

A Regularized Maximum Likelihood Estimator for the Period of a Cyclostationary Process

David Ramírez and Peter J. Schreier
Signal and System Theory Group
University of Paderborn, Germany
{david.ramirez,peter.schreier}@sst.upb.de

Javier Vía and Ignacio Santamaría
Advanced Signal Processing Group
University of Cantabria, Spain
{jvia,nacho}@gtas.dicom.unican.es

Louis L. Scharf
Department of Mathematics
Colorado State University, United States
Louis.Scharf@ColoState.edu

Abstract—We derive an estimator of the cycle period of a univariate cyclostationary process based on an information-theoretic criterion. Transforming the univariate cyclostationary process into a vector-valued wide-sense stationary process allows us to obtain the structure of the covariance matrix, which is block-Toeplitz, and its block size depends on the unknown cycle period. Therefore, we sweep the block size and obtain the ML estimate of the covariance matrix, required for the information-theoretic criterion. Since there are no closed-form ML estimates of block-Toeplitz matrices, we asymptotically approximate them as block-circulant. Finally, some numerical examples show the good performance of the proposed estimator.

I. INTRODUCTION

Cyclostationary (CS) processes [1], [2] are commonly encountered in science and engineering since they can be used to model periodic phenomena in climatology, meteorology, astronomy, economics, medicine, mechanics and communications, among others. A process is said to be CS if its statistical properties vary periodically. Since we will be considering zero-mean processes and second-order cyclostationarity, only the covariance function is assumed to be periodic in its global time variable.

Most techniques that exploit cyclostationarity assume that the cycle period is known a priori. For instance, this is a typical assumption for detectors of cyclostationarity [3]–[6], beamforming [7] and many other applications (see [2] and references therein). In practice, however, the cycle period must usually be estimated from the measurements.

Some estimators of the cycle period have been proposed in the past. The authors in [8] obtained the estimate from the Loève spectrum [9]. The estimator is based on a result in [10], which states that the support of a discrete CS process in the dual-frequency domain is composed by a finite number of parallel lines, which are separated by the cycle frequency. Hence, this estimator obtains the cycle period by estimating this separation. The authors in [11] proposed an estimator based on a measure, named variability. Basically, this technique obtains the value of the variability for different cycle periods and selects the period that maximizes it. However, the variability-based estimator does not work with the so-called *ill* cyclostationary signals, which includes many practically relevant signals, such as a communications signal with rectangular shaping. Finally, another estimator based on the CS detector in [4] was proposed in [11]. This estimator computes the detector statistic of [4] for different cycle periods and picks the best one. This principle of searching a detector statistic over candidate

values of the cycle period may also be used with other detector statistics such as that of [5], [6].

All these techniques provide ad-hoc CS period estimators suited for specific scenarios, but they are not based on sounded estimation approaches such as maximum likelihood (ML), maximum a posteriori probability or minimum mean square error, etc. To fill this gap, in this work, we obtain the ML estimator for the cycle period of a CS process. The main tool we use is the relationship between the scalar-valued CS process and a vector-valued wide-sense stationary (WSS) process, obtained by vectorizing the CS time-series with a vector length given by the cycle period [10]. Since the cycle period is unknown, we must sweep the block size (vector length) to find the optimal value. Asymptotically, this seemingly complex search is actually a simple search in the frequency domain.

Assuming Gaussianity and for a given block-size, we must obtain the ML estimate of the covariance matrix, which is block-Toeplitz. Since there are no closed-form solutions for the ML estimate of Toeplitz matrices [12], we asymptotically approximate it as block-circulant. It was shown in [13] that this likelihood converges in mean square to the likelihood with the block-Toeplitz matrix. With the ML estimate of the block-circulant matrix for a fixed block size, we must sweep the block size to find the ML estimate of the cycle period. However, as the value of the investigated cycle period increases, the number of parameters to estimate also increases, thus yielding a larger likelihood. It is therefore necessary to add a penalty term. Following [14], we propose to use an information theoretic criterion to penalize large values of the cycle period. Specifically, we used the minimum description length criterion [15], which provides us a better estimate than the ML approach, and, as we will see, this estimator outperforms previously proposed estimators.

