N\ln\pi - N\sum_{q=1}^{k}\ln\sigma_q^2 \\ &- N\left(N_\mathrm{b}-k\right)\ln\sigma^2 \\ &- N\sum_{q=1}^{k}\frac{\widehat{\sigma}_q^2}{\sigma_q^2} - \frac{N}{\sigma^2}\mathrm{tr}\left\{\mathbf{N}_{(k)}\right\} \end{aligned} \tag{4}$$

where $(\sigma_1^2,\ldots,\sigma_k^2) = \mathrm{diag}\left\{\mathbf{\Sigma}_{(k)}\right\}$ and $(\widehat{\sigma}_1^2,\ldots,\widehat{\sigma}_k^2) = \mathrm{diag}\left\{\mathbf{S}_{(k)}\right\}$. In (4) we explicitly show the dependence of the log-likelihood on $\mathbf{\Theta}^{(k)}$, which is the vector that contains the unknown parameters, whose number, $\phi(k)$, depends on k. In our problem we have $\mathbf{\Theta}^{(k)} = \left(\sigma_1^2,\ldots,\sigma_k^2,\sigma^2\right)$, and thus $\phi(k) = k+1$.

III. Information Theoretic Criteria for Wideband Spectrum Sensing

For solving the MOS problem we adopt ITC, a statistical approach for choosing among a family of possible models the one that better describes the observed data [33], [34]. Considering that at least one frequency bin contains only noise, there are N_b possible models, where the kth model corresponds to the case in which the first k ordered bins are occupied, with $k \in \{0,\ldots,N_\mathrm{b}-1\}$.[2]

[1] Also note that considering OFDM transmissions, the presence of cyclic prefix causes uncorrelated frequency bins.

[2] We refer to the kth model also as the kth hypothesis.

According to ITC, the model that better fits the observed data is the one that minimizes the penalized likelihood metric [34], [35]

$$\mathsf{ITC}(k) = -2 \ln f\left(\widetilde{\mathbf{X}}; \widehat{\mathbf{\Theta}}^{(k)}\right) + \mathscr{P}(k) \tag{5}$$

where $f\left(\widetilde{\mathbf{X}}; \widehat{\mathbf{\Theta}}^{(k)}\right)$ is the likelihood, $\widehat{\mathbf{\Theta}}^{(k)}$ is the vector of the estimated parameters in the kth hypothesis, and $\mathscr{P}(k)$ is the ITC penalty.[3] Thus the order of the model is estimated as[4]

$$\widehat{k} = \arg\min_{k} \mathsf{ITC}(k)\,, \quad \text{with} \quad k \in \{0, 1, \ldots, N_{\mathrm{b}} - 1\}\,. \tag{6}$$

The choice of the penalty in (5) defines the criterion adopted, determining its performance and complexity [33]. In the next section we review the most common techniques which have been also adopted throughout the paper.

In the WS problem considered the unknown parameters vector can be estimated as $\widehat{\mathbf{\Theta}}^{(k)} = \left(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_k^2, \widehat{\sigma}^2\right)$ where $\widehat{\sigma}_q^2 = (N) \sum_{i=1}^{N} |x_{i,q}|^2$, and

$$\widehat{\sigma}^2 = \frac{\mathrm{tr}\left\{\mathbf{N}_{(k)}\right\}}{(N_{\mathrm{b}} - k)}. \tag{7}$$

Therefore, the log-likelihood in (5) can be expressed as

$$\ln f\left(\widetilde{\mathbf{X}}; \widehat{\mathbf{\Theta}}^{(k)}\right) = -N \sum_{q=1}^{k} \ln \widehat{\sigma}_q^2 - N\,(N_{\mathrm{b}} - k) \ln \widehat{\sigma}^2 + \mathrm{c} \tag{8}$$

where c is a term that does not depend on k and thus has no effect on the minimization in (6). Then, using (8) in (5) and (6) the estimated number of occupied frequency bins is represented by $\widehat{k}$. Also note that from (7) we obtain the noise power estimate $\mathrm{tr}\left\{\mathbf{N}_{(\widehat{k})}\right\} / \left(N_{\mathrm{b}} - \widehat{k}\right)$.

The advantage of using ITC is that they do not require any a priori knowledge of parameters such as the noise power or PUs characteristics. In addition, they do not require the setting of thresholds, avoiding problems such as deriving a threshold selection rule. The unique requirement of the proposed WS approach is to have a model of the frequency bins statistic. However, since that ITC can be also applied when the true model of the observed data does not belong to the family set considered [34], [37]–[39], in Section V we show that the proposed WS technique is effective also when only an approximate model of the frequency bins statistic is available. This situation is generally referred as a MOS problem with misspecified models [39].

A. Review of fundamental criteria

In [35] Akaike first proposed an information theoretic criterion for statistical model selection deriving what is now called the Akaike information criterion (AIC), in which the penalty is

$$\mathscr{P}_{\mathrm{AIC}}(k) = 2\,\phi(k) \tag{9}$$

where $\phi(k)$ is the number of free parameters in $\mathbf{\Theta}^{(k)}$. Although the AIC metric provides an unbiased estimate of the Kullback-Leibler (K-L) divergence between the likelihood in the kth hypothesis and the likelihood in the k^* case, in many situations it overestimates the order of the model, even asymptotically [33], [40], [41].

Alternative ITC can be derived adopting the Bayesian approach, which chooses the model maximizing the posterior probability $\mathbb{P}\left\{\mathbf{\Theta}^{(k)} \middle| \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right\}$. In this context, the most simple criterion is the Bayesian information criterion (BIC) with penalty [42]

$$\mathscr{P}_{\mathrm{BIC}}(k) = \phi(k)\, \ln N. \tag{10}$$

For large enough samples BIC coincides with the minimum description length (MDL) criterion, which attempts to construct a model that permits the shortest description of the data [43]. The AIC and BIC are the most popular ITC adopted in statistical and engineering problems (see e.g. [24], [33], [41], [44]–[47]).

Besides (9) and (10), many other ITC present a penalty in the form

$$\mathscr{P}_{\mathrm{GIC}}(k) = \phi(k)\, \nu \tag{11}$$

where ν can be a constant (as in (9)) or a function of the system parameters (as in (10)). We refer to this approach as the GIC [33]. Other examples of GIC are given, e.g., in [40], where $\nu = \ln N + 1$ is used, and in [46], where $\nu = 2 \ln N$ has been adopted.

B. MOS performance metrics

In this section we introduce the performance metrics that will be used in Section IV for the penalty design and in Section V for the numerical results. The performance of WS is evaluated in terms of probability to correctly estimate k^*, $\mathrm{P_c} = \mathbb{P}\left\{\widehat{k} = k^*\right\}$, and the probability of correctly identifying the set of occupied frequency bins, $\mathrm{P_{id}}$. The error probabilities are evaluated in terms of probability of overestimation, $\mathrm{P_{over}} = \mathbb{P}\left\{\widehat{k} > k^*\right\}$, and probability of underestimation, $\mathrm{P_{under}} = \mathbb{P}\left\{\widehat{k} < k^*\right\}$, which can be used to express $\mathrm{P_c}$ as

$$\mathrm{P_c} = 1 - \mathrm{P_{over}} - \mathrm{P_{under}}. \tag{12}$$

Note that the false alarm probability, commonly adopted in binary hypothesis testing, coincides with $\mathrm{P_{over}}$ when no frequency bins are occupied, i.e.

$$\mathrm{P_{FA}} = \mathrm{P_{over}}\big|_{k^*=0}\,. \tag{13}$$

An additional performance metric useful in practice is the probability of detection related to the qth bin, $\mathrm{P_D^q}$.

