On distributed composite tests with spatially dependent observations in sensor networks

Juan Augusto Maya *Member, IEEE,* and Leonardo Rey Vega, *Member, IEEE.*

***Abstract*—We consider a distributed detection problem with statistically spatial dependent measurements at each sensor, for which there is no a central processing unit. Thus, each node takes some measurements, does some processing, exchanges messages with its neighbors and finally makes a decision (typically the same for all nodes) about the phenomenon of interest. A cooperative algorithm is proposed for reducing the number of communications between sensors and thus make an efficient use of the energy budget of a wireless sensor network (WSN). The problem is formulated as a composite hypothesis test using a general probability density function with unknown parameters leading naturally to the use of the Generalized Likelihood Ratio (GLR) test. As the sensors observe statistically spatial dependent samples, which makes difficult the implementation of fully distributed detection procedures, we propose a simpler algorithm for making a decision about the true hypothesis. We also compute its asymptotic distribution to characterize its performance. Interestingly, despite the fact that our proposal is more simple and efficient to implement than the GLR test, we find relevant scenarios for which it outperforms the latter, even in finite length regimes. We also test the proposed approach in a real-world application as spectrum sensing for cognitive radios.**

I. Introduction

In the near past, Wireless Sensor Networks (WSN) have received considerable attention from the research and industrial community because of their remote monitoring and control capabilities [1]–[3]. More recently, they have become an essential part of the emerging technology of Internet of Things (IoT) [4]–[6]. Among the different tasks to be done by WSNs, distributed detection is an active research topic [7]–[9].

In a distributed detection problem, geographically distributed sensors collect measurements from the phenomenon of interest, make some processing, exchange information with their neighbors and, finally, execute some consensus or diffusion algorithm to achieve a common final decision. This option is robust against node failures, and the communications between nodes are done locally, over typically short distances, saving energy and also bandwidth, by employing spatial reuse of the frequency bands.

Many works have considered fully distributed detection architectures [10]–[12]. Nevertheless, most of the works found in the literature assumes that the spatial measurements are independent [13]–[16]. However, in many applications of interest, the measurements taken by spatially distributed nodes are statistically dependent [17]. For example, this is the case of WSNs sensing physical variables like temperature and humidity [18], acoustic signals [19], [20] or even devices transmitting communication signals in the context of spectrum sensing for cognitive radio [21].

In this work we deal with a composite hypothesis testing problem where sensors take spatially dependent observations under each hypothesis, described by a general probability density function (pdf). We consider the approach of the generalized likelihood ratio (GLR) test, frequently used in these cases where some parameters of pdf of the observations are unknown. In the classical GLR statistic, the unknown parameters are estimated using the maximum likelihood estimator (MLE). Although it is possible to implement a MLE in a decentralized scenario, it generally requires a high number of messages exchanges between the network nodes, resulting in a elevated energy, bandwidth or delay costs to achieve a decision about the nature of the data. To alleviate this cost, we propose a statistic cooperatively built with the product of the marginal pdfs at each node that uses only local measurements. This is relevant, not only for the simplicity of the distributed computation over the network of the test statistic, but also for the estimation of the unknown parameters of the probability distributions. We also observe that this statistic has an attractive factorization structure that facilitates its computation in a distributed scenario, resulting in low implementation costs in terms of network resources.

In order to characterize its performance and evaluate possible penalties introduced by the proposed strategy, we derive the theoretical asymptotic distribution of the statistic which allows us to compute the error probabilities (type I or II) of the test. Although these results are strictly valid in the asymptotic scenario, we show with numerical examples that they offer good results also in the finite length regime. Surprisingly at first glance, the proposed statistic performs better than the GLR in several scenarios of practical interest. The theoretical results allow us to explain the main reasons for that behavior.

The paper is organized as follows. In the next section we introduce a general model to deal with. In Sec. III we present the proposed fully distributed statistic and in Sec IV, we derive its asymptotic performance. In Sec. V, we analyze a simple example to gain insight into the results, and in Sec VI, we evaluate the algorithm in the context of spectrum sensing for cognitive radio networks. We finally draw the main conclusion in Sec. VII. We let the proof of the asymptotic results and some other technical details to the appendices.

The two authors are with Universidad de Buenos Aires, Buenos Aires, Argentina, and CSC-CONICET, Buenos Aires, Argentina. Emails: {jmaya, lrey}@fi.uba.ar. This work was partially supported by the projects UBACyT 20020170200283BA and PICT-2017-4533.

Notation: Vectors and matrices are written in bold letters. $\mathbf{I}_P$ is the identity matrix of size $P \times P$. The gradient of a scalar function p with respect to (wrt) the vector $\boldsymbol{x} \in \mathbb{R}^{M_1}$ is noted as $\frac{\partial p}{\partial \boldsymbol{x}}$ and assumed to be a column vector. If $\boldsymbol{y} \in \mathbb{R}^{M_2}$, $\frac{\partial^2 p}{\partial \boldsymbol{x} \partial \boldsymbol{y}}$ is a $M_1 \times M_2$ matrix with its (i, j)-th component being $\frac{\partial^2 p}{\partial x_i \partial y_j}$, $i \in [1 : M_1]$, and $j \in [1 : M_2]$. $\mathbb{E}_{\boldsymbol{\phi}}(\cdot)$ denotes the expectation wrt the pdf $p(\cdot, \boldsymbol{\phi})$ with parameter $\boldsymbol{\phi}$. A Gaussian vector $\boldsymbol{n}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is notated $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. diag$(\mathbf{A})$ and diag$(\boldsymbol{a})$ are diagonal matrices built with the diagonal elements of the square matrix $\mathbf{A}$, or the elements of the vector $\boldsymbol{a}$, respectively. $\overline{\text{diag}}(\mathbf{A})$ is a vector built with the diagonal elements of $\mathbf{A}$.

II. Model

We consider a composite binary hypothesis testing problem in a WSN with N nodes. Assume that each sensor takes L observations, which are statistically independent and identically distributed (iid) in time, but possible dependent across the sensors. Let $\boldsymbol{z}_l \equiv [z_1(l), \ldots, z_N(l)]^T \in \mathbb{R}^N$ be the observations taken by all nodes at the l-th slot time, $l \in [1 : L]$, and $\boldsymbol{z} \equiv \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_L\}$ which includes all the network measurements. We assume that the hypothesis testing problem can be expressed as a parameter test [22]. We let the joint pdf of $\boldsymbol{z}_l$ with vector parameter $\boldsymbol{\theta} \in \mathbb{R}^M$ be $p(\boldsymbol{z}_l; \boldsymbol{\theta})$. The true vector parameter under the hypothesis $\mathcal{H}_i$ is $\boldsymbol{\theta}^i$, $i = 0, 1$. The test is

$$\begin{cases} \mathcal{H}_0 : & \boldsymbol{z}_l \overset{iid}{\sim} p(\boldsymbol{z}_l; \boldsymbol{\theta}^0) \\ \mathcal{H}_1 : & \boldsymbol{z}_l \overset{iid}{\sim} p(\boldsymbol{z}_l; \boldsymbol{\theta}^1), \; l \in [1 : L], \end{cases} \tag{1}$$

