Bayesian Multiantenna Sensing for Cognitive Radio

J. Manco-Vásquez, M. Lazaro-Gredilla†, D. Ramírez*, J. Vía and I. Santamaria
Dept. of Communications Engineering, University of Cantabria, Spain
†Dept. Signal Theory and Commun., Univ. Carlos III de Madrid, Spain
*Dept. of Electrical Engineering and Inf. Technology, University of Paderborn, Germany

***Abstract*—In this paper, the problem of multiantenna spectrum sensing in cognitive radio (CR) is addressed within a Bayesian framework. Unlike previous works, our Bayesian model places priors directly on the spatial covariance matrices under both hypotheses, as well as on the probability of channel occupancy. Specifically, we use inverse-gamma and complex inverse-Wishart distributions as conjugate priors for the null and alternative hypotheses, respectively; and a Bernoulli distribution as the prior for channel occupancy. At each sensing period, Bayesian inference is applied and the posterior of channel occupancy is thresholded for detection. After a suitable approximation, the posteriors are employed as priors for the next sensing frame, which can be beneficial in slowly time-varying environments. By means of simulations, the proposed detector is shown to outperform the Generalized Likelihood Ratio Test (GLRT) detector.**

***Index Terms*—Cognitive Radio, Multiantenna Spectrum Sensing, Bayesian Inference, Generalized Likelihood Ratio Test (GLRT), Bayesian Forgetting.**

I. Introduction

Spectrum sensing is one of the critical operations that cognitive radio (CR) secondary users must perform to identify whether a wireless communication channel is in use by a licensed primary user (PU) or not [1]. Recently, detectors employing multiple antennas have received increased attention because they do not require prior knowledge about the PU signalling scheme or the noise powers and are able to work with asynchronously sampled signals [2]. These multiantenna detectors exploit the fact that under the null hypothesis (empty channel) the signals received at the different antennas are spatially uncorrelated, whereas the presence of a PU induces some correlation and/or additional structure in the spatial covariance matrix. Since the binary hypothesis testing problem involves some unknown parameters (e.g., noise variance and channel state information), the generalized likelihood ratio test (GLRT) approach has been typically followed to find detectors in several scenarios [2], [3], [4].

As an alternative to the GLRT approach, Bayesian detectors for cognitive radios exploiting prior information in the receiver about the noise and channel statistics have been recently studied in [5], [6], [7]. In particular, [5] considers single antenna detectors and assumes a Gaussian prior for the fading channel, an inverse chi-squared prior for the noise variance and a uniform prior for the transmitted symbols. With this prior knowledge, an iterative detector that uses the Maximum A Posteriori (MAP) estimates for the unknown parameters is proposed. In [6], a multiple-input multiple-output (MIMO) cognitive channel is considered and the density under the alternative hypothesis is marginalized by integrating over the probability space of random matrices with Gaussian i.i.d. entries.

In this paper we also follow a Bayesian approach to derive a multiantenna CR detector that, unlike previously proposed detectors, exploits prior information obtained from past sensing periods. A motivation for this approach is that the time scale of variation of the statistical parameters involved in the detection problem (for instance, noise variance or space-time PU activity patterns) can be long compared to the sensing period. For instance, channel access patterns for primary users have been characterized as slowly time-varying in [8], which allows us to predict the spectrum usage. However, one-shot detectors like those considered in this paper have to quickly identify the spectrum occupancy using to this end as few samples as possible. It is therefore reasonable to think that the statistical information obtained from past decisions can be a suitable prior for the current sensing period.

Unlike [5], [6], [7], our Bayesian model places priors directly on the spatial covariance matrices under both hypotheses, as well as on the probability of channel occupancy. Specifically, we use inverse-gamma and complex inverse-Wishart distributions as conjugate priors for the null ($\mathcal{H}_0$) and alternative hypothesis ($\mathcal{H}_1$), respectively; and a Bernoulli distribution as the prior for channel occupancy. At each sensing period, exact Bayesian inference is applied and the posterior for the channel occupancy given the observations is thresholded for detection. Since the posteriors summarize the information gathered so far about the actual CR scenario, they are used as priors for the next sensing period. We show in the paper how this simple Bayesian procedure can be modified to introduce a forgetting mechanism to cope with non-stationary environments.