Evaluating the MOS error performance generally requires to take into account both underestimation and overestimation events. This holds in particular in low SNR regions in which

[3] Using the notation $\mathscr{P}(k)$ we emphasize that the penalty depends on k, which is important for the minimization in (6). Note that in general $\mathscr{P}(k)$ could also depend on further parameters, e.g. N_{b}, N, and other functions of the observation.

[4] Note that ITC, defined by (5) and (6), can be considered as extensions of the maximum likelihood (ML) principle in the form of penalized likelihoods, where the penalty is introduced as a cost for the increased complexity of the model, related to the presence of unknown parameters that must be estimated [35], [36].

$\mathrm{P}_{\text{under}}$ and P_{over} are both non negligible. However, increasing the SNR numerical results show that for finite sample sizes $\mathrm{P}_{\text{under}}$ goes to zero, while P_{over} generally converges to a non zero value. Several works in literature, such as [47] and [48], report this behavior and in Section V we show some simulation results that confirm it. Thus for high SNR an incorrect detection always consists in an overestimation, and the probability of correct model selection can be expressed as

$$\mathrm{P}_{\text{c}} \simeq 1 - \mathrm{P}_{\text{over}} \quad \text{(high SNR regime)}. \tag{14}$$

Note that this is a very favorable property in CR scenarios, implying that ITC never misdetect the presence of PUs if the SNR is sufficiently high.

IV. Design of the Generalized Information Criterion

Although ITC have been widely adopted for solving MOS problems, there are relatively few works that address the derivation of the penalty as a design problem. Most of them study the consistency of MOS, defining conditions under which asymptotically, for large N, the correct model order is selected [37], [49]. In particular, considering GIC, it has been proved that it is required for N that goes to infinity that $\nu/N \to 0$ to avoid underestimation and $\nu/\ln\ln N \to +\infty$ to avoid overestimation [37]. However, in practice, where finite sample sizes are used, they are not sufficient for controlling the error probabilities.[5] In this section we address the design of GIC for finite samples problems, following the approach proposed in [50].

A. Design strategy to control sensing performance

Considering (5) and (6), given that $\mathscr{P}(k)$ is an increasing function of k, it is easy to see that for large penalties the selection of a small model order is favored, and thus a lower P_{over} is provided. However, this gives a higher $\mathrm{P}_{\text{under}}$ and thus choosing the penalty implies a tradeoff between overestimation and underestimation. See Section V-A for some numerical examples. Given a target maximum overestimation probability, $\mathrm{P}_{\text{over}}^{\text{MAX}}$, we choose the penalty that minimizes $\mathrm{P}_{\text{under}}$ while guaranteeing $\mathrm{P}_{\text{over}} \leq \mathrm{P}_{\text{over}}^{\text{MAX}}$. Considering the overestimation-underestimation trade-off, $\mathrm{P}_{\text{under}}$ is minimized when P_{over} is maximum, and thus our sensing design strategy can be expressed as

$$\widetilde{\nu} = \arg\max_{\nu} \left\{ \mathrm{P}_{\text{over}} \middle| \, \mathrm{P}_{\text{over}} \leq \mathrm{P}_{\text{over}}^{\text{MAX}} \right\}. \tag{15}$$

Note that P_{over} is in general a function of k^* and SNR, which are unknown. Thus, adopting a worst case approach, we modify the design rule (15) including the maximization on these parameters obtaining

$$\begin{aligned}\widetilde{\nu} &= \arg\max_{\nu} \max_{k^*,\mathsf{SNR}} \left\{ \mathrm{P}_{\text{over}}\left(k^*, \mathsf{SNR}, \nu\right) \middle| \, \mathrm{P}_{\text{over}} \leq \mathrm{P}_{\text{over}}^{\text{MAX}} \right\} \\ &= \arg\max_{\nu} \max_{k^*} \left\{ \mathrm{P}_{\text{over}}\left(k^*, \infty, \nu\right) \middle| \, \mathrm{P}_{\text{over}} \leq \mathrm{P}_{\text{over}}^{\text{MAX}} \right\} \end{aligned} \tag{16}$$

[5] In CR scenarios the time available for sensing is severely constrained making the finite samples analysis crucial.

where the last form is due to the fact that P_{over} increases with SNR. In the next section we study the high SNR expression of P_{over} for GIC. This analysis allows to provide an explicit form for the computation of (16).

B. Probability of overestimation

The function $\mathsf{ITC}(k)$ defined in (5) is in general a sum of two monotonic functions in k. Indeed, minus two the log-likelihood of the observation decreases with k, while the penalty is an increasing function. In most of selection problems with nested models if the penalty is properly designed $\mathsf{ITC}(k)$ is a convex function with a unique minimum, that in case of correct model selection corresponds to $k = k^*$ [34]. Therefore, in order to analyze overestimation it has been shown that in many MOS problems it is sufficient to study which is the minimum for $k = \{k^*, k^*+1\}$, giving [47], [48], [51], [52]

$$\mathrm{P}_{\text{over}} \simeq \mathbb{P}\{\mathsf{ITC}(k^*) > \mathsf{ITC}(k^*+1)\} \tag{17}$$

where P_{over} is defined for $k^* \in \{0, 1, \ldots, N_{\text{b}} - 2\}$. In the following we adopt (17) for our problem in agreement with numerical results presented in Section V-A, that confirm the tight approximation for high SNR.

Using (5) and (8) in (17) we obtain (18) that can be rewritten as

$$\mathrm{P}_{\text{over}} \simeq \mathbb{P}\Big\{ y\left(1-y\right)^{N_{\text{b}}-k^*-1} < \xi_{k^*} \Big\} \tag{19}$$

with

$$y = \frac{\widehat{\sigma}^2_{k^*+1}}{\sum_{q=k^*+1}^{N_{\text{b}}} \widehat{\sigma}^2_q} \tag{20}$$

and

$$\xi_{k^*} = \frac{\left(N_{\text{b}} - k^* - 1\right)^{N_{\text{b}}-k^*-1}}{\left(N_{\text{b}} - k^*\right)^{N_{\text{b}}-k^*}} \exp\left(\frac{\mathscr{P}(k^*) - \mathscr{P}(k^*+1)}{2N} \right). \tag{21}$$

The equation $y\left(1-y\right)^{N_{\text{b}}-k^*-1} = \xi_{k^*}$ has one single root in $\left[1/\left(N_{\text{b}} - k^*\right), 1\right]$, the range of y, which can be easily computed using standard root finding algorithms. Denoting this root with y, (18) can be expressed by the CDF of y, $F_y(\cdot)$, i.e.

$$\mathrm{P}_{\text{over}} = 1 - F_y(\mathsf{y}). \tag{22}$$

In the Appendix we propose two approximations of $F_y(\cdot)$ for computing P_{over}.