where we assume that $\boldsymbol{\theta}^0$ is known, and $\boldsymbol{\theta}^1 \neq \boldsymbol{\theta}^0$ is unknown. We define the local vector parameter $\boldsymbol{\theta}_k^{\text{loc}} \in \mathbb{R}^{M_k}$, $k \in [1 : N]$, as the parameter that completely describes de marginal pdf of the k-th node, i.e., $\int \cdots \int p(\boldsymbol{z}_l; \boldsymbol{\theta}) dz_1(l) \ldots dz_{k-1}(l) dz_{k+1}(l) \ldots dz_N(l) = p_k(z_k(l); \boldsymbol{\theta}_k^{\text{loc}})$. We let $\boldsymbol{\theta}^{\text{loc}} \equiv \{\boldsymbol{\theta}_1^{\text{loc}}, \ldots, \boldsymbol{\theta}_N^{\text{loc}}\}$, $\boldsymbol{\theta}^{\text{loc}} \in \mathbb{R}^P$, $P = \sum_{k=1}^N M_k$. In general, $\boldsymbol{\theta}^{\text{loc}}$ is a subset of $\boldsymbol{\theta}$ or a function of it. In either case, $\boldsymbol{\theta}^{\text{loc}}$ is the set of parameters that are *observable* at individual nodes and can be estimated locally without knowledge of the samples taken at other nodes.

III. Proposed statistics

A. A brief review of the GLR test

To perform the test (1) we consider the GLR test approach, frequently used in the literature and whose asymptotic performance can be computed analytically [22], [23]. The classical GLR statistic is $T_{\text{G}}(\boldsymbol{z}) \equiv \frac{p(\boldsymbol{z}; \boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{\text{G-MLE}})}{p(\boldsymbol{z}; \boldsymbol{\theta} = \boldsymbol{\theta}^0)}$, where $\hat{\boldsymbol{\theta}}_{\text{G-MLE}}$ is the (global) MLE[1] of $\boldsymbol{\theta}_1$. The asymptotic distribution of the global GLR statistic T_{G} under *weak* conditions (i.e., there exists a constant c such that $\|\boldsymbol{\theta}^1 - \boldsymbol{\theta}^0\| \leq c/\sqrt{L}$) is well known [22]:

$$2 \log T_{\text{G}}(\boldsymbol{z}) \overset{a}{\sim} \begin{cases} \chi_M^2 & \text{under } \mathcal{H}_0 \\ \chi_M'^2(\lambda_g) & \text{under } \mathcal{H}_1, \end{cases} \tag{2}$$

[1]We call it *global* MLE to differentiate it from the local one defined next.

where the symbol $\overset{a}{\sim}$ means "asymptotically distributed as when L tends to infinity", χ_M^2 is the chi-square distribution with M degrees of freedom and $\chi_M'^2(\lambda_g)$ is the non-central chi-square distribution with M degrees of freedom and non-centrality parameter

$$\lambda_g = L(\boldsymbol{\theta}^1 - \boldsymbol{\theta}^0)^T \boldsymbol{i}(\boldsymbol{\theta}^0)(\boldsymbol{\theta}^1 - \boldsymbol{\theta}^0), \tag{3}$$

where $\boldsymbol{i}(\boldsymbol{\theta}^0)$ is the Fisher information matrix at $\boldsymbol{\theta}^0$ [24].

B. Local MLE estimate

In many cases the global MLE is difficult to compute in a distributed scenario. The reason for this is two-fold. In first place, depending of the structure of $p(\boldsymbol{z}_l; \boldsymbol{\theta})$, the maximization problem can be very hard from a computational point of view, even in a centralized scenario. On the other hand, computing the global MLE would require in general a large number of sensor communications (of the sensor measurements or some statistics of them) imposing a serious practical constraint in terms of energy, bandwidth and/or delay. Looking for a simpler approach to compute the MLE we consider a *local* MLE in the sensor node k which only use the locally sensed values $\{z_k(l)\}_{l=1}^L$, whose distribution under $\mathcal{H}_1$ is $\prod_{l=1}^L p_k(z_k(l), \boldsymbol{\theta}_k^{\text{loc},1})$. Thus, the k-th local parameter estimate $\hat{\boldsymbol{\theta}}_k^{\text{loc}}$ of $\boldsymbol{\theta}_k^{\text{loc},1}$ is defined as, $k \in [1 : N]$,

$$\hat{\boldsymbol{\theta}}_k^{\text{loc}} \equiv \arg\max_{\boldsymbol{\theta}_k^{\text{loc}}} \tfrac{1}{L} \textstyle\sum_{l=1}^L \log p_k(z_k(l); \boldsymbol{\theta}_k^{\text{loc}}). \tag{4}$$

It is important to note that $\hat{\boldsymbol{\theta}}_k^{\text{loc}}$ is still an asymptotically consistent estimator of the true parameter $\boldsymbol{\theta}_k^{\text{loc},1}$ given that it is a maximum likelihood estimator. However, the concatenation of all local vector parameters $\boldsymbol{\theta}^{\text{loc}}$ could loss the property of asymptotic *efficiency* of the global MLE [24] for the same set of parameters, given that $\{\hat{\boldsymbol{\theta}}_k^{\text{loc}}\}_{k=1}^N$ are estimated through the corresponding marginal distributions instead of the joint probability. Clearly, this estimation efficiency loss could impact negatively in the detection performance of a test that uses the local estimator. Thus, a theoretical performance characterization is important, among others things, to evaluate this effect. We will cover this in the next section.

C. A fully distributed statistic

From a point of view of the distributed implementation of an algorithm, it would be beneficial to factorize the joint pdf of the observations such that each node can compute a part of the whole statistic using only local measurements and then, use a simple cooperative scheme to compute the final statistic, exchanging only a reduced amount of information between the nodes. Therefore, we propose to build a statistic using the marginal distributions instead of the joint distribution of the data. Specifically, we define the following statistic:

$$\log T_{\text{L-MP}}(\boldsymbol{z}) \equiv \sum_{k=1}^N \log \frac{p_k(\{z_k(l)\}_{l=1}^L; \boldsymbol{\theta}_k^{\text{loc}} = \hat{\boldsymbol{\theta}}_k^{\text{loc}})}{p_k(\{z_k(l)\}_{l=1}^L; \boldsymbol{\theta}_k^{\text{loc}} = \boldsymbol{\theta}_k^{\text{loc},0})} \tag{5}$$

where the subscript refers to the statistic that uses the local estimation of the parameters (L) and the joint pdf is replaced by the product of the marginal (MP) pdfs under each hypothesis. In the next section we will obtain the asymptotic distribution of this statistic, and in Sec. V and VI, we will present some examples for which $T_{\text{L-MP}}$ has better performance than the full GLR statistic T_{G}. This can be explained by the fact that the widely used GLR statistic, in general, has not optimality guarantees for composite hypothesis testing problems [22], [23], despite the fact that it uses the full dependence structure of $p(\boldsymbol{z};\boldsymbol{\theta})$.

Additionally, the structure of $T_{\text{L-MP}}$ opens opportunities to save valuable resources in a WSN such as energy and bandwidth for communicating the quantities required to run the detection algorithms. Considering $\log T_{\text{L-MP}}(\boldsymbol{z})$ we see that each sensor is able to compute its corresponding term in the sum and then, share this quantity to the rest of the sensors to obtain $\log T_{\text{L-MP}}$ via a simple consensus algorithm.