II. PROBLEM FORMULATION

We would like to estimate the cycle period, P , of the cyclostationary (CS) time series $u[n]$. We assume that $u[n]$ is a scalar-valued and zero-mean process, and the cyclostationarity is exhibited in the covariance function, i.e.,

$$r_{uu}[n, m] = E[u[n]u^*[n - m]] = r_{uu}[n + P, m],$$

where the period P is an unknown natural number. Moreover, we also assume that $P \geq 2$, since the case $P = 1$ corresponds to wide-sense stationarity.

To derive an estimator based on an information-theoretic criterion we first need the ML estimator. In this section, following our previous work in [3], we formulate the problem in a way that makes this possible. We assume that $u[n]$ is proper complex Gaussian with zero mean.

We first assume that the cycle period is known. Thus, let us arrange $u[n]$ in blocks of size P , yielding the *vector-valued* time series

$$\mathbf{x}[n] = [u[nP] \ \cdots \ u[(n+1)P-1]]^T \in \mathbb{C}^P,$$

which is WSS [10]. We have therefore transformed a scalar-valued CS process into a vector-valued WSS process. We can now stack N realizations of $\mathbf{x}[n]$ into the vector

$$\mathbf{y} = [\mathbf{x}^T[0] \ \mathbf{x}^T[1] \ \cdots \ \mathbf{x}^T[N-1]]^T \in \mathbb{C}^{NP},$$

which is a stack of NP samples of $u[n]$, i.e.,

$$\mathbf{y} = [u[0] \ u[1] \ \cdots \ u[NP-1]]^T.$$

The former expression simplifies the derivation of the covariance matrix of \mathbf{y} , which is [3], [13]

$$\mathbf{R} = \begin{bmatrix} \mathbf{Q}[0] & \mathbf{Q}[-1] & \cdots & \mathbf{Q}[-N+1] \\ \mathbf{Q}[1] & \mathbf{Q}[0] & \cdots & \mathbf{Q}[-N+2] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}[N-1] & \mathbf{Q}[N-2] & \cdots & \mathbf{Q}[0] \end{bmatrix},$$

where $\mathbf{Q}[m] = E[\mathbf{x}[n]\mathbf{x}^H[n-m]] \in \mathbb{C}^{P \times P}$ is the matrix-valued WSS covariance function of $\mathbf{x}[n]$. The covariance matrix $\mathbf{R} \in \mathbb{C}^{NP \times NP}$ is a block-Toeplitz covariance matrix with block-size P . The observations are distributed as

$$p(\mathbf{y}; \mathbf{R}) = \frac{1}{\pi^{NP} \det(\mathbf{R})} \exp \{ -\mathbf{y}^H \mathbf{R}^{-1} \mathbf{y} \}.$$

III. DERIVATION OF THE ESTIMATOR

In this section we derive an estimator for the cycle period P based on an information theoretic criterion. First, we obtain the ML estimator and show that it cannot be directly applied since a larger P implies more degrees of freedom, and, therefore, a larger likelihood. Hence, we penalize the likelihood following an information theoretic approach, similar to that in [14].

A. Maximum Likelihood Estimator

To obtain the ML estimator we proceed as follows: For a fixed P we obtain the ML estimate of the covariance matrix \mathbf{R} . Then, we sweep P and select the value that maximizes the (compressed) likelihood. Assuming M independent and identically distributed realizations of \mathbf{y} , $\{\mathbf{y}_m\}_{m=0}^{M-1}$, the log-likelihood of the observations is given by

$$\log p(\mathbf{y}_0, \dots, \mathbf{y}_{M-1}; \mathbf{R}) = -NPM \log \pi - M \log \det(\mathbf{R}) - M \text{tr}(\mathbf{R}^{-1} \hat{\mathbf{R}}), \quad (1)$$

where the sample covariance matrix is

$$\hat{\mathbf{R}} = \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{y}_m \mathbf{y}_m^H.$$

We first obtain the ML estimate of \mathbf{R} for a fixed P . That is, we need to obtain the ML estimator of a block-Toeplitz

covariance matrix with block size P . However, it was shown in [12] that there is no closed-form solution for the ML estimate of Toeplitz matrices. We must therefore follow a different approach, which is based on the following theorem.