C. Penalty design for DFT-based WS

From (22) and (38) we can see that P_{over} is a decreasing function of k^* and thus the maximum probability of overestimation is reached for $k^* = 0$. Then the penalty design (16) reduces to

$$\widetilde{\nu} = \arg\max_{\nu} \left\{ \mathrm{P}_{\text{over}}\left(0, \infty, \nu\right) \middle| \, \mathrm{P}_{\text{over}} \leq \mathrm{P}_{\text{over}}^{\text{MAX}} \right\}. \tag{23}$$

Using (22), we have $\mathrm{P}_{\text{over}}\left(0, \infty, \nu\right) = 1 - F_{y|_{k^*=0}}(\mathsf{y})$ and thus the value of y that corresponds to $\mathrm{P}_{\text{over}}^{\text{MAX}}$ is

$$\widetilde{\mathsf{y}} = F^{-1}_{y|_{k^*=0}}\left(1 - \mathrm{P}_{\text{over}}^{\text{MAX}}\right). \tag{24}$$

$$\mathrm{P}_{\text{over}} \simeq \mathbb{P}\left\{2\ln f\left(\widetilde{\mathbf{X}};\widehat{\boldsymbol{\Theta}}^{(k^*+1)}\right) - 2\ln f\left(\widetilde{\mathbf{X}};\widehat{\boldsymbol{\Theta}}^{(k^*)}\right) > \mathscr{P}(k^*+1) - \mathscr{P}(k^*)\right\}$$
$$= \mathbb{P}\left\{\ln\widehat{\sigma}^2_{k^*+1} + (N_\text{b}-k^*-1)\ln\left(\frac{\sum_{q=k^*+2}^{N_\text{b}}\widehat{\sigma}^2_q}{N_\text{b}-k^*-1}\right) - (N_\text{b}-k^*)\ln\left(\frac{\sum_{q=k^*+1}^{N_\text{b}}\widehat{\sigma}^2_q}{N_\text{b}-k^*}\right) < \frac{\mathscr{P}(k^*)-\mathscr{P}(k^*+1)}{2N}\right\}. \quad (18)$$

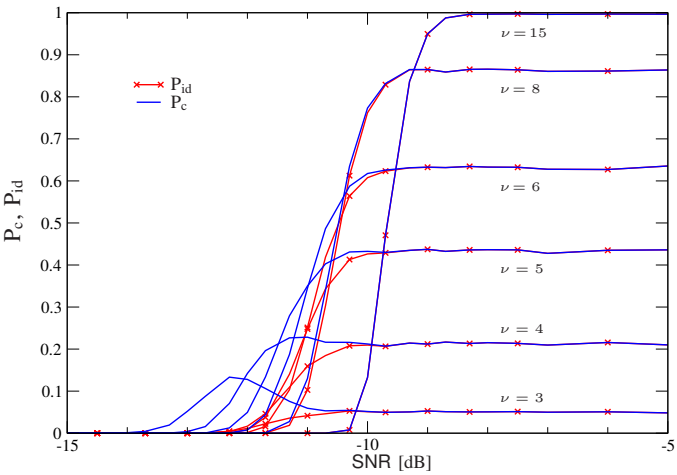


Fig. 3. Probability to correctly estimate the number of occupied frequency bins, $\mathrm{P_c}$ (blue lines), and probability to detect the occupied frequency bins, $\mathrm{P_{id}}$ (red lines with crosses), as function of the SNR for GIC. The number of occupied bins is $k^* = 64$, $N_\text{b} = 128$, and $N = 1000$.

Considering $k^* = 0$, the argument of (18) is

$$\widetilde{\mathrm{y}}\,(1-\widetilde{\mathrm{y}})^{N_\text{b}-1} = \frac{(N_\text{b}-1)^{N_\text{b}-1}}{N_\text{b}^{N_\text{b}}}\exp\left(\frac{\mathscr{P}_\text{GIC}(0)-\mathscr{P}_\text{GIC}(1)}{2N}\right)$$
$$= \frac{(N_\text{b}-1)^{N_\text{b}-1}}{N_\text{b}^{N_\text{b}}}\exp\left(-\frac{\widetilde{\nu}}{2N}\right) \quad (25)$$

from which we finally derive the desired GIC penalty

$$\widetilde{\nu} = -2N\ln\left(\frac{{N_\text{b}}^{N_\text{b}}}{(N_\text{b}-1)^{N_\text{b}-1}}\widetilde{\mathrm{y}}\,(1-\widetilde{\mathrm{y}})^{N_\text{b}-1}\right). \quad (26)$$

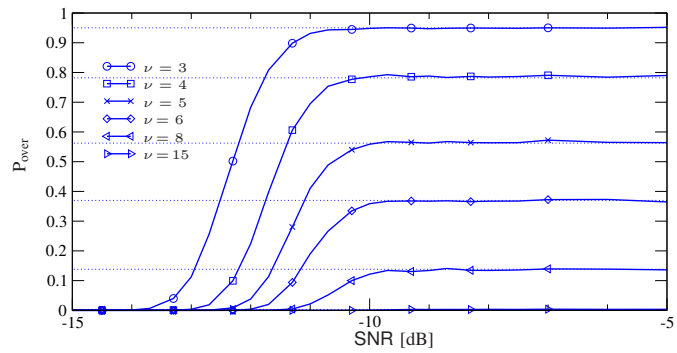
Note that (23) reveals that the design approach based on $\mathrm{P}^{\text{MAX}}_{\text{over}}$, thanks to (13), is equivalent to satisfy a requirement on the probability of false alarm. Thus our design strategy is analogous to the Neyman-Pearson criterion adopted in binary hypothesis testing, in which $\mathrm{P}_{\text{under}}$ corresponds to the probability of misdetection.

V. Numerical results

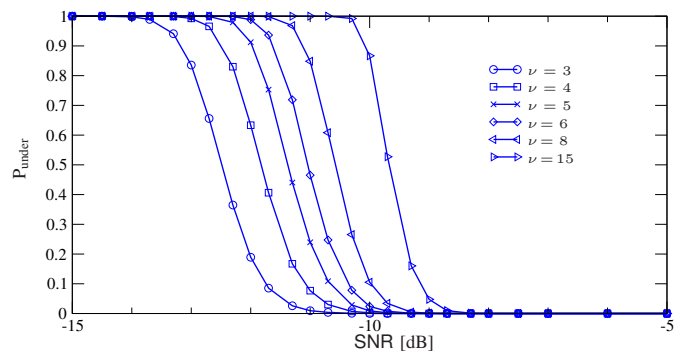
A. Performance analysis of GIC-based WS

In this section, we show some numerical results of the performance metrics $\mathrm{P_c}$, $\mathrm{P_{id}}$, $\mathrm{P_{over}}$, and $\mathrm{P_{under}}$ for the proposed DFT-based WS approach, focusing on the GIC for different values of ν. To evaluate the performance of the algorithm, we consider Gaussian band-limited signals with flat PSD.[6]

[6]Note that, using real signals, the number of occupied bins also depends on sideband leakage.



(a) Probability of overestimation



(b) Probability of underestimation

Fig. 4. Probabilities to overestimate and underestimate the number of occupied bins as function of the SNR for GIC. The number of occupied bins is $k^* = 64$, $N_\text{b} = 128$, and $N = 1000$. The dotted lines correspond to $\mathrm{P_{over}}$ for high SNR computed using (22) and (37).