D. Spatial averaging

The statistic $\log T_{\text{L-MP}}$ in (5) requires the computation of a *spatial* sum $\sum_{k=1}^{N}(\cdot)$ over all the sensors in the network. Next $\bar{a} \equiv \sum_{k=1}^{N} a_k$ will represent that sum. Each sensor node generates locally a scalar value $a_k \in \mathbb{R}$, $k \in \mathcal{N} \equiv [1:N]$ and it is desired to compute the average $\tilde{a} = \frac{1}{N}\sum_{k=1}^{N} a_k$ (or the sum $\bar{a} = N\tilde{a}$) at each node in a distributed manner and with minimal resources allocated to the exchanges between the nodes.

The spatial averages can be computed via a consensus procedure such as in [12], [25], [26]. Consider a network (modeled as a connected graph) $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ consisting of a set of nodes $\mathcal{N}$ and a set of edges $\mathcal{E}$, where each edge $\{i,j\} \in \mathcal{E}$ is an unordered pair of distinct nodes. The set of neighbors of node i is denoted by $\mathcal{N}_i = \{j \in \mathcal{N} | \{i,j\} \in \mathcal{E}\}$. The average value $\tilde{a}$ can be computed iteratively as, $t \in \mathbb{N}$:

$$a_k(t) = W_{kk} a_k(t-1) + \textstyle\sum_{j\in\mathcal{N}_k} W_{kj} a_j(t-1), \quad k \in \mathcal{N}, \tag{6}$$

where $a_k(t)$ is the average after t iterations (or message exchanges between the nodes), $a_k(0) = a_k$ is the initial value and W_{kj} is the weight on $a_j(t-1)$ at the node k. Considering local transmissions only, i.e., each node broadcasts its local value at iteration t only to the nodes in its neighborhood, we have that for each $k \in \mathcal{N}$, $W_{kj} = 0$ for $j \notin \mathcal{N}_k$ and $j \neq k$.

Among all the existing possibilities for selecting the weights, we will consider a simple but effective algorithm called *local-degree weights* distributed averaging algorithm [25]. Its convergence to the required average is guaranteed given that graph is not bipartite. The weights are assumed to be symmetric with value $W_{kj} = W_{jk} = 1/\max(d_k, d_j)$, where d_k is the degree of node k, i.e., the number of neighbors of the node k. Algorithm 1 summarizes the steps required to compute the statistic $T_{\text{L-MP}}$. Several stopping criteria can be considered for the iterative computation of the spatial average (6). Here we consider a fixed number of exchanges N_{it}.

IV. Asymptotic performance analysis

The asymptotic pdf of $T_{\text{L-MP}}$ is presented next. The proof is based on classical tools used in the GLR theory, and it is relegated to Appendix A.

Algorithm 1 Distributed implementation of $T_{\text{L-MP}}$

```
for k = 1, ..., N do (simultaneously at each sensor)
    Compute the local estimate θ̂_k^loc using eq. (4).
    Compute log T_k ≡ log [ p_k({z_k(l)}_{l=1}^L; θ_k = θ̂_k^loc) / p_k({z_k(l)}_{l=1}^L; θ_k = θ_k^{loc,0}) ].
    log T_{L-MP,k} = SpatialSum({log T_j}_{j∈N_k ∪ {k}})
    if log T_{L-MP,k} < γ then Sensor k decides H_0,
    else Sensor k decides H_1.
    end if          ▷ γ is the predefined threshold of the test.
end for
function SpatialSum({a_k}_{k∈N_k ∪ {k}})      ▷ Compute ā_k.
    t = 0, a_k(0) = a_k          ▷ Initial condition for t = 0.
    while t < N_it do
        t = t + 1
        Compute the spatial average a_k(t) using (6).
    end while
    return N a_k(t)              ▷ Return the sum ā_k
end function
```

Lemma 1. Assume i) the first and second-order derivatives of the log-likelihood function are well defined and continuous functions. ii) $\mathbb{E}[\partial \log p_k(z_k(l);\boldsymbol{\theta}_k^{\text{loc}})/\partial\boldsymbol{\theta}_k^{\text{loc}}] = \mathbf{0}$, $\forall l$, $k \in [1:N]$. iii) the matrix $\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})$ defined in (9) is nonsingular. Then, the asymptotic distribution of $\hat{\boldsymbol{\theta}}^{\text{loc}}$ under $\mathcal{H}_i$, $i=0,1$, is

$$\hat{\boldsymbol{\theta}}^{\text{loc}} \overset{a}{\sim} \mathcal{N}\left(\boldsymbol{\theta}^{\text{loc},i}, \tfrac{1}{L}\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})^{-1}\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},i})\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})^{-1}\right), \tag{7}$$

where the (k,j) $M_k \times M_j$ sub-matrix of $\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},i})$ and $\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})$ are respectively defined by, $k,j \in [1:N]$,

$$[\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},i})]_{kj} \equiv \mathbb{E}_{\boldsymbol{\theta}^i}\left(\frac{\partial \log p_k(z_k(l);\boldsymbol{\theta}_k^{\text{loc}})}{\partial\boldsymbol{\theta}_k^{\text{loc}}}\frac{\partial \log p_j(z_j(l);\boldsymbol{\theta}_j^{\text{loc}})^T}{\partial\boldsymbol{\theta}_j^{\text{loc}}}\right), \tag{8}$$

$$[\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})]_{kj} \equiv -\mathbb{E}_{\boldsymbol{\theta}_k^{\text{loc},i}}\left(\frac{\partial^2 \log p_k(z_k(l);\boldsymbol{\theta}_k^{\text{loc}})}{\partial\boldsymbol{\theta}_k^{\text{loc}}\partial\boldsymbol{\theta}_j^{\text{loc}}}\right), \tag{9}$$

where the expectations are taken with respect to $p(\boldsymbol{z}_l;\boldsymbol{\theta}^i)$, and the marginal pdf $p_k(z_k(l);\boldsymbol{\theta}_k^{\text{loc},i})$, respectively. Also, the asymptotic distribution of $T_{\text{L-MP}}$ under $\mathcal{H}_i$ is:

$$2\log T_{\text{L-MP}}(\boldsymbol{z}) \overset{a}{\sim} f_P(\boldsymbol{\mu}_{\text{MP},i}, \boldsymbol{\Sigma}_{\text{MP},i}) \;\; i=0,1, \tag{10}$$

where $\boldsymbol{\mu}_{\text{MP},0} = \mathbf{0}$, $\boldsymbol{\mu}_{\text{MP},1} = \sqrt{L}\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},1})^{\frac{1}{2}}(\boldsymbol{\theta}^{\text{loc},1} - \boldsymbol{\theta}^{\text{loc},0})$, and $\boldsymbol{\Sigma}_{\text{MP},i} = \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},i})^{\frac{1}{2}}\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})^{-1}\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},i})\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})^{-1}\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},i})^{\frac{1}{2}}$, $p_{\text{MP}}(\boldsymbol{z}_l;\boldsymbol{\theta}^{\text{loc},i}) \equiv \Pi_{k=1}^{N} p_k(z_k(l);\boldsymbol{\theta}_k^{\text{loc},i})$ is the product of the marginal pdfs used for building $T_{\text{L-MP}}$, and

$$\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},i}) \equiv \mathbb{E}_{\boldsymbol{\theta}^{\text{loc},i}}\left(\frac{\partial \log p_{\text{MP}}(\boldsymbol{z}_l;\boldsymbol{\theta}^{\text{loc},i})}{\partial\boldsymbol{\theta}^{\text{loc}}}\frac{\partial \log p_{\text{MP}}(\boldsymbol{z}_l;\boldsymbol{\theta}^{\text{loc},i})^T}{\partial\boldsymbol{\theta}^{\text{loc}}}\right),$$