The rest of the paper is organized as follows. The CR detection problem is formulated in Section II. The Bayesian model, the inference procedure and the proposed scheme to handle time-varying scenarios are presented in Section III. The simulation results will be presented in Section IV and finally the concluding remarks are given in Section V.

II. Multiantenna sensing for cognitive radio

We consider a multiantenna secondary receiver, which, during the t-th sensing period or frame, is able to sense the medium N times through L independent antennas, thus

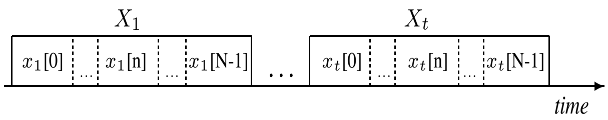

Fig. 1. Sensing process: At each sensing frame, N i.i.d. L-dimensional observations are collected. The channel is constant within a frame and varies smoothly between frames.

collecting $L \times N$ observations in an observation matrix $\mathbf{X}_t = [\mathbf{x}_t[1], \ldots, \mathbf{x}_t[N]]$. We use $\mathbf{x}_t[n] \in \mathbb{C}^L$ to denote the L observations at the n-th time instant within the t-th sensing period or frame. The channel may vary from frame to frame, but it is assumed constant within a frame, similarly to the block fading model typically considered in packet-based wireless communication systems. The sensing strategy is depicted in Figure 1.

When a PU is active, the signal received by the cognitive user is given by

$$\mathbf{x}_t[n] = \mathbf{H}_t \mathbf{s}_t[n] + \mathbf{v}_t[n], \tag{1}$$

where $\mathbf{s}_t[n] \in \mathbb{C}^P$ is a vector composed of P transmitted sources (this models either a single P-antenna transmitter or P single-antenna transmitters), $\mathbf{H}_t \in \mathbb{C}^{L\times P}$ describes the channel between the transmitter and the receiver, and $\mathbf{v}_t[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{D}_t)$ models the additive Gaussian noise with zero mean and diagonal covariance matrix $\mathbf{D}_t$. In this way we assume independent noises (with possible different unknown variances) at each antenna. The noise variance associated to each antenna and the channel remain constant within a single frame, but may vary between frames. We will assume that, whenever this variation exists, it is smooth. We finally assume that $\mathbf{s}_t[n]$ can be modeled as i.i.d. zero-mean circular complex normal with covariance matrix $\sigma_s^2\mathbf{I}$.

The detection problem at sensing period t amounts to deciding between two different structures for the covariance matrix of $\mathbf{x}_t[n]$:

$$\begin{aligned}
&\mathcal{H}_0 : \mathbf{x}_t[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{D}_t),\ n = 0, \ldots, N-1,\\
&\mathcal{H}_1 : \mathbf{x}_t[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_t),\ n = 0, \ldots, N-1;
\end{aligned}$$

where we assume that $\mathbf{R}_t$ is an arbitrary positive definite covariance matrix and $\mathbf{D}_t$ is the previously defined diagonal covariance matrix.

According to the previous analysis, observations within a single sensing frame can be regarded as i.i.d. draws (signals are considered temporally white) from a zero-mean complex normal with some covariance matrix, and the absence or presence of a transmitter is equivalent to testing whether this covariance matrix is diagonal or not, respectively.

III. A Bayesian detector

We first address the detection problem in the case of a single frame and introduce an algorithm to perform approximate Bayesian inference. Then, we extend this result to multiple frames (time-varying environments) by introducing a forgetting mechanism.

A. One-shot detector for a single frame

We first consider the detection problem for a single sensing frame and a stationary environment. Therefore, to simplify notation, we drop the frame indicator t in the following.

1) Likelihood of $\mathbf{R}$ *and* $\mathbf{D}$*:* We can write the likelihood of $\mathbf{R}$ and $\mathbf{D}$ given $\mathbf{X}$ according to the description of Section II:

$$p(\mathbf{X}|z=0,\mathbf{D}) = \prod_{n=1}^{N} \mathcal{CN}(\mathbf{x}[n]|z=0,\mathbf{D}) \tag{2a}$$

$$p(\mathbf{X}|z=1,\mathbf{R}) = \prod_{n=1}^{N} \mathcal{CN}(\mathbf{x}[n]|z=1,\mathbf{R}) \tag{2b}$$

where z is a binary hidden variable that indicates whether a transmitter is present ($z=1$) or not ($z=0$). All our information about $\mathbf{R}$ and $\mathbf{D}$ is that they are some unknown covariance matrices, respectively full Hermitian and diagonal.