In Fig. 3 we consider the WS strategy proposed with $N_\text{b} = 128$, in presence of a signal that occupies $k^* = 64$ bins and AWGN. The number of DFT outputs collected is $N = 1000$. For all criteria $\mathrm{P_{id}}$ presents a sigmoidal dependence on the SNR, reaching a maximum value that increases with ν. Note that the difference between $\mathrm{P_c}$ and $\mathrm{P_{id}}$ is small, ensuring that generally when the algorithm correctly estimates k^* it also correctly identifies the occupied set as well. It is also evident that a high ν moves the rise of $\mathrm{P_c}$ towards higher SNRs. Considering the same scenario, in Fig. 4 we show the corresponding probabilities of incorrect order selection. We can see that for high SNR $\mathrm{P_{under}}$ goes to zero and an incorrect detection always consists in an overestimation. Fig. 4(a) also confirms that the maximum $\mathrm{P_{over}}$ is correctly predicted using (22) and (37) (horizontal dotted lines).

B. Performance comparison among ITC and ED

The GIC with penalty designed as described in Sections IV-A is now compared with AIC, BIC (adopted in [24],

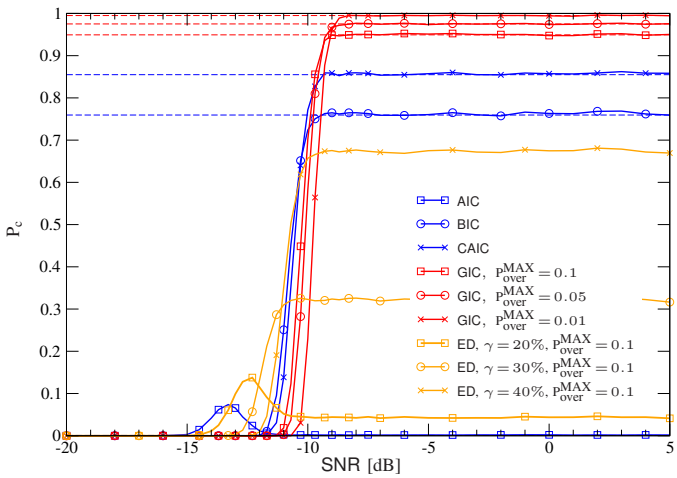


Fig. 5. Probability to correctly estimate the number of occupied bins as function of the SNR. The number of occupied bins is $k^* = 64$, $N_b = 128$, and $N = 1000$. The dashed lines correspond to the maximum P_c computed using (14), (22) and (37).

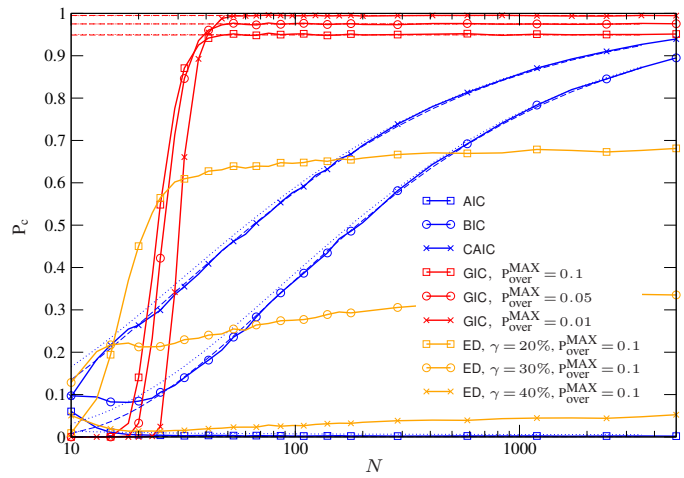


Fig. 6. Probability to correctly estimate the number of occupied bins as function of the number of observations N. The simulation has been performed considering $k^* = 64$, $N_b = 128$, and $\mathsf{SNR} = 0\,\text{dB}$. The dashed and dotted lines correspond to P_c computed from (14) and (22) using (37) and (38), respectively.

[25]), and the consistent AIC (CAIC) with $\nu = 1 + \ln N$. Unless otherwise specified we consider $N = 1000$, which gives $\nu = 6.90$ and 7.90 for BIC and CAIC, respectively. For GIC the penalty is designed considering $P_{over}^{MAX} = 0.1$, 0.05 and 0.01, which gives $\nu \simeq 9.88$, 11.18 and 14.14, respectively. The performance of ITC-based WS is also compared with a detection scheme where an ED is adopted for each frequency bin, as proposed in [22]. In this case, according to [22] and [23], we assume that at least a known fraction of the N_b bins contains only noise, so that the $\gamma \cdot N_b$ bins with lower received power are used for noise power estimation. Then, the ED threshold is set in order to keep the overestimation probability in the $\gamma \cdot N_b$ noise-only bins below P_{over}^{MAX}.

In the same setting of Section V-A, Fig. 5 shows the probability P_c as a function of the SNR for ITC-based and ED-based WS. As can be seen the high SNR performance is always correctly predicted, confirming the efficacy of the design strategy and the analysis proposed in Sections III-B and IV-B. Note that maximum P_c for BIC is about 0.76, while the AIC confirms its non consistent behavior providing almost zero probability of correct MOS. Also note that the ED-based approach with $\gamma = 20\%$, 30%, and 40% performs quite poorly due to the insufficient accuracy of noise power estimation[7] [4].

In Fig. 6 we show P_c as function of the number of collected DFT outputs N with $k^* = 64$, $N_b = 128$, and $\mathsf{SNR} = 0\,\text{dB}$. The curves confirm that the performance of the consistent ITC (i.e. all except AIC) improves with N. In the same figure we also show the theoretical values of P_c computed using (37) (dashed lines) and (38) (dotted lines). When N is sufficiently high (approximatively $N > 25$) using the gamma approximation we correctly predict the P_c obtained by simulations. The discrepancy between theoretical and experimental curves for small N is justified by the incorrect ordering caused by the inaccurate estimates $\left\{\widehat{\sigma}_q^2\right\}$, i.e., the power of the frequency bins. Note that the simpler approximation proposed in the Appendix (dotted lines) matches the simulated curves for high values of P_c.

The probability of correct bin selection as function of the actual number of occupied bins k^* is plotted in Fig. 7, where the SNR for the single bin, $\mathsf{SNR}^{(\text{bin})} = \mathbb{E}\left\{|x_{q,i}|^2\right\}/\sigma^2$, is assumed to be constant for all i and q, and equal to -5 or $0\,\text{dB}$. Also in this case we can see that the proposed GIC outperforms AIC, BIC and CAIC. Note that for all the ITC the probability of correct model selection increases, demonstrating that the worst case design with $k^* = 0$ provides a high performance until some high value of k^*, beyond which P_c drops rapidly. This is mainly related to imperfect noise power estimation because for high k^* the frequency bins used to estimate σ^2 (see (7)) are too few. Note, however, that increasing the SNR this effect is less evident, with a drop that moves to higher k^* values.

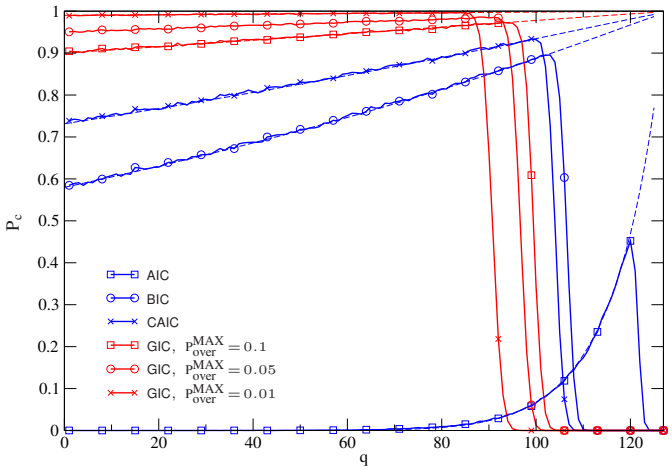
In Fig. 8 we perform the same analysis of Fig. 5 in a Rayleigh fading scenario. Here we consider time-frequency block fading where each block size is 16 bins $\times$ 250 time samples. With respect to the AWGN case the effect of fading on the performance of all the algorithms (both ITC and ED based) is to increase the SNR value at which the target probability of correct detection is reached.