with the expectation taken respect to $p_{\text{MP}}(\boldsymbol{z}_l, \boldsymbol{\theta}^{\text{loc},i})$. We also define $f_P(\boldsymbol{\mu}, \boldsymbol{\Sigma})$[2] as the pdf of $\|\boldsymbol{n}\|^2$ when $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Corollary 1. If the signal to be tested is weak, i.e., there exists a constant c such that $\|\boldsymbol{\theta}^{\text{loc},1} - \boldsymbol{\theta}^{\text{loc},0}\| \leq \frac{c}{\sqrt{L}}$, then, as $L \to \infty$,

$$\|\boldsymbol{\mu}_{\text{MP},1}\|^2 = L(\boldsymbol{\theta}^{\text{loc},1} - \boldsymbol{\theta}^{\text{loc},0})^T \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},0})(\boldsymbol{\theta}^{\text{loc},1} - \boldsymbol{\theta}^{\text{loc},0}) \tag{11}$$

[2]Note that $f_P(\mathbf{0}_P, I_P)$ ($f_P(\boldsymbol{\mu}, I_P)$) is the (non-central) chi-2 pdf with P degrees of freedom (with non-centrality parameter $\|\boldsymbol{\mu}\|^2$).

The proof is based on a Taylor expansion of $\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},1})$ around $\boldsymbol{\theta}^{\text{loc},0}$, where the higher order terms are discarded using the weak signal condition as $L \to \infty$.

Remark 1. If the spatial measurements are independent, i.e., $p(\boldsymbol{z}_l;\boldsymbol{\theta}) = p_{\text{MP}}(\boldsymbol{z}_l;\boldsymbol{\theta}^{\text{loc}})$, and the local parameter vector equals to the global parameter $\boldsymbol{\theta}^{\text{loc}} = \boldsymbol{\theta}$, then the Fisher information matrix satisfies $\boldsymbol{i}(\boldsymbol{\theta}^i) = \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^i) = \tilde{\boldsymbol{i}}(\boldsymbol{\theta}^i) = \boldsymbol{j}(\boldsymbol{\theta}^i)$, $\boldsymbol{\Sigma}_{\text{MP},i} = \mathbf{I}_M$, $i = 0, 1$. Using Corollary 1, we have that $\|\boldsymbol{\mu}_{\text{MP},1}\|^2 = \lambda_g$. Thus, the asymptotic performance of $T_{\text{L-MP}}$ and the GLR test (2) is the same. We show this result to demonstrate consistency of Lemma 1 for this particular case, though in this paper we are clearly interested in the case where the spatial random variables are dependent.

The pdf f_P and its corresponding cumulative distribution function have not closed form expressions but can be tightly approximated using the Lugannani-Rice approximation [27] given the fact that the corresponding moment generating function is easy to obtain. We will use this approximation for evaluating the miss-detection and false alarm probabilities of the statistic $T_{\text{L-MP}}$. For a given statistic T and a predefined threshold τ, the miss-detection and false alarm probabilities are, respectively, $P_{\text{md}} = \mathbb{P}(T < \tau|\mathcal{H}_1)$, and $P_{\text{fa}} = \mathbb{P}(T > \tau|\mathcal{H}_0)$.

The deflection coefficient of a statistic is typically used as a simplified measure of performance when the error probabilities are difficult to compute. However, it is a true distance measure when the statistic is Gaussian distributed [23]. In this paper, we will use it in order to interpret and try to understand the performance of the proposed statistic. The deflection coefficient of T is defined as $D_T^2 = \frac{(\mathbb{E}_1(T)-\mathbb{E}_0(T))^2}{\mathbb{V}\text{ar}_0(T)}$, where $\mathbb{V}\text{ar}_0$ is the variance operator under $\mathcal{H}_0$. In some detectors where certain parameters are unknown and must be estimated through the data, even the mean or the variance can be rather cumbersome to compute. Therefore, we use the asymptotic characterization of the distributions to obtain the deflection coefficient. Next we provide both, the GLR and L-MP deflection coefficients:

$$D_{\text{G}}^2 = \frac{\lambda_g^2}{2M},\ D_{\text{L-MP}}^2 = \frac{(\|\boldsymbol{\mu}_{\text{MP},1}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\text{MP},1} - \boldsymbol{\Sigma}_{\text{MP},0}))^2}{2\text{tr}(\boldsymbol{\Sigma}_{\text{MP},0}^2)}, \tag{12}$$

where we have use (2), Lem. 1, and that, asymptotically, $\mathbb{E}_{\boldsymbol{\theta}^{\text{loc},i}}(2\log T_{\text{L-MP}}(\boldsymbol{z})) = \|\boldsymbol{\mu}_{\text{MP},i}\|^2 + \text{tr}(\boldsymbol{\Sigma}_{\text{MP},i})$, $i = 0, 1$, and $\mathbb{V}\text{ar}_0(2\log T_{\text{L-MP}}(\boldsymbol{z})) = 2\text{tr}(\boldsymbol{\Sigma}_{\text{MP},0}^2)$.

V. Example

Consider the following hypothesis testing problem:

$$\begin{cases} \mathcal{H}_0: & \boldsymbol{z}_l \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{C}) \\ \mathcal{H}_1: & \boldsymbol{z}_l \overset{iid}{\sim} \mathcal{N}(\boldsymbol{\mu}_1, \mathbf{C}),\ l \in [1:L]. \end{cases} \tag{13}$$

The parameter test consists in testing the mean $\boldsymbol{\theta} = \boldsymbol{\theta}^0 = \mathbf{0} \in \mathbb{R}^N$ against $\boldsymbol{\theta} \neq \mathbf{0}$ with statistically dependent observations across the sensors. In this case $\boldsymbol{\theta}^1 = \boldsymbol{\mu}_1$ is the unknown vector parameter, $\boldsymbol{\theta}^{\text{loc}} = \boldsymbol{\theta} \in \mathbb{R}^N$, $M_k = 1\ \forall k$, $M = P = N$. We consider two cases. First, we assume that the covariance matrix $\mathbf{C}$ is known, and then, in the second example of this section we will consider it unknown.

A. Covariance matrix known

It is easy to show that the global MLE is the sample mean of the observations, and that it coincides with the local MLE: $\hat{\boldsymbol{\theta}}^{\text{loc}} = \frac{1}{L}\sum_{l=1}^{L}\boldsymbol{z}_l$. Thus, in this case, the correlation does not introduce a penalty in estimating the parameter locally. Then,

$$2\log T_{\text{G}}(\boldsymbol{z}) = L(\hat{\boldsymbol{\theta}}^{\text{loc}})^T\mathbf{C}^{-1}\hat{\boldsymbol{\theta}}^{\text{loc}}. \tag{14}$$

We now compute the parameters to characterize the asymptotic performance of $T_{\text{L-MP}}$. For the Gaussian case, the (k,j)-th component of the Fisher information matrix can be computed as [24]

$$\begin{aligned}[\boldsymbol{i}(\boldsymbol{\theta})]_{kj} = {} & \frac{\partial\boldsymbol{\mu}_1(\boldsymbol{\theta})^T}{\partial\theta_k}\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial\boldsymbol{\mu}_1(\boldsymbol{\theta})}{\partial\theta_j} \\ & + \frac{1}{2}\text{tr}\left[\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\theta_k}\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\theta_j}\right].\end{aligned} \tag{15}$$