Theorem 1: As the number of samples N tends to infinity, the log-likelihood in (1) converges in the mean square sense to a log-likelihood with a particular block-circulant covariance matrix, i.e.,

$$\lim_{N \rightarrow \infty} E \left[\left[\frac{1}{N} \log p(\mathbf{y}_0, \dots, \mathbf{y}_{M-1}; \mathbf{R}) - \log p(\mathbf{y}_0, \dots, \mathbf{y}_{M-1}; \mathbf{C}) \right]^2 \right] = 0,$$

where $\mathbf{C} \in \mathbb{C}^{NP \times NP}$ is the block-circulant covariance matrix obtained as

$$\mathbf{C} = (\mathbf{F}_N \otimes \mathbf{I}_P) \mathbf{V} (\mathbf{F}_N \otimes \mathbf{I}_P)^H. \quad (2)$$

Here, \mathbf{F}_N is the Fourier matrix with elements given by $[\mathbf{F}_N]_{lm} = e^{-j2\pi lm/N} / \sqrt{N}$, and \mathbf{V} is a block-diagonal matrix, whose k th block is given by the discrete Fourier transform (DFT) of the covariance sequence

$$\mathbf{V}_k = \sum_{n=0}^{N-1} \mathbf{Q}[n] e^{-j\theta_k n} \in \mathbb{C}^{P \times P}, \quad (3)$$

where $\theta_k = 2\pi k/N$, with $k = 0, \dots, N-1$.

Proof: The proof is very similar to that in [13]. ■

Taking Theorem 1 into account, we may transform the observations as $\mathbf{z} = (\mathbf{F}_N \otimes \mathbf{I}_P)^H \mathbf{y}$, which is simply the stack of N frequencies of the discrete Fourier transform of $\mathbf{x}[n]$. Thus, the random vector \mathbf{z} is asymptotically distributed as a proper complex Gaussian with zero mean and covariance matrix \mathbf{V} . The covariance matrix is block-diagonal, and it is therefore relatively easy to estimate. In the following lemma, we present the ML estimator.

Lemma 1: The ML estimate of \mathbf{V} is given by

$$\hat{\mathbf{V}} = \text{diag}_P(\hat{\mathbf{S}}),$$

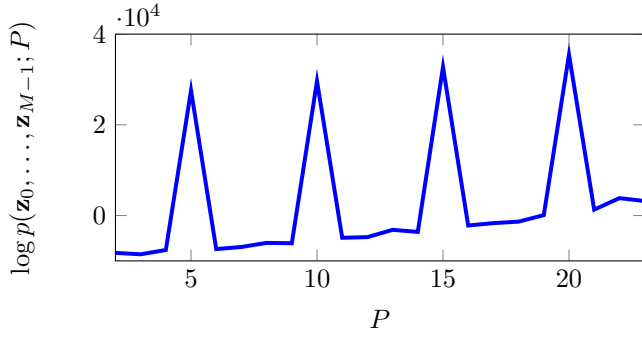
where $\text{diag}_P(\hat{\mathbf{S}})$ is the operator that builds a block-diagonal matrix with block sizes P from the blocks on the diagonal of $\hat{\mathbf{S}}$, which is the sample covariance matrix of the observations \mathbf{z}_m , i.e.,

$$\hat{\mathbf{S}} = \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{z}_m \mathbf{z}_m^H = (\mathbf{F}_N \otimes \mathbf{I}_P)^H \hat{\mathbf{R}} (\mathbf{F}_N \otimes \mathbf{I}_P).$$

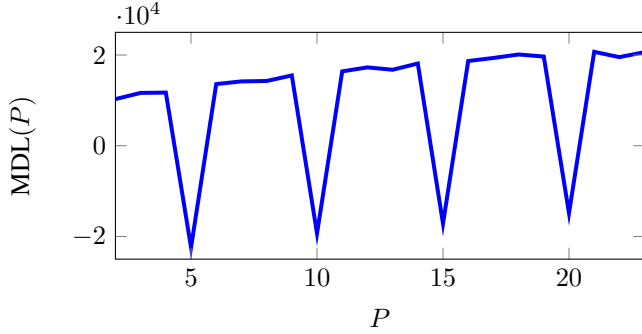
Proof: See [16]. ■

Based on the previous lemma, the *asymptotic* compressed likelihood is

$$\log p(\mathbf{z}_0, \dots, \mathbf{z}_{M-1}; P) = -NPM(\log \pi + 1) - M \sum_{k=0}^{N-1} \log \det(\hat{\mathbf{S}}_k), \quad (4)$$