2) Prior distribution of z*,* $\mathbf{R}$ *and* $\mathbf{D}$*:* Following a proper Bayesian treatment, we must place a prior distribution on all the unknown parameters of the model. We will use the following

$$p(z) = \text{Bernoulli}(z|\breve{\pi}) = \breve{\pi}^z(1-\breve{\pi})^{1-z} \tag{3a}$$

$$\begin{aligned}
p(\mathbf{R}) &= \mathcal{CW}^{-1}(\mathbf{R}|\breve{n}, \breve{\mathbf{R}})\\
&= \frac{|\breve{\mathbf{R}}|^{\frac{\breve{n}}{2}} |\mathbf{R}|^{-\frac{\breve{n}+L+1}{2}} \exp(-\frac{1}{2}\operatorname{trace}(\mathbf{R}^{-1}\breve{\mathbf{R}}))}{2^{\frac{\breve{n}L}{2}}\Gamma_L(\frac{\breve{n}}{2})}
\end{aligned} \tag{3b}$$

$$\begin{aligned}
p(\mathbf{D}) &= \mathcal{G}_L^{-1}(\mathbf{D}|\breve{m}, \breve{\mathbf{D}})\\
&= \prod_{l=1}^{L} \mathcal{G}^{-1}([\mathbf{D}]_{ll}|\breve{m}/2, [\breve{\mathbf{D}}]_{ll}/2)\\
&= \frac{|\breve{\mathbf{D}}|^{\frac{\breve{m}}{2}} |\mathbf{D}|^{-\frac{\breve{m}+L+1}{2}} \exp(-\frac{1}{2}\operatorname{trace}(\mathbf{D}^{-1}\breve{\mathbf{D}}))}{2^{\frac{\breve{m}L}{2}}\Gamma^L(\frac{\breve{m}}{2})}
\end{aligned} \tag{3c}$$

where we have included the definitions of the Bernoulli distribution, the complex inverse-Wishart ($\mathcal{CW}^{-1}$) and the product of L independent inverse-gamma ($\mathcal{G}_L^{-1}$). Note the difference between $\Gamma_L(\cdot)$ (used to denote the multivariate gamma function) and $\Gamma^L(\cdot)$ (the standard gamma function raised to the L-th power). We denote the parameters of the prior distributions as $\breve{\pi}, \breve{n}, \breve{\mathbf{R}}, \breve{m}$ and $\breve{\mathbf{D}}$.

The main argument for these choices is analytical tractability: The complex inverse-Wishart distribution placed on $\mathbf{R}$ and the product of univariate inverse-gamma distributions placed on $\mathbf{D}$ are the conjugate priors for the distribution of full-rank covariance matrices and a diagonal covariance matrices, respectively, when the observations follow a complex multivariate Gaussian. This is very convenient, since in this way Bayesian inference can be performed exactly without resorting to numerical integration methods.

3) Exact posterior distribution of z*,* $\mathbf{R}$ *and* $\mathbf{D}$*:* Given the hidden variable, z, priors are conjugate and therefore posterior distributions have the same form as the prior, but with different parameters. When z is marginalized, each posterior is a convex combination of the posteriors for each hypothesis, yielding

$$p(z|\mathbf{X}) = \text{Bernoulli}(z|\hat{\pi}) \tag{4a}$$

$$p(\mathbf{R}|\mathbf{X}) = \hat{\pi}\mathcal{CW}^{-1}(\mathbf{R}|\hat{n}, \hat{\mathbf{R}}) + (1-\hat{\pi})\mathcal{CW}^{-1}(\mathbf{R}|\breve{n}, \breve{\mathbf{R}}) \tag{4b}$$

$$p(\mathbf{D}|\mathbf{X}) = \hat{\pi}\mathcal{G}_L^{-1}(\mathbf{D}|\breve{m}, \breve{\mathbf{D}}) + (1-\hat{\pi})\mathcal{G}_L^{-1}(\mathbf{D}|\hat{m}, \hat{\mathbf{D}}). \tag{4c}$$