Thus, $\boldsymbol{i}(\boldsymbol{\theta}^i) = \mathbf{C}^{-1}$, independently of $\boldsymbol{\theta}^i$, $i = 0, 1$. It can be shown that $\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^i) = \text{diag}(\mathbf{C})^{-1}\mathbf{C}\ \text{diag}(\mathbf{C})^{-1}$ and that $\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^i) = \boldsymbol{j}(\boldsymbol{\theta}^i) = \text{diag}(\mathbf{C})^{-1}$. Then, the asymptotic covariance of the local estimator given in (7) is $\frac{1}{L}\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})^{-1}\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},i})\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},i})^{-1} = \frac{1}{L}\mathbf{C}$ and coincides with that one of the global MLE. That is, the local and global MLE are asymptotically equivalent, something expected given the equivalence of both estimators for finite data size. The $T_{\text{L-MP}}$ statistic is easily computed as:

$$2\log T_{\text{L-MP}}(\boldsymbol{z}) = L(\hat{\boldsymbol{\theta}}^{\text{loc}})^T\text{diag}(\mathbf{C})^{-1}\hat{\boldsymbol{\theta}}^{\text{loc}}, \tag{16}$$

with asymptotic parameters $\boldsymbol{\mu}_{\text{MP},1} = \sqrt{L}\text{diag}(\mathbf{C})^{-\frac{1}{2}}\boldsymbol{\theta}^1$, and $\boldsymbol{\Sigma}_{\text{MP},i} = \text{diag}(\mathbf{C})^{-\frac{1}{2}}\mathbf{C}\text{diag}(\mathbf{C})^{-\frac{1}{2}}$ for $i = 0, 1$.

The behavior of the statistics depends on the parameter vector $\boldsymbol{\theta}^1$ and the covariance matrix $\mathbf{C}$, fundamentally through the variance in the direction of $\boldsymbol{\theta}^1$ given by $\mathbf{C}$. This is clearly understood if we look at the corresponding deflection coefficients (12). For the present example, they result in[3]:

$$D_{\text{G}}^2 = \frac{(L(\boldsymbol{\theta}^1)^T\mathbf{C}^{-1}\boldsymbol{\theta}^1)^2}{2N},\ D_{\text{L-MP}}^2 = \frac{(L(\boldsymbol{\theta}^1)^T\text{diag}(\mathbf{C})^{-1}\boldsymbol{\theta}^1)^2}{2\text{tr}((\mathbf{C}\text{diag}(\mathbf{C})^{-1})^2)}.$$

In general, for a fixed $\mathbf{C}$, the deflection coefficients have the following extremes as a function of $\boldsymbol{\theta}^1$. In the case of D_{G}^2, the maximum is $\frac{L^2\|\boldsymbol{\theta}^1\|^4}{2N\lambda_{\text{min}}^2}$ attained for $\boldsymbol{\theta}^1 = \|\boldsymbol{\theta}^1\|\boldsymbol{v}_{\text{min}}$, where λ_{min} is the minimum eigenvalue of $\mathbf{C}$ and $\boldsymbol{v}_{\text{min}}$ its corresponding eigenvector. On the other hand, the minimum D_{G}^2 is $\frac{L^2\|\boldsymbol{\theta}^1\|^4}{2N\lambda_{\text{max}}^2}$ attained for $\boldsymbol{\theta}^1 = \|\boldsymbol{\theta}^1\|\boldsymbol{v}_{\text{max}}$, where λ_{max} is the maximum eigenvalue of $\mathbf{C}$ and $\boldsymbol{v}_{\text{max}}$ its corresponding eigenvector.

In the case of $D_{\text{L-MP}}^2$, the maximum is $\frac{L^2\|\boldsymbol{\theta}^1\|^4}{2\text{tr}((\mathbf{C}\text{diag}(\mathbf{C})^{-1})^2))\sigma_{\text{min}}^4}$ attained for $\boldsymbol{\theta}^1 = \|\boldsymbol{\theta}^1\|\boldsymbol{e}_{\text{min}}$, where σ_{min}^2 is the minimum variance in the diagonal of $\mathbf{C}$, and $\boldsymbol{e}_{\text{min}}$ is the canonical vector with

[3] In this case, the deflection coefficient can also be easily computed using the non-asymptotic data statistics given that $\hat{\boldsymbol{\theta}}^{\text{loc}}$ is a Gaussian vector, and it results the same as using the asymptotic statistics.

all zero components but 1 in the component corresponding to the minimum variance. On the other hand, the minimum $D^2_{\text{L-MP}}$ is $\frac{L^2\|\boldsymbol{\theta}^1\|^4}{2\text{tr}((\mathbf{C}\text{diag}(\mathbf{C})^{-1})^2))\sigma^4_{\text{max}}}$ attained for $\boldsymbol{\theta}^1 = \|\boldsymbol{\theta}^1\|\boldsymbol{e}_{\text{max}}$, where σ^2_{max} is the maximum variance in the diagonal of $\mathbf{C}$, and $\boldsymbol{e}_{\text{max}}$ is the corresponding canonical vector.

The covariance matrix could be arbitrary but we set it as follows in order to be controlled by a single parameter ρ:

$$\mathbf{C} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{N-1} \\ \rho & 1 & \rho & \dots & \rho^{N-2} \\ & & \ddots & & \\ \rho^{N-1} & \rho^{N-2} & \dots & \rho & 1 \end{bmatrix}. \tag{17}$$

In Fig. 1 we plot the quotient $Q \equiv D^2_{\text{L-MP}}/D^2_{\text{G}}$ in dB, for $N = 10$ and $N = 100$, as a function of the correlation parameter ρ, for $\mathbf{C}$ in (17). As all the elements of the diagonal of $\mathbf{C}$ are equal in this case, $D^2_{\text{L-MP}}$ does not depend on the direction of $\boldsymbol{\theta}^1$. Thus, Q will be determined by the direction of $\boldsymbol{\theta}^1$ in the expression of D^2_{G}. Q_{max} and Q_{min} are the maximum and minimum of Q, while Q_1 and Q_{-1} are are obtained evaluating Q at $\boldsymbol{\theta}^1/\|\boldsymbol{\theta}^1\| = [1, 1, \dots, 1]^T/\sqrt{N}$ and $\boldsymbol{\theta}^1/\|\boldsymbol{\theta}^1\| = [1, -1, \dots, 1]^T/\sqrt{N}$, respectively. As observed, when all the components of $\boldsymbol{\theta}_1$ are equal, Q is very near to its maximum, suggesting a better performance for $T_{\text{L-MP}}$ as compared to T_{G}.