(a) Compressed likelihood for different values of P



(b) MDL criterion for different values of P

Fig. 1: Compressed likelihood and MDL

where $\hat{\mathbf{S}}_k$ denotes the k th $P \times P$ block on the diagonal of $\hat{\mathbf{S}}$. Moreover, the (asymptotic) ML estimator of P is given by the value of P that maximizes (4). That is,

$$\hat{P} = \arg \max_{P=1, \dots, P_{\max}} \log p(\mathbf{z}_0, \dots, \mathbf{z}_{M-1}; P), \quad (5)$$

where P_{\max} is the maximum period that is checked.

We now show why the ML estimator does not directly solve this problem. Consider an example with $P = 5$, $N = 128$ and $M = 25$. As we may see in Figure 1a, there are clear peaks at the cycle period and its multiples (5, 10, 15, ...). However, there is also increasing trend with P , which precludes the use of the maximum as the estimator of the period. This trend may be explained as follows: Keeping the size of the covariance matrix fixed, which we denote by T , a larger P leaves a smaller N , since $T = NP$. Therefore, the covariance matrix has more structure as P increases. As we will see later, the degrees of freedom vary between T and T^2 , for $P = 1, N = T$ and $N = 1, P = T$, respectively. Here, $N = 1$ is the extreme case, where the covariance matrix is no longer a block-Toeplitz (or block-circulant) matrix, but just a positive definite matrix. Hence, a penalty term is necessary.

B. Minimum Description Length

We propose to use the minimum description length (MDL) criterion [15] to overcome the overfitting problem. Other criteria, such as the Akaike information criterion [17], could also be applied.

The MDL criterion for our problem is given by

$$\text{MDL}(P) = -\log p(\mathbf{z}_0, \dots, \mathbf{z}_{M-1}; P) + \frac{1}{2}\nu \log M, \quad (6)$$

where ν is the number of degrees of freedom of the model, which depends on P . We now count the number of degrees of freedom. One may be tempted to count the number of degrees of freedom of a block-Toeplitz covariance matrix. However, we are actually estimating the unknown parameters of a block-circulant matrix, that is, the elements in the block-diagonal matrix \mathbf{V} . For a given P , the matrix \mathbf{V} is composed by N blocks of dimension $P \times P$. Moreover, since \mathbf{V} is Hermitian without further structure, each block is also Hermitian without further structure. Thus, the number of degrees of freedom of each block is

$$P + 2 \frac{P(P-1)}{2} = P^2,$$

where the first P comes from the P real elements in the main diagonal of the k th block, and the second term is two times the number of complex elements above the main diagonal. Thus, since there are N blocks, the total number of degrees of freedom is $\nu = NP^2$. Plugging ν and (4) into (6), the MDL criterion becomes

$$\begin{aligned} \text{MDL}(P) = & NPM(\log \pi + 1) \\ & + M \sum_{k=0}^{N-1} \log \det(\hat{\mathbf{S}}_k) + \frac{1}{2}NP^2 \log M, \end{aligned} \quad (7)$$

and the MDL-based estimator is

$$\hat{P} = \arg \min_{P=1, \dots, P_{\max}} \text{MDL}(P). \quad (8)$$

Figure 1b shows the MDL criterion for the experiment presented in the previous subsection. In this figure, we can observe that the period that minimizes the criterion is $\hat{P} = 5$.

IV. NUMERICAL SIMULATIONS

In this section we evaluate the performance of the proposed estimator, and compare it with that of previous detectors. Specifically, we consider the following model for the CS process

$$u[n] = (h * s)[n] + w[n],$$

where $h[n]$ is a Rayleigh channel with 10 taps (at the symbol rate) and a flat power delay profile, $w[n]$ is an additive Gaussian noise, generated by a moving average model of order 19 (at the sampling frequency). The transmitted signal $s[n]$ is QPSK with rectangular shaping. Moreover, the symbol rate is 600 Ksymbols/second and the sampling frequency is 2.4 MHz, which yields a cycle period $P = 4$. Finally, $M = 10$ realizations and $N = 64$, which results in $4 \cdot 10 \cdot 64 = 2560$ samples available at the estimator.