The posterior parameters are given by

$$\hat{n} = \breve{n} + N \tag{5a}$$

$$\hat{\mathbf{R}} = \breve{\mathbf{R}} + \mathbf{S} \tag{5b}$$

$$\hat{m} = \breve{m} + N \tag{5c}$$

$$\hat{\mathbf{D}} = \breve{\mathbf{D}} + \text{diag}(\mathbf{S}) \tag{5d}$$

$$\hat{\pi} = \frac{p(\mathbf{X}|z=1)p(z=1)}{p(\mathbf{X}|z=1)p(z=1) + p(\mathbf{X}|z=0)p(z=0)} \tag{5e}$$

where $\mathbf{S} = \mathbf{X}\mathbf{X}^H/N$ is the sample covariance matrix and the $\text{diag}(\cdot)$ operator sets all the elements of a matrix to zero except those in the main diagonal. Observe that we use a breve ($\breve{}$) to denote the parameters of the *prior* distribution, whereas we use a hat ($\hat{}$) to denote the parameters of the *posterior* distribution. Finally, the marginal likelihood $p(\mathbf{X}|z)$ can be obtained analytically as

$$p(\mathbf{X}|z=1) = \int p(\mathbf{X}|z=1,\mathbf{R})p(\mathbf{R})\text{d}\mathbf{R} = \frac{|\breve{\mathbf{R}}|^{\frac{\breve{n}}{2}}\Gamma_L(\frac{\hat{n}}{2})}{\pi^{\frac{NL}{2}}|\hat{\mathbf{R}}|^{\frac{\hat{n}}{2}}\Gamma_L(\frac{\breve{n}}{2})} \tag{6a}$$

$$p(\mathbf{X}|z=0) = \int p(\mathbf{X}|z=0,\mathbf{D})p(\mathbf{D})\text{d}\mathbf{D} = \frac{|\breve{\mathbf{D}}|^{\frac{\breve{m}}{2}}\Gamma(\frac{\hat{m}}{2})^L}{\pi^{\frac{NL}{2}}|\hat{\mathbf{D}}|^{\frac{\hat{m}}{2}}\Gamma(\frac{\breve{m}}{2})^L}. \tag{6b}$$

After the posterior has been computed, the probability of a transmitter being present given observations $\mathbf{X}$ is simply $p(z=1|\mathbf{X}) = \hat{\pi}$. Thus, we can occupy the channel when the collision probability $\hat{\pi}$ is below some threshold.

B. Multiple frames

The general idea to repeat the process in the next sensing period is that the posterior at frame t can be used as prior at frame $t+1$. However, after integrating out the hidden variable, the posteriors of $\mathbf{R}$ and $\mathbf{D}$ are linear combinations of densities given by (4b) and (4c). To keep the process simple and scalable, it would be convenient to find an approximation of the posteriors within the family of each respective prior. A simple approximation can be obtained by truncating $\hat{\pi}$ to either 0 or 1, whichever is closer. When this is done, Eq. (4b) and (4c) directly yield a posterior in the same family as the prior. In that case, when $\mathcal{H}_1$ is more probable, the posterior is obtained by performing only updates (5a) and (5b) , whereas in the opposite case, only updates (5c) and (5d) are needed. A more rigorous approach would be to find the approximation of the posterior within the family of the prior that minimizes the Kullback-Leibler distance. This alternative approach is omitted in this paper for brevity.

Using the truncated approximation, the posterior after processing frame t summarizes (approximately) all statistical information observed so far. Since the channel may vary smoothly between consecutive frames, it is interesting to introduce a mechanism within the Bayesian framework to forget past data and hence be able to operate in a non-stationary environment. Since a dynamical model of the evolution of $\mathbf{R}_t$ and $\mathbf{D}_t$ is not available, we resort to the idea of Bayesian λ-forgetting [9]. To that end, we define the prior distributions for frame $t+1$ as a "smoothed" version of the posterior distributions obtained after processing frame t and the original prior distributions for $\mathbf{R}$ and $\mathbf{D}$ given by (3), i.e.,

$$p(\mathbf{D}_{t+1}|\mathbf{X}_t) \propto p(\mathbf{D}_t|\mathbf{X}_t)^\lambda p(\mathbf{D})^{1-\lambda} \tag{7a}$$

$$p(\mathbf{R}_{t+1}|\mathbf{X}_t) \propto p(\mathbf{R}_t|\mathbf{X}_t)^\lambda p(\mathbf{R})^{1-\lambda}, \tag{7b}$$