C. Case study with multiple PUs and different spectral estimates

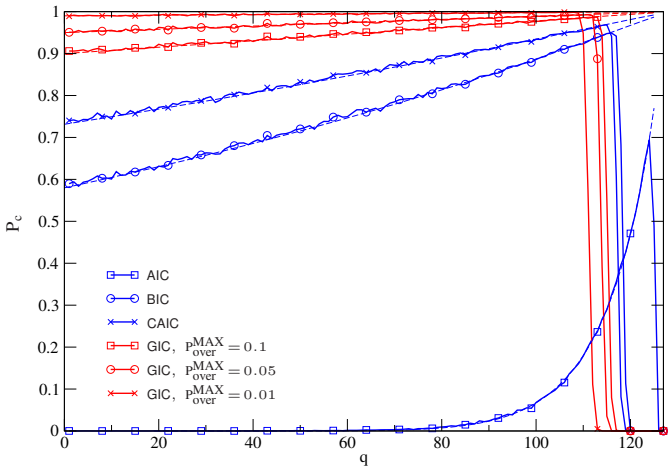
We now analyze a multiband scenario in which three OFDM-like signals are present in the observed bandwidth. The PSD of the three signals is depicted in Fig. 9. In the following we denote with SNR the SNR of each of the two lower frequency signals, while the higher frequency signal has a SNR drop of $-3\,\text{dB}$. In Fig. 9 we also plot the noise spectral density for the $\mathsf{SNR} = -10\,\text{dB}$ case, which is adopted in the following.

1) DFT-based wideband sensing: In Fig. 10 we show P_D^q when the wideband algorithm proposed in Section IV

[7]The low ED performance is also accompanied by the need to a priori know the number of noise-only bins, which makes the approach not blind.



(a) $\mathsf{SNR}^{(\mathrm{bin})} = -5\,\mathrm{dB}$



(b) $\mathsf{SNR}^{(\mathrm{bin})} = 0\,\mathrm{dB}$

Fig. 7. Probability to correctly estimate the number of occupied bins as function of the actual number of occupied bins k^*. The simulation has been performed considering $N_b = 128$ and $N = 1000$. The dashed lines correspond to the maximum P_c computed using (14), (22) and (37).

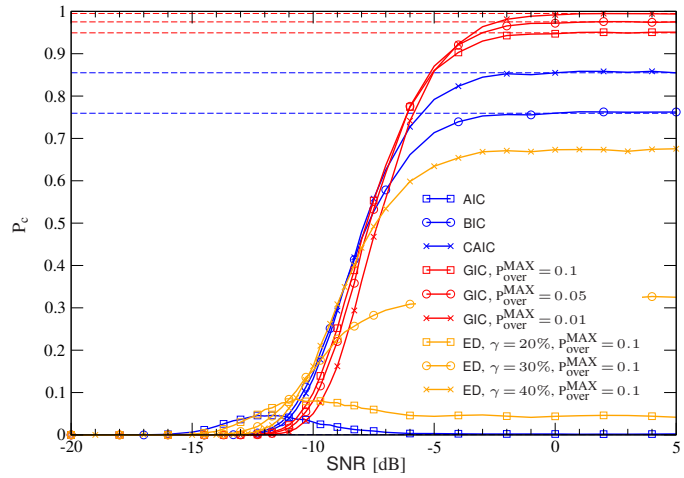


Fig. 8. Probability to correctly estimate the number of occupied bins as function of the SNR in Rayleigh fading. The number of occupied bins is $k^* = 64$, $N_b = 128$, and $N = 1000$. The dashed lines correspond to the maximum P_c computed using (14), (22) and (37).

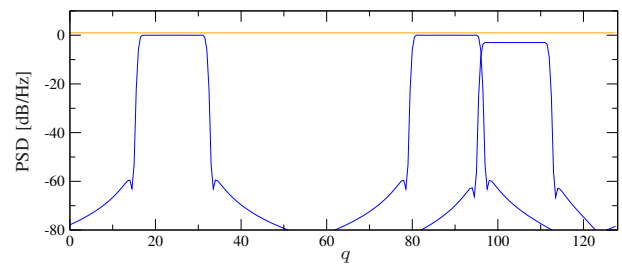


Fig. 9. PSD of the signals in a scenario with multiple PUs as a function of the frequency bin index, q. PSDs are normalized to 0 dB/Hz while the higher frequency signal has a SNR drop of $-3\,\mathrm{dB}$. The noise PSD level for the $\mathsf{SNR} = -10\,\mathrm{dB}$ case is also plotted (orange line).

is adopted. Note that compared to other criteria the AIC provides a high probability of detection also for out-of-band frequencies.[8] This is in accordance with the fact that, as noted in Section III-A, AIC tends to overestimate the number of occupied bins. We can also see that all ITC provide a high P_c at $\mathsf{SNR} = -10\,\mathrm{dB}$.

2) MOS based WS in case of misspecified models: ITC are conceived for being statistical approaches that choose the model that best represents data among a family of models. It is interesting therefore to analyze if they provide a good detection performance also when the true model of the observed signal is not in the considered model set. This analysis is interesting for practical situations in which either the exact statistical description of the collected data is not known, or it is too complex to calculate the likelihood in (5), and thus algorithms derived for simpler models can be adopted as approximated approaches. Here we consider the case in which the spectral representation is obtained with the WOSA [53] or the MT [7] spectrum estimates, and we apply the WS technique derived for the DFT in Section IV.[9] This can be considered as an approximated strategy in which these spectrum estimates, that in general are chi squared distributed, are approximated to Gaussian r.v.s [55].[10] In Fig. 11(a) and Fig. 11(b) we can see that WOSA and MT estimates provide very good detection performances over the DFT-based approach for low SNR levels. Then we benefit from the adoption of better spectrum estimates (DFT has a non negligible spectral leakage). Note that also in this case AIC presents a high probability of detection in the noise only bands.

VI. Conclusions

In this paper we proposed a WS technique based on MOS. We derived a general approach that can be applied to any spectral representation and discussed in detail the case in which DFT is used. This WS technique is completely blind since it does not require any knowledge about the noise power and the characteristics of the signals present in the observed

[8]Note that in general frequency occupancy should also consider the problem of out-of-band emissions, such as OFDM sidelobes.

[9]Also note that using DFT with different windows, such as e.g. Hanning, Kaiser, Chebychev windows, could be an alternative approach [54].

[10]Note that this approximation is valid when the chi squared distribution has a high number of d.o.f..