We have chosen the following cycle period estimators for comparison. The first one, presented in [8], is based on the Loève spectrum [9]. Basically, it uses a smoothed periodogram to estimate the Loève spectrum, which is normalized by the estimated power spectral density. Then, the magnitude squared of the normalized spectrum, or coherence, is integrated along lines parallel to the stationary manifold, and the fundamental cycle frequency is the one that maximizes this measure. In particular, we estimate the smoothed periodogram with a length 2048 FFT and a smoothing window of length 256. The second competitor is based on the detector of cyclostationarity presented in [4]. Basically, it obtains the statistic for different

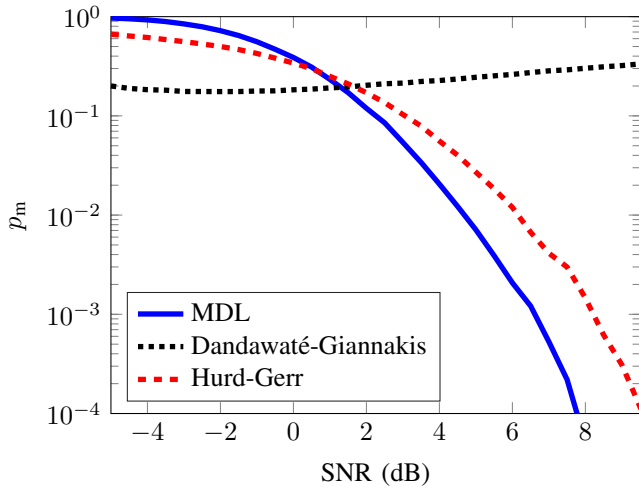


Fig. 2: Probability of missed detection vs. SNR for an experiment with $P = 4$, $M = 10$ snapshots and $N = 64$

cycle periods, and selects the period with the largest value of the statistic. In our implementation, we use the first four lags of the cyclic covariance to build the detector in [4], and a Kaiser window of length 257 to estimate the cyclic power spectral density.

Figure 2 shows the probability of missed detection vs. the signal-to-noise ratio (SNR) for the proposed estimator and the two competitors. As can be seen, the proposed estimator yields the best results for moderate and high SNRs. It is also worth noticing that the estimator based on the detector in [4] does not improve with the SNR. To shed some light on this observation we plot 1000 realizations of the statistic used in [4] against P for $\text{SNR} = 10$ dB, and its mean (as the thick line). As can be seen the value corresponding to $P = 2$ also presents high values. This phenomenon is due to the fact that $P = 2$ captures the cyclostationarity exhibited in the second harmonic of the fundamental cycle frequency. For some realizations the statistic corresponding to $P = 2$ may be larger than for $P = 4$, and that explains why the performance of the estimator does not improve with the SNR.

Two final comments are in order. First, the approach based on the detector [4] provides very good performance for cycle periods that are prime numbers, since the aforementioned effect does not occur for such values. Second, the signal in our example is an “ill-CS” signal¹ according to the definition given in [11]. This prevents a comparison with the variability technique [11].

V. CONCLUSIONS

We have obtained the ML estimator for the period of a CS process. The first ingredient to obtain the ML estimator is the relationship between the scalar-valued CS process and a vector-valued wide-sense stationary (WSS) process, obtained by arranging the CS process in blocks of size given by the cycle period. The resulting WSS process has a block-Toeplitz

¹A ill-CS signal is a CS signal with all the diagonal elements of $\mathbf{Q}[m]$ equal.