Observe that according to this definition, when $\lambda = 0$, all the information obtained from previous data is forgotten and the process considers each frame independently (as the GLRT does), which is reasonable if abrupt changes occur in $\mathbf{R}_t$ and $\mathbf{D}_t$ between frames. When $\lambda = 1$, no forgetting occurs and the new posterior corresponds to the standard Bayesian posterior when $\mathbf{D}_t$ and $\mathbf{R}_t$ are constant across frames $\mathbf{D}_t = \mathbf{D}$, $\mathbf{R}_t = \mathbf{R}$ $\forall_t$, which is reasonable under stationary conditions. Values $\lambda \in [0,1]$ are therefore appropriate to model different evolution speeds in the channel, without having to define a concrete dynamical model. In another perspective, Eq. (7) represents a change of the posterior in the direction of the prior: this has also been named as "back-to-the-prior" forgetting in [10].

With this forgetting step, the parameters of the prior distributions to be used for Bayesian inference at $t+1$ are given by

$$\breve{n}_{t+1} = \lambda\hat{n}_t + (1-\lambda)\breve{n}_0 \tag{8a}$$

$$\breve{\mathbf{R}}_{t+1} = \lambda\hat{\mathbf{R}}_t + (1-\lambda)\breve{\mathbf{R}}_0 \tag{8b}$$

$$\breve{m}_{t+1} = \lambda\hat{m}_t + (1-\lambda)\breve{m}_0 \tag{8c}$$

$$\breve{\mathbf{D}}_{t+1} = \lambda\hat{\mathbf{D}}_t + (1-\lambda)\breve{\mathbf{D}}_0. \tag{8d}$$

The posterior for frame $t+1$ can then be computed using (5) and (6). The whole process is summarized in Algorithm 1. Since the algorithm only requires updating and storing $\hat{\mathbf{R}}_t$, $\hat{n}_t$, $\hat{\mathbf{D}}_t$, $\hat{m}_t$ from one frame to the next, it requires a fixed amount of memory and computation per sensing frame, which is $\mathcal{O}(L^2)$.

IV. Simulation results

In this section, the performance of the proposed Bayesian detector is compared against the GLRT by means of simulations. We have considered $L = 5$ antennas at the receiver, $N = 50$ observations for each sensing frame and a signal-to-noise ratio SNR $= -8\,\text{dB}$ and $-10\,\text{dB}$. The probability of detection, P_D, for a fixed false alarm rate $P_{FA} = 0.1$ is estimated after running 10^5 independent realizations. Fig.2

Algorithm 1 Bayesian Multiantenna Sensing

Parameters: λ, $\breve{\mathbf{R}}_0$, $\breve{n}_0$, $\breve{\mathbf{D}}_0$, $\breve{m}_0$
for Frame $t = 1, 2, \ldots$ **do**
  Sense the medium N times through L antennas to get $\mathbf{X}_t$
  Exact posterior: Compute $\hat{\pi}_t$, $\hat{\mathbf{R}}_t$, $\hat{n}_t$, $\hat{\mathbf{D}}_t$, $\hat{m}_t$ using (5) and (6)
  Output $\hat{\pi}_t$, probability of a transmitter being present during t
  Compute the approximated posterior by truncating $\hat{\pi}_t$
  Forget: Compute $\breve{\mathbf{R}}_t$, $\breve{n}_t$, $\breve{\mathbf{D}}_t$, $\breve{m}_t$ using Eqs. (8)
end for

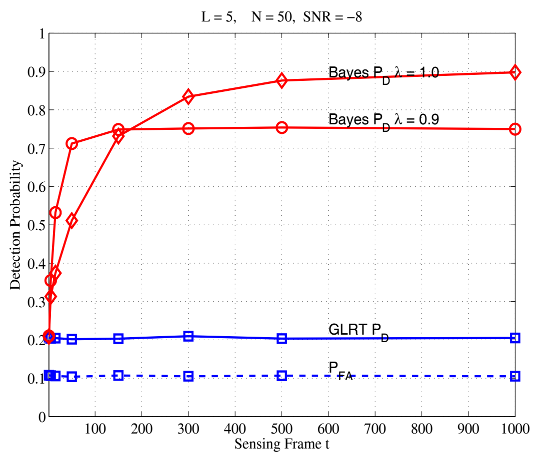

Fig. 2. P_D and P_{FA} for the Bayesian detector and the GLRT with a time-invariant channel. $L = 5$ antennas, $N = 50$ and SNR $= -8$ dB.

shows the results obtained by our Bayesian detector and the GLRT in a stationary environment in which the channel remains constant over all sensing frames. After sensing just a few frames the Bayesian detector provides much higher detection probability than the one-shot GLRT. We apply the Bayesian detector with $\lambda = 1$ and $\lambda = 0.9$, although we are in a stationary environment, the use of $\lambda = 0.9$ provides initially better results than $\lambda = 1$ due to the frequent detection errors that occur during the first sensing frames.