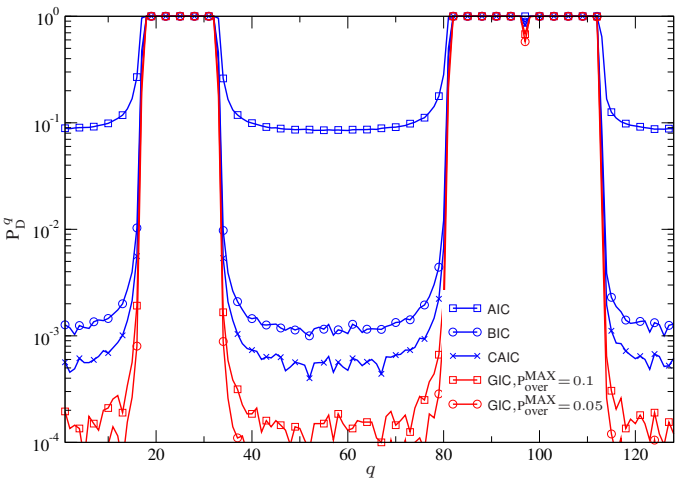
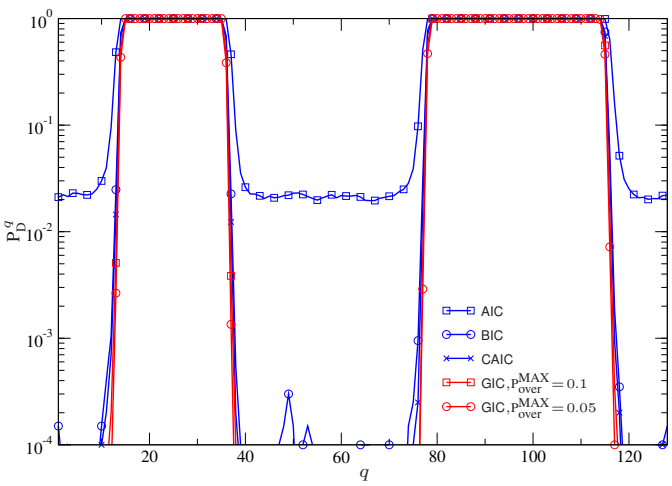
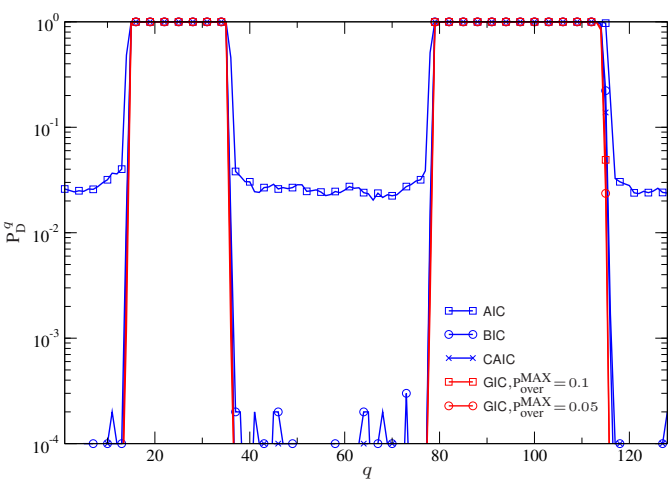


Fig. 10. Probability of detection for each frequency bin, when DFT is adopted as a spectral estimate with $N_b = 128$, and $N = 1000$. The GIC penalty is designed for $P_{over}^{MAX} = 0.1$, and 0.05.



(a) P_D^q with WOSA spectrum estimate



(b) P_D^q with MT spectrum estimate

Fig. 11. Probability of detection for each frequency bin when 128-points spectrum estimates based on WOSA and MT methods are adopted with $N = 1000$. The GIC penalty is designed for $P_{over}^{MAX} = 0.1$, and 0.05.

band. In particular, we proposed a performance analysis in the high SNR regime which led to a tight approximation of the probability of detecting the occupied bins for all the criteria that can be enclosed in the GIC class. Such analysis has been used to design the penalty that balances the trade-off between overestimation and underestimation. Finally, we showed that the DFT-based algorithm derived can be successfully applied as an approximated approach in situations in which the exact distribution of the spectral observation is unknown or too complex, such as in the case of MT spectrum estimation.

Appendix

According to (22) the probability of overestimation can be evaluated using the statistic of the parameter y defined in (20). The high SNR regime considered in Section IV-B implies perfect ordering, i.e., ordering performed under reliable estimation of the received power in each bin, $\left\{\widehat{\sigma}_q^2\right\}_{q=1,\ldots,N_b}$. Note that the estimation of $\widehat{\sigma}_q^2$ does not depend only on the SNR but also on N. However, it is always verified that perfect ordering is guaranteed if the SNR is sufficiently high. Therefore the last $N_b - k^*$ bins in the ordered vectors are noise only bins, and thus the distribution of y can be derived by the noise statistic. Although deriving the exact distribution of y is a very difficult task, the evaluation of its exact moments can be performed as described below. Thus, an approximated distribution for y can be derived using the method of moments [46], [56], [57]. In the following we derive the exact moments of y and two approximations of $F_y(\cdot)$ based on the maximum of a set of $N_b - k^*$ independent, identically distributed (i.i.d.) r.v.s $z_q = \mu_q/t$, where $2N\mu_q \sim \chi^2_{2N}$, $t = \sum_{q=1}^{N_b-k^*} \mu_q$, and $q = \{1, \ldots, N_b - k^*\}$.

A. Moments

The joint distribution of the set of chi squared distributed r.v.s $\{\mu_q\}_{q=\{1,\ldots,N_b-k^*\}}$ is given by

$$\frac{(N_b - k^*)!\, N^{N(N_b-k^*)}}{(\Gamma(N))^{N_b-k^*}} \prod_{q=1}^{N_b-k^*} \mu_q^{N-1} \exp\left(-N\mu_q\right).$$

Using the change of variables $\mu_q = z_q \cdot t$ we obtain the distribution

$$\frac{(N_b - k^*)!\, N^{N(N_b-k^*)}}{(\Gamma(N))^{N_b-k^*}}\, t^{N(N_b-k^*)-1} e^{-Nt} \prod_{q=1}^{N_b-k^*} z_q^{N-1} \quad (27)$$

where the Jacobian of the transformation is $t^{N_b-k^*}$. The factorized form of (27) implies that t is independent from the set $\{z_q\}_{q=\{1,\ldots,N_b-k^*\}}$, and thus the pth moment of μ_q are given by

$$\mathbb{E}\left\{\mu_q^p\right\} = \mathbb{E}\left\{z_q^p\right\} \cdot \mathbb{E}\left\{t^p\right\}. \quad (28)$$

Thus the moments of y can be expressed as

$$\mathbb{E}\left\{y^p\right\} = \mathbb{E}\left\{\tilde{\mu}_1^p\right\} / \mathbb{E}\left\{t^p\right\} \quad (29)$$

where $\tilde{\mu}_1 = \max \mu_q$.

The exact moments of $\tilde{\mu}_1$ are given by [58]

$$\mathbb{E}\left\{\tilde{\mu}_1^p\right\} = \frac{(N_\mathrm{b}-k^*)\Gamma(p+N(N_\mathrm{b}-k^*))}{N^{p+N_\mathrm{b}-k^*-1}\Gamma(N)^{N_\mathrm{b}-k^*}} \times F_A^{N_\mathrm{b}-k^*-1}\left(p+N\left(N_\mathrm{b}-k^*\right);\ N,\ldots,N;\ N+1,\ldots,N+1;-1,\ldots,-1\right) \quad (30)$$

where F_A^n is a Lauricella function of type A in n variables. Given that the implementation of the Lauricella function can be cumbersome, alternative forms for the computation of the moments of $\tilde{\mu}_1$ are given, e.g., by the expressions proposed in [59]. Note, however, that these forms can be computationally quite expensive, especially for large number of d.o.f. and large sets. Alternatively, we can numerically compute $\mathbb{E}\left\{\tilde{\mu}_1^p\right\}$ using the integral representation of the moments given by [60]

$$\mathbb{E}\left\{\tilde{\mu}_1^p\right\} = (N_\mathrm{b}-k^*)\int_0^1 \left(F_{\mu_q}^{-1}(x)\right)^p x^{N_\mathrm{b}-k^*-1}dx \quad (31)$$

where $F_{\mu_q}(x) = \gamma(N, Nx)$ is the CDF of μ_q. Some approximated methods for the computation of (31) are provided in [61].