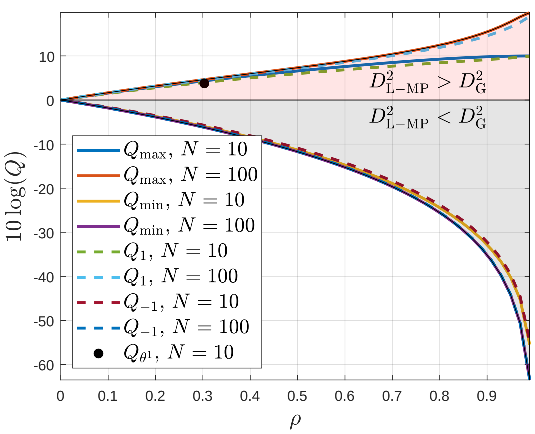
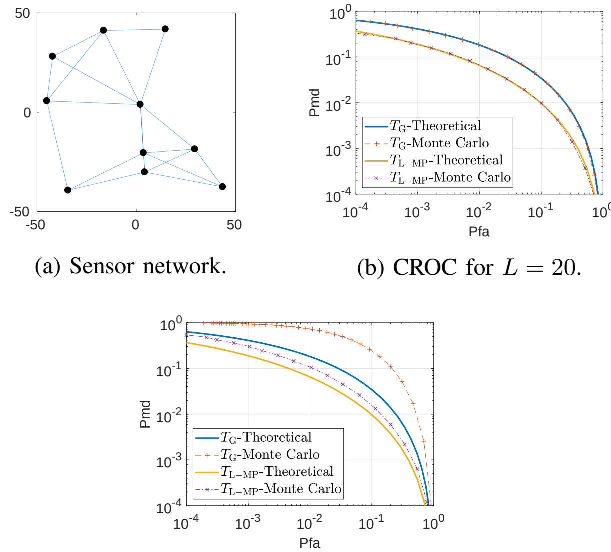


Figure 1: Deflection coefficients comparison for $N = 10, 100$.

Consider now the network represented through its graph shown in the Fig. 2a with $N = 10$ nodes and $|\mathcal{E}| = 20$ edges. In this example, the graph is used to define the neighbours of each node in order to run Algorithm 1. To build the network, we randomly generated 10 nodes, uniformly distributed on a square of 100×100 m^2. We impose that two nodes are connected by an edge if their distance is less than a predefined threshold. Then we increase this threshold until the total number of edges is 20 and check that the resulting graph is connected. Consider also $N_{it} = 20$, $\rho = 0.3$, 10^5 Monte Carlo realizations of (13), and $\boldsymbol{\theta}^1 = [.33, .55, .32, .57, .42, .46, .52, .43, .35, .32]^T$ which was uniformly and randomly generated with components in $[.35, .6]$.



(a) Sensor network.

(b) CROC for $L = 20$.

(c) $\mathbf{C}$ unknown, $L = 20$.

Figure 2: Performance of test for a spatially correlated Gaussian distribution with unknown mean.

In Fig. 1 we plot as a black dot the quotient Q for this case, denoted as $Q_{\boldsymbol{\theta}^1}$. In Fig. 2b we plot the complementary receiver operating characteristic (CROC). Note that the asymptotic theoretical curves match perfectly the Monte Carlo simulations. The reason is that in this case the asymptotic distribution corresponds also to test distribution in the finite length regime. A remarkable fact is that the $T_{\text{L-MP}}$ statistic outperforms the GLR statistic. It is in concordance with the deflection coefficients analysis.

B. *Covariance matrix unknown*

If the covariance is unknown, it must be estimated, at least part of it, in order to implement any of the statistics considered in this work. As this parameter is unknown but the same under each hypothesis it is a *nuisance* parameter $\boldsymbol{\theta}_s$ [22]. The parameter test is, in this case:

$$\begin{cases} \mathcal{H}_0: & \boldsymbol{\theta}_r = \boldsymbol{\theta}_r^0, \boldsymbol{\theta}_s \\ \mathcal{H}_1: & \boldsymbol{\theta}_r \neq \boldsymbol{\theta}_r^0, \boldsymbol{\theta}_s, \end{cases} \tag{18}$$

where $\boldsymbol{\theta}_r^0 = \mathbf{0}_N$ and $\boldsymbol{\theta}_r^1 = \boldsymbol{\mu}_1 \in \mathbb{R}^N$ and $\boldsymbol{\theta}^{\text{loc}} = \boldsymbol{\theta}$. Let us begin with the GLR statistic. In this case all the elements of $\mathbf{C} \in \mathbb{R}^{N\times N}$ must be estimated. Considering that the covariance matrices are symmetric, it is sufficient to estimate the lower diagonal and the diagonal elements of it. Thus, the nuisance parameter is $\boldsymbol{\theta}_s = \text{vech}(\mathbf{C}) \in \mathbb{R}^{\frac{1}{2}N(N+1)}$, where the vech operator concatenates all but the supra diagonal elements of $\mathbf{C}$. To evaluate the performance of the test it is required to compute the Fisher information matrix of the vector parameter $\boldsymbol{\theta} \equiv [\boldsymbol{\theta}_r^T, \boldsymbol{\theta}_s^T]^T \in \mathbb{R}^{N+\frac{1}{2}N(N+1)}$. For the Gaussian case, the

(k,j)-th component of the Fisher information matrix can be computed as [24]

$$[\boldsymbol{i}(\boldsymbol{\theta})]_{kj} = \frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\theta})^T}{\partial \theta_k} \mathbf{C}(\boldsymbol{\theta})^{-1} \frac{\partial \boldsymbol{\mu}_1(\boldsymbol{\theta})}{\partial \theta_j} + \frac{1}{2}\text{tr}\left[\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_k}\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \theta_j}\right]. \quad (19)$$

Then, $\boldsymbol{i}(\boldsymbol{\theta}) \in \mathbb{R}^{(N+\frac{1}{2}N(N+1))\times(N+\frac{1}{2}N(N+1))}$ is a block diagonal matrix where the first $N \times N$ sub-matrix is $\mathbf{C}^{-1}$. The fact that $\boldsymbol{i}(\boldsymbol{\theta})$ is a diagonal block makes the asymptotic performance of the test with or without nuisance parameters the same [22, Ch. 6.5]. This happens because in the case of nuisance parameters, the non-centrality parameter of the non-central chi-square distribution depends on a quadratic form that includes the Schur complement of the second $\frac{1}{2}N(N+1) \times \frac{1}{2}N(N+1)$ block matrix of $\boldsymbol{i}(\boldsymbol{\theta})$. Thus, if the out of diagonal block matrices are zero, the non-central parameter does not change with or without nuisance parameters.

Something similar happens with the statistic $T_{\text{L-MP}}$. The difference is that for this case the nuisance vector parameter $\boldsymbol{\theta}_s \in \mathbb{R}^N$ has as components only the diagonals element of $\mathbf{C}$ (which are the only parameters that are observable at each sensor node). Following a similar reasoning to [22, App. 6B], it can be shown that the matrix that controls the asymptotic behavior is $\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc}})$ instead of $\boldsymbol{i}(\boldsymbol{\theta})$, given that is the one that appears in the factorization (34). Considering the partition of the $\boldsymbol{\theta}$ in $\boldsymbol{\theta}_r$ and $\boldsymbol{\theta}_s$, $\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc}})$ can be partitioned as $[\boldsymbol{i}_{rr}, \boldsymbol{i}_{rs}; \boldsymbol{i}_{sr}, \boldsymbol{i}_{ss}] \in \mathbb{R}^{2N\times 2N}$, where for this example $\boldsymbol{i}_{rr}, \boldsymbol{i}_{rs}, \boldsymbol{i}_{sr}, \boldsymbol{i}_{ss} \in \mathbb{R}^{N\times N}$. Then if $\boldsymbol{i}_{rs}, \boldsymbol{i}_{sr} \in \mathbf{0}$, the asymptotic performance does not change if there are or not nuisance parameter. This happens, as in the present example, if the local parameters $\{\boldsymbol{\theta}^{\text{loc}}_{r,k}\}_{k=1}^N$ and the local nuisance parameters $\{\boldsymbol{\theta}^{\text{loc}}_{s,k}\}_{k=1}^N$ are independent parameters, i.e., $\partial\boldsymbol{\theta}^{\text{loc}}_{r,k}/\partial\boldsymbol{\theta}^{\text{loc}}_{s,n} = 0$, $n, k \in [1:N]$.