In the second example we consider a slowly time-varying environment in which the channel evolves from frame to frame as $\mathbf{H}_{t+1} = 0.9\mathbf{H}_t + \mathbf{P}_{t+1}$, with $\mathbf{P}_{t+1}$ a complex Gaussian noise matrix with i.i.d. zero-mean and scaled in order to have unit variance entries in $\mathbf{H}_{t+1}$. The results for different values of λ are depicted in Fig. 3.

V. Conclusion

We have derived a new Bayesian detector for multiantenna cognitive receivers that exploits statistical information obtained from past sensing frames. This information is summarized by the posterior densities on the covariance matrices under both hypotheses and on the probability of channel occupancy. Using suitable conjugate priors allowed us to perform exact Bayesian inference. We also extended this result to multiple frames (time-varying environments) by approximating the posterior within the family of the prior and introducing a forgetting mechanism. The proposed detector outperforms the one-shot GLRT detector in slowly time-varying environments. Future work will consider the use of Markov models to capture the space-time activity patterns of primary users.

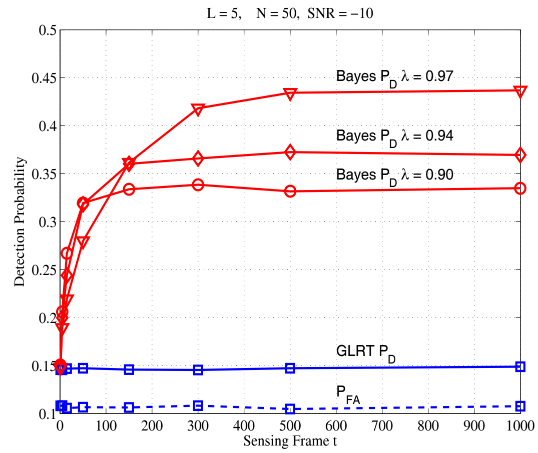

Fig. 3. P_D and P_{FA} for the Bayesian detector and the GLRT with a time-varying channel. $L = 5$ antennas, $N = 50$ and SNR $= -10$ dB.

Acknowledgment

The research leading to these results has received funding from the Spanish Government (MICINN) under projects TEC2010-19545-C04-03 (COSIMA) and CONSOLIDER-INGENIO 2010 CSD2008-00010 (COMONSENS).

References

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

[2] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1791–1800, May 2010.

[3] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of Rank-P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3764–3774, 2011.

[4] D. Ramirez, J. Via, I. Santamaria, and L. L. Scharf, "Detection of spatially correlated gaussian times series," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5006–5015, Oct. 2010.

[5] J. Font-Segura and X. Wang, "GLRT-based spectrum sensing for cognitive radio with prior information," *IEEE Transactions on Communications*, vol. 58, no. 7, pp. 2137–2146, 2010.

[6] R. Couillet and M. Debbah, "A Bayesian framework for collaborative multi-source signal sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5186–5195, Oct. 2010.

[7] E. Axel and E. G. Larsson, "A Bayesian approach to spectrum sensing, denoising and anomaly detection," in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc. 2009*, April 2009, pp. 2333–2336.

[8] T. Clancy and B. Walker, "Predictive dynamic spectrum access," in *Proc. SDR Forum Technical Conference*, Orlando, FL, USA, Nov 2006.

[9] R. Kulhavý and M. B. Zarrop, "On general concept of forgetting," *International Journal of Control*, vol. 58, no. 4, pp. 905–924, 1993.

[10] M. Lazaro-Gredilla, S. Van Vaerenbergh, and I. Santamaria, "A Bayesian approach to tracking with kernel recursive least-squares," in *IEEE Int. Workshop on Machine Learning for Signal Processing*, Beijing, China, Sept. 2011.