The moments $\mathbb{E}\left\{t^p\right\}$ can be derived considering that t is proportional to a chi squared distributed r.v. with $2N\left(N_\mathrm{b}-k^*\right)$ d.o.f., and thus we have

$$\mathbb{E}\left\{t^p\right\} = \frac{\Gamma(p+N\left(N_\mathrm{b}-k^*\right))}{N^p\,\Gamma\left(N\left(N_\mathrm{b}-k^*\right)\right)}. \quad (32)$$

Therefore the moment of y can be computed using (29), (30) (or (31)), and (32).

B. Moment matching approach

Once the moments of y are computed using (29), we approximate the distribution of y choosing a mathematically tractable distribution and setting its parameters to match the first moments. We choose to approximate y to a shifted Gamma distributed r.v. as [57]

$$y+\alpha \sim \mathcal{G}(\kappa,\theta) \quad (33)$$

where κ, θ and the shift α are computed as [57]

$$\kappa = \frac{4\left(\mathsf{m}_2-\mathsf{m}_1^2\right)^3}{\left(\mathsf{m}_3^3-3\mathsf{m}_1\mathsf{m}_2-4\mathsf{m}_1^3\right)^2} \quad (34)$$

$$\theta = \frac{\mathsf{m}_3^3-3\mathsf{m}_1\mathsf{m}_2-4\mathsf{m}_1^3}{2\left(\mathsf{m}_2-\mathsf{m}_1^2\right)} \quad (35)$$

$$\alpha = \frac{2\left(\mathsf{m}_2-\mathsf{m}_1^2\right)^2}{\mathsf{m}_3^3-3\mathsf{m}_1\mathsf{m}_2-4\mathsf{m}_1^3} \quad (36)$$

with $\mathsf{m}_p = \mathbb{E}\left\{y^p\right\}$ and $p=1,2,3$. Thus the CDF of y is approximated as

$$F_y(x) \simeq \begin{cases} \gamma\left(\kappa, \frac{x+\alpha}{\theta}\right), & x > -\alpha \\ 0, & x \leq -\alpha \end{cases} \quad (37)$$

where $\gamma(a,z) \triangleq \frac{1}{\Gamma(a)}\int_0^z y^{a-1}e^{-y}dy$ is the incomplete gamma function and $\Gamma(a) \triangleq \int_0^\infty y^{a-1}e^{-y}dy$ is the gamma function. Note that in principle any distribution can be used for moment matching. We consider the Gamma distribution because y is a positive r.v., and we chose, in particular, a shifted Gamma distribution because it ensures a better matching (involving three moments) with respect to the conventional Gamma distribution (where only two moments are used).

C. Approximation based on the independence assumption

An alternative approximation for $F_y(x)$ can be derived under the assumption that the r.v.s μ_q are independent, which holds for large $N_\mathrm{b}-k^*$. This is equivalent to assume that z_q are i.i.d. beta distributed r.v.s with parameters N and $N\left(N_\mathrm{b}-k^*\right)$, respectively. Thus the approximated CDF can be derived as

$$F_y(x) \simeq \left(I_x(N, N\left(N_\mathrm{b}-k^*-1\right))\right)^{N_\mathrm{b}-k^*} \quad (38)$$

where $I_x(a,b) = \frac{1}{B(a,b)}\int_0^x z^{a-1}(1-z)^{b-1}dz$ is the incomplete beta function and $B(a,b) = \Gamma(a)\,\Gamma(b)\,/\Gamma(a+b)$ is the beta function with integer parameters.

The goodness of the two approximations (37) and (38) is investigated in Section V.

References

[1] CISCO, VNI Forecast Widget. [Online]. Available: http://www.ciscovni.com/vni_forecast/advanced.html.

[2] S. Kandeepan and A. Giorgetti, *Cognitive Radios and Enabling Techniques*. Boston, USA: Artech House Publishers, Nov. 2012.

[3] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 1, pp. 116–130, 2009.

[4] A. Mariani, A. Giorgetti, and M. Chiani, "Effects of noise power estimation on energy detection for cognitive radio applications," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3410–3420, Dec. 2011.

[5] M. Renfors, F. Bader, L. Baltar, D. Le Ruyet, D. Roviras, P. Mege, M. Haardt, and T. H. Stitz, "On the use of filter bank based multicarrier modulation for professional mobile radio," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring 2013)*, Dresen, Germany, Jun. 2013.

[6] A. Gorcin and H. Arslan, "Signal identification for adaptive spectrum hyperspace access in wireless communications systems," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 134–145, Oct. 2014.

[7] D. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[8] T. Erpek, A. Leu, and B. Mark, "Spectrum sensing performance in TV bands using the multitaper method," in *Proc. IEEE Signal Process. and Commun. Applications Conf. (SIU 2007)*, Eskisehir, Turkey, Jun. 2007.

[9] Q. T. Zhang, "Theoretical performance and thresholds of the multitaper method for spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2128–2138, Jun. 2011.

[10] D. Cabric, S. Mishra, and R. Brodersen, "Implementation issues in spectrum sensing for cognitive radios," in *Proc. Asilomar IEEE Conf. on Signals, Syst. and Comput.*, vol. 1, Pacific Grove, CA, USA, Nov. 2004, pp. 772–776.

[11] I. Akyildiz, W.-Y. Lee, M. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Communication Networks Journal (Elsevier)*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[12] V. Blaschke, T. Renk, and F. Jondral, "A cognitive radio receiver supporting wide-band sensing," in *Proc. IEEE Int. Conf. on Commun. (ICC 2008)*, Beijing, China, May 2008, pp. 499–503.

[13] N. Krasner, "Efficient search methods using energy detectors–maximum probability of detection," *IEEE J. Sel. Areas Commun.*, vol. 4, no. 2, pp. 273–279, Mar. 1986.

[14] S. Kandeepan, R. Piesiewicz, T. Aysal, A. Biswas, and I. Chlamtac, "Spectrum sensing for cognitive radios with transmission statistics: Considering linear frequency sweeping," *EURASIP J. on Wireless Commun. and Networking*, vol. 2010, no. 6, pp. 1–13, Mar. 2010.

[15] B. Farhang-Boroujeny, "Filter bank spectrum sensing for cognitive radios," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1801–1811, May 2008.

[16] T. Sato and M. Umehira, "A new spectrum sensing scheme using overlap FFT filter-bank for dynamic spectrum access," in *Proc. ICST Int. Conf. on Cognitive Radio Oriented Wireless Netw. and Commun. (CROWNCOM 2011)*, Osaka, Japan, 2011, pp. 6–11.

[17] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wireless Commun. Mag.*, vol. 20, no. 2, pp. 74–81, Apr. 2013.