In Fig. 2c we show the CROC. Note that the GLR statistic asymptotic performance overestimate the performance for the finite length data set. Nonetheless, the $T_{\text{L-MP}}$ statistic asymptotic performance presents a good agreement with the Monte Carlo estimation for this finite length data set. The difference between the theoretical and the Monte Carlo curves are attributed to the presence of the nuisance parameter and the relatively few data set length ($L = 20$). It is also shown that the performance of $T_{\text{L-MP}}$ is better than that of the GLR, about an order of magnitude in the miss detection probability when the false alarm probability is 10^{-2}.

Finally, the estimation of the diagonal elements of $\mathbf{C}$ does not require any transmissions, so the number of transmissions needed is the same as in the case of $\mathbf{C}$ known.

C. Communication energy analysis

One of the tasks which requires a substantial part of the energy budget of a WSN is the communication between sensors for exchanging information [28], [29]. Each summand of $\log T_{\text{L-MP}}$ in (5) is estimated locally at each node without communicating with other nodes. Thus, the network need to run only one time the consensus algorithm to compute $\log T_{\text{L-MP}}$ as indicated in Algorithm 1. Therefore, NN_{it} transmissions in the whole network are required for computing this statistic. This is the case, in general, for any parameter test.

Statistic	# total transmissions in the network
$T_{\text{L-MP}}$	NN_{it}
T_{G}, $\mathbf{C}$ known	$N(N+1)N_{it}$
T_{G}, $\mathbf{C}$ unknown	$N((N+1)N_{it}+L)$

Table I: Number of transmissions for computing the statistics.

Now we consider T_{G} in (14) when $\mathbf{C}$ is known. In this the case, the GLR can be expressed using the inverse of the square root matrix of the covariance matrix $\mathbf{R} \equiv \mathbf{C}^{-\frac{1}{2}}$, assumed to be precomputed and stored at each node, as $2\log T_{\text{G}} = L\|\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{G-MLE}}\|^2 = \sum_{k=1}^N (\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{G-MLE}})_k^2$. This means that N runs of a consensus algorithm are needed for computing each component of $\mathbf{R}\hat{\boldsymbol{\theta}}_{\text{G-MLE}}$ and one more run for computing the sum in the index k. Then, the previous statistic is computed in $N(N+1)N_{it}$ local broadcast communications.

Now we consider a case that will be presented in the next section, in which the distribution is still Gaussian, but both the mean and the covariance matrix $\mathbf{C}$ are unknown and must be estimated. The global MLE of the mean and covariance matrix for a Gaussian distribution is known to be the sample mean and the sample covariance matrix. The sample covariance matrix expression it is not amenable to be implemented in a distributed scenario. Then, consider the approach given in [30] for estimating the precision matrix (the inverse of the covariance matrix), assumed sparse for a Gaussian graphical model. This technique is shown to be applicable with good accuracy even in scenarios where the actual precision matrix is not sparse, while the sparsity pattern is obtained by thresholding the true precision matrix. Therefore, we consider this procedure only as a reference for computing the number of transmissions needed for estimating $\mathbf{C}^{-1}$ in a more favorable case. In this case, each node must broadcast its L observations to its neighbors, so the number of transmissions in the network for estimating $\mathbf{C}^{-1}$ is NL. For computing T_{G}, we have to add the corresponding transmissions already computed, resulting in a total of $N((N+1)N_{it}+L)$ transmissions. See Table I for a summary.

As it was shown, the implementation of T_{G} requires a network energy budget scaling with N^2 in both cases ($\mathbf{C}$ known and unknown), while the scaling factor of $T_{\text{L-MP}}$ is N, an order of magnitude less. This is a notorious advantage in favor of $T_{\text{L-MP}}$ when the performance achieved by it is adequate, specially for large sensor networks. When $\mathbf{C}$ is unknown, even in small networks $T_{\text{L-MP}}$ can save valuable network resources when L is large.

VI. Application to Spectrum Sensing

A. Model

In this section we consider as an application example a cognitive radio (CR) system, which emerged several years ago as a possible solution for the spectrum shortage (see [31]–[33] and references therein). In CR systems, unlicensed, or secondary users (SU), sense the spectrum in a particular place and time and wish to detect the presence or absence of licensed, or primary users (PU), in order to use the spectrum when it is available. In Fig. 3, we show a possible scenario of CR. One or several PUs transmit to their intended receivers located inside the corresponding primary range R_p, which defines the coverage area of this licensed system. The SUs, located at positions $\boldsymbol{r}_n^{\mathrm{SU}}$, $n = 1, \ldots, N$ in the two-dimensional space, sense the spectrum and decide if they are out of the *protected region* of the primary system which would allow them to use the spectrum without causing harmful interference to the PUs. R_s is defined as the interference range of the SUs. Consider a system with multiple PUs that could alternate

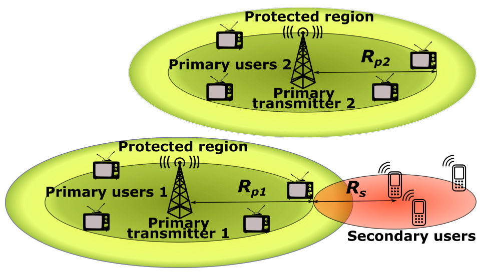


Figure 3: A CR network with spectrum sensing devices.

between active and inactive states. Let K be the amount of PUs each of which is indexed by $m = 1, \ldots, K$. Let s_m indicate the state of the m-th PU, located at $\boldsymbol{r}_m^{\mathrm{PU}}$ in the two-dimensional space. If the m-th PU transmits a signal (it is active), then $s_m = 1$, and if it is silent, $s_m = 0$. Let $\boldsymbol{s} = [s_1, \ldots, s_K]^T$ be the PUs state vector.

We assume that the SUs implement energy detectors for a time time duration τ, which is a common assumption. If w is the sensed bandwidth, the energy detector measurement can be modeled as the energy of $w\tau$ baseband complex signal samples taken in a time interval of duration τ. Let $y_n(i)$ denote the i-th signal sample taken by the n-th SU. The signal samples result from the addition of the signals corresponding to all active PUs and the thermal noise, i.e.,

$$y_n(i) = \sum_{m=1}^{K} s_m h_{mn} x_m(i) + v_n(i), \tag{20}$$

where h_{mn} denotes the complex channel attenuation from the m-th PU to the n-th SU, $x_m(i)$ is the (zero-mean) signal sample transmitted by the m-th PU, and $v_n(i)$ is the thermal noise sample at the n-th SU, modeled as zero-mean white circular complex Gaussian noise with variance N_0, assumed to be independent of $x_m(i)$. The channel is modeled as:

$$h_{mn} = \mathrm{PL}(\|\boldsymbol{r}_m^{\mathrm{PU}} - \boldsymbol{r}_n^{\mathrm{SU}}\|)\eta_{mn},$$

where $\|\cdot\|$ is the Euclidean distance, $\mathrm{PL}(d) = d^{-\alpha/2}$ is the path-loss amplitude attenuation for a distance d and a path-loss exponent α, and η_{mn} represents the shadowing-fading and the multi-path fading effects. We assume that the PUs and the SUs are static during the observation time interval, thus, η_{mn} is constant during that time. The energy detector of the n-th SU delivers the energy level normalized by the noise power spectral density N_0, that is:

$$z_n = \frac{1}{N_0} \sum_{i=1}^{w\tau} |y_n(i)|^2, \quad n = 1, \ldots, N. \tag{21}$$

At this point we need to compute the distribution of the energy vector $\boldsymbol{z} \equiv [z_1, \ldots, z_N]^T$. The PUs signal is naturally a random process, given that it is an information signal that the base station wants to communicate to its intended receivers. Thus, $x_m(i)$ are considered uncorrelated random variables.