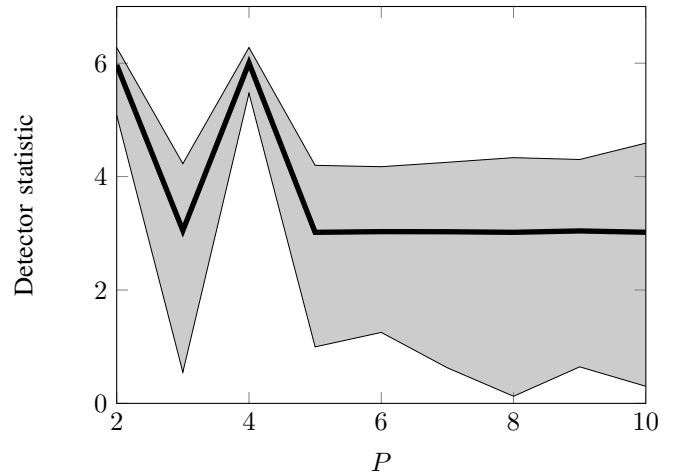


Fig. 3: Dandawaté-Giannakis statistic for different values of P

covariance matrix, for which there is no closed-form ML estimate. We thus asymptotically approximate it by a block-circulant matrix, for which we may obtain a closed-form estimate. The ML estimate of the covariance matrix assumes that the cycle period is known. Since this is not true, we must therefore sweep its value and select the value that maximizes the compressed likelihood as the ML estimator. However, we showed that the ML estimator is not a valid estimator, since the larger the cycle period is, the more degrees of freedom there are, and the likelihood obviously becomes larger. To solve this issue, we penalized large values of the cycle period using the MDL criterion. Finally, simulations showed the good performance of the proposed estimator.

ACKNOWLEDGMENT

The work of P. Schreier was supported by the Alfried Krupp von Bohlen und Halbach Foundation, under its program “Return of German scientists from abroad”. The work of I. Santamaría and J. Vía was supported by the Spanish Government, Ministerio de Ciencia e Innovación (MICINN), under project RACHEL (TEC2013-47141-C4-3-R). The work of L. Scharf was supported by the Airforce Office of Scientific Research under contract FA9550-10-1-0241.

REFERENCES

- [1] W. A. Gardner, A. Napolitano, and L. Paura, “Cyclostationarity: Half a century of research,” *Signal Process.*, vol. 86, no. 4, pp. 639–697, 2006.
- [2] E. Serpedin, F. Panduru, I. Sar, and G. B. Giannakis, “Bibliography on cyclostationarity,” *Signal Process.*, vol. 85, no. 12, pp. 2233–2303, 2005.
- [3] D. Ramírez, L. L. Scharf, J. Vía, I. Santamaría, and P. J. Schreier, “An asymptotic GLRT for the detection of cyclostationary signals,” in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, Florence, Italy, May 2014.
- [4] A. V. Dandawaté and G. B. Giannakis, “Statistical tests for presence of cyclostationarity,” *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2355–2369, Sep. 1994.
- [5] J. Riba, J. Font-Segura, J. Villares, and G. Vazquez, “Frequency-domain GLR detection of a second-order cyclostationary signal over fading channels,” *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 1899–1912, Apr. 2014.

- [6] S. Enserink and D. Cochran, "On detection of cyclostationary signals," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Detroit, USA, May 1995, pp. 2004–2007.
- [7] L. Castedo and A. R. Figueiras-Vidal, "An adaptive beamforming technique based on cyclostationary signal properties," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1637–1650, Jul. 1995.
- [8] H. L. Hurd and N. L. Gerr, "Graphical methods for determining the presence of periodic correlation," *Journal of Time Series Analysis*, vol. 12, no. 4, pp. 337–350, 1991.
- [9] M. Loève, *Probability Theory II*, 4th ed. New York: Springer, 1978.
- [10] E. D. Gladyshev, "Periodically correlated random sequences," *Soviet Math. Dokl.*, vol. 2, pp. 385–388, 1961.
- [11] J. Wang, T. Chen, and B. Huang, "Cyclo-period estimation for discrete-time cyclo-stationary signals," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 83–94, Jan 2006.
- [12] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison - Wesley, 1991.
- [13] D. Ramírez, G. Vázquez-Vilar, R. López-Valcarce, J. Vía, and I. Santamaría, "Detection of rank- P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, Aug. 2011.
- [14] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. on Acoustics, Speech and Signal Process.*, vol. ASSP-33, no. 2, p. 387, Apr. 1985.
- [15] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, pp. 461–464, 1978.
- [16] D. Ramírez, J. Vía, I. Santamaría, and L. L. Scharf, "Detection of spatially correlated Gaussian time series," *IEEE Trans. Signal Process.*, vol. 58, no. 10, Oct. 2010.
- [17] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *2nd Int. Symp. Inf. Theory*, 1973, pp. 267–281.