[18] Z. Tian and G. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2007)*, vol. IV, Apr. 2007, pp. 1357–1360.

[19] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[20] H. Sun, W.-Y. Chiu, J. Jiang, A. Nallanathan, and H. V. Poor, "Wideband spectrum sensing with sub-Nyquist sampling in cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 6068–6073, Nov. 2012.

[21] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.

[22] A. Taherpour, S. Gazor, and M. Nasiri-Kenari, "Invariant wideband spectrum sensing under unknown variances," *IEEE Trans. Wireless Commun.*, vol. 8, no. 5, pp. 2182–2186, May 2009.

[23] A. J. Coulson, "Blind detection of wideband interference for cognitive radio applications," *EURASIP J. Advances Signal Process.*, vol. 2009, no. 8, pp. 1–12, 2009.

[24] S. Liu, J. Shen, R. Zhang, Z. Zhang, and Y. Liu, "Information theoretic criterion-based spectrum sensing for cognitive radio," *IET Communications*, vol. 2, no. 6, pp. 753–762, Feb. 2008.

[25] M. Fujii and Y. Watanabe, "A study on interference detection scheme using AIC for UWB MB-OFDM systems," in *Proc. Int. Workshop on Multi-Carrier Systems Solutions (MC-SS 2011)*, Herrsching, Germany, May 2011.

[26] B. Sadhu, M. Sturm, B. M. Sadler, and R. Harjani, "Analysis and design of a 5 GS/s analog charge-domain FFT for an SDR front-end in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 48, no. 5, pp. 1199–1211, May 2013.

[27] ——, "Building an on-chip spectrum sensor for cognitive radios," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 92–100, Apr. 2014.

[28] F. Rivet, Y. Deval, J. Begueret, D. Dallet, P. Cathelin, and D. Belot, "The experimental demonstration of a SASP-based full software radio receiver," *IEEE J. Solid-State Circuits*, vol. 45, no. 5, pp. 979–988, May 2010.

[29] S. M. Kay, *Fundamentals of Statistical Processing, Volume 1: Estimation Theory*. Prentice-Hall Signal Processing Series, 1993.

[30] A. Mariani, "Spectrum sensing algorithms for cognitive radio applications," Ph.D. dissertation, Philosophy Doctoral Program in Electronics, Computer Science and Telecommunications, Alma Mater Studiorum University of Bologna, Cesena, Italy, 2013.

[31] A. J. Coulson, "Do wireless data signals exhibit spectral autocorrelation?" in *Proc. IEEE Australian Commun. Theory Workshop (AusCTW 2008)*, Christchurch, New Zealand, Jan./Feb. 2008.

[32] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise*. San Diego: Academic Press, 1995.

[33] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[34] K. Burnham and D. Anderson, *Model selection and multi-model inference: a practical information-theoretic approach*. Springer Science & Business Media, 2002.

[35] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. Int. Symp. on Inf. Theory*, 1972, pp. 267–281.

[36] M. H. Hansen and B. Yu, "Model selection and the principle of minimum description length," *Journal of the American Statistical Association*, vol. 96, no. 454, pp. 746–774, Jun. 2001.

[37] R. Nishii, "Maximum likelihood principle and model selection when the true model is unspecified," *Journal of Multivariate Analysis*, vol. 27, no. 2, pp. 392–403, Nov. 1988.

[38] R. Shibata, "Asymptotically efficient selection of the order of the model for estimating parameters of a linear process," *The Annals of Statistics*, vol. 8, no. 1, pp. 147–164, Jan. 1980.

[41] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, pp. 387–392, Apr. 1985.

[39] C. J. Flynn, C. M. Hurvich, and J. S. Simonoff, "Efficiency for regularization parameter selection in penalized likelihood estimation of misspecified models," *Journal of the American Statistical Association*, vol. 108, no. 503, pp. 1031–1043, Sep. 2013.

[40] H. Bozdogan, "Akaike's information criterion (AIC): The general theory and its analytical extensions," *Psychometrica*, vol. 52, no. 3, pp. 345–370, Sep. 1987.

[42] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, Mar. 1978.

[43] J. Rissanen, "An introduction to the MDL principle," 2004. [Online]. Available: www.mdl-research.org

[44] M. Chiani and M. Z. Win, "Estimating the number of signals observed by multiple sensors," in *Proc. IEEE Int. Workshop on Cognitive Inf. Process. (CIP 2010)*, Elba Island, Italy, Jun. 2010, pp. 156–161.

[45] A. Giorgetti, A. Mariani, and M. Chiani, "Spectrum holes detection by information theoretic criteria," in *Proc. ACM Int. Conf. on Cognitive Radio and Advanced Spectrum Management (COGART 2011)*, Barcelona, Spain, Oct. 2011, Invited Paper.

[46] A. Giorgetti and M. Chiani, "Time-of-arrival estimation based on information theoretic criteria," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 1869–1879, Apr. 2013.

[47] Q. T. Zhang, K. M. Wong, P. C. Yip, and J. P. Reilly, "Statistical analysis of the performance of information theoretic criteria in the detection of the number of signals in array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 1557–1567, Oct. 1989.

[48] E. Fishler, M. Grossman, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1027–1036, May 2002.

[49] R. V. Foutz and R. C. Srivastava, "The performance of the likelihood ratio test when the model is incorrect," *The annals of Statistics*, vol. 5, no. 6, pp. 1183–1194, Nov. 1977.

[50] A. Mariani, A. Giorgetti, and M. Chiani, "Model order selection based on information theoretic criteria: design of the penalty," *IEEE Trans. Signal Process.*, vol. 63, no. 11, pp. 2779–2789, Jun. 2015.

[51] A. P. Liavas and P. A. Regalia, "On the behaviour of information theoretic criteria for model order selection," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1689–1695, Aug. 2001.

[52] A. Mariani, A. Giorgetti, and M. Chiani, "Wideband spectrum sensing for cognitive radio: a model order selection approach," in *Proc. IEEE Int. Conf. on Commun. (ICC 2014)*, Sydney, Australia, Jun. 2014, pp. 1385–1390.

[53] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, May 1967.

[54] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson/Prentice Hall Upper Saddle River, NJ, 2005.

[55] D. B. Percival and A. T. Walden, *Spectral analysis for physical applications*. Cambridge University Press, 1993.

[56] A. Mariani, A. Giorgetti, and M. Chiani, "Test of independence for cooperative spectrum sensing with uncalibrated receivers," in *Proc. IEEE Global Commun. Conf. (GLOBECOM 2012)*, Anaheim, CA, USA, Dec. 2012, pp. 1374–1379.

[57] M. Chiani, "Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy-Widom distribution," *Journal of Multivariate Analysis*, vol. 129, pp. 69–81, Aug. 2014.

[58] S. Nadarajah, "Explicit expressions for moments of order statistics," *Statistics & Probability Letters*, vol. 78, no. 2, pp. 196–205, Feb. 2008.

[59] S. S. Gupta, "Order statistics from the gamma distribution," *Technometrics*, vol. 2, no. 2, pp. 243–262, May 1960.

[60] H. A. David and H. N. Nagaraja, *Order statistics*. Wiley Online Library, 2004.

[61] F. N. David and N. L. Johnson, "Statistical treatment of censored data part I. Fundamental formulae," *Biometrika*, vol. 41, no. 1/2, pp. 228–240, Jun. 1954.