On the other hand, the bandwidth-time product is typically $w\tau \gg 1$, therefore the central limit theorem (CLT) can be used to describe $\boldsymbol{z}$ as a Gaussian random vector. All these considerations produce

$$\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{z}}, \mathbf{C}_{\boldsymbol{z}}), \tag{22}$$

where the components of $\boldsymbol{\mu}_{\boldsymbol{z}}$ and $\mathbf{C}_{\boldsymbol{z}}$ are, respectively,

$$\begin{aligned}
(\boldsymbol{\mu}_{\boldsymbol{z}})_n &= \frac{w\tau}{N_0}\left((\boldsymbol{s} \circ \boldsymbol{E}_x)^T \boldsymbol{g}_n + N_0\right), \text{and} \\
(\mathbf{C}_{\boldsymbol{z}})_{n,n'} &= \frac{w\tau}{N_0^2}\{(\boldsymbol{s} \circ \boldsymbol{E}_x \circ \boldsymbol{E}_x \circ (\boldsymbol{\kappa}_x - 2\mathbf{1}))^T(\boldsymbol{g}_n \circ \boldsymbol{g}_{n'}) \\
&\quad + \left((\boldsymbol{s} \circ \boldsymbol{E}_x \circ \boldsymbol{h}_n^* \circ \boldsymbol{h}_{n'})(\boldsymbol{s} \circ \boldsymbol{E}_x \circ \boldsymbol{h}_n \circ \boldsymbol{h}_{n'}^*)^T\right)_+ \\
&\quad + (2N_0(\boldsymbol{s} \circ \boldsymbol{E}_x)^T \boldsymbol{g}_n + N_0^2)\delta_{nn'}\},
\end{aligned} \tag{23}$$

where $n, n' = 1 \ldots, N$, $\circ$ is the element-wise (Hadamard) product, $\boldsymbol{h}_n = [h_{1n}, \ldots, h_{Kn}]^T$, $\boldsymbol{g}_n = \boldsymbol{h}_n \circ \boldsymbol{h}_n^*$ is the vector channel power gain, $(\boldsymbol{E}_x)_m = \frac{1}{w\tau}\sum_{i=1}^{w\tau} \mathbb{E}(|x_m(i)|^2)$ is the average energy of the m-th PU transmitted signal, $(\boldsymbol{\kappa}_x)_m = \frac{1}{w\tau}\sum_{i=1}^{w\tau} \mathbb{E}(|x_m(i)|^4)/\mathbb{E}(|x_m(i)|^2)^2$ is the average kurtosis of the m-th PU transmitted signal, $\mathbf{1}$ is a vector of ones of dimension K, $(\mathbf{A})_+$ means sum all the elements of the matrix $\mathbf{A}$ and $\delta_{nn'}$ is the Kronecker delta (1 if $n = n'$ and 0 otherwise). Appendix B has the details on these computations. It is clear that when at least one transmitter is active, $\boldsymbol{s} \neq \mathbf{0}$, the covariance matrix is not diagonal and the measurements from different SUs are correlated. Next we consider L time intervals of duration τ, for a total observation time window of $L\tau$. The energy observations from all nodes at the l-th τ-window $\boldsymbol{z}_l \equiv [z_{1,l}, \ldots, z_{N,l}]^T$, $l = 1, \ldots, L$ are assumed to be iid drawn from (22).

Remark 2. In other works as in [34], the signal transmitted by the PUs $x_m(i)$ is considered deterministic (and unknown). Thus, the distribution of z_n conditioned to the PUs state vector $\boldsymbol{s} = \mathbf{s}$ is a non-central chi-squared distribution of $2w\tau$ degrees of freedom with a the non-centrality parameter given by the energy of first term of (20). However, in most scenarios found in practice, this energy varies between the time intervals of length τ. Thus, the non-centrality parameter is not constant and follows a given distribution. For this reason, we consider a more appropriate model that one presented previously. Additionally, the way we treat the transmitted signal here reveals that the measurements taken by different SUs (nodes) are correlated, something expected given that all nodes observes the same PUs signals.

B. Asymptotic performance

In this section we compute the parameters to evaluate analytically the asymptotic performance of both T_{G} and $T_{\text{L-MP}}$ statistics. Here we consider a general scenario where under the null hypothesis $\mathcal{H}_0$ the parameters of (22) are known and the user activation is given by $\boldsymbol{s} = \mathrm{s}_0$. Note that $\mathrm{s}_0 \neq \mathbf{0}$ models a case where some PUs are active but far away enough such that the channel can be considered clear to be used by the SUs. On the other hand, under $\mathcal{H}_1$ the user activation is $\boldsymbol{s} = \mathrm{s}_1$, and the parameters are unknown and have to be estimated to run any of the statistics. Thus, the GLR parameter test is:

$$\left\{ \begin{array}{ll} \mathcal{H}_0: & \boldsymbol{\theta} = \boldsymbol{\theta}^0 \\ \mathcal{H}_1: & \boldsymbol{\theta} \neq \boldsymbol{\theta}^0, \end{array} \right. \tag{24}$$

where $\boldsymbol{\theta} = [\boldsymbol{\mu}_z^T, \text{vech}(\mathbf{C}_z)^T]^T \in \mathbb{R}^M$, $\text{vech}(\cdot)$ is the operator that concatenates all but the supra diagonal elements of $\mathbf{C}_z$, and $\boldsymbol{\theta}^0$ and $\boldsymbol{\theta}^1$ are obtained evaluating $\boldsymbol{\theta}$ at $\boldsymbol{s} = \mathrm{s}_0$ and $\boldsymbol{s} = \mathrm{s}_1$, respectively.

The performance of the GLR test given by (2). The degrees of freedom of the central and non-central chi-2 distribution is the dimension of the vector $\boldsymbol{\theta}$: $M = N + \frac{1}{2}N(N+1)$, and the non-centrality parameter λ_g depends on the Fisher information matrix $\boldsymbol{i}(\boldsymbol{\theta}^0)$ (see (3)). This matrix is a block diagonal matrix, whose first $N \times N$ sub-matrix is $\mathbf{C}_z^{-1}$, and the second $\frac{1}{2}N(N+1) \times \frac{1}{2}N(N+1)$ sub-matrix is as follows. Let c_{ij} be the (i,j)-th element of $\mathbf{C}_z$. There exists a one-to-one mapping noted $(i,j) \leftrightarrow k$ such that $c_{ij} = \theta_k$, $i,j = 1, \ldots, N$, and $k = N+1, \ldots, N + \frac{1}{2}N(N+1)$. Using (19) we have that:

$$\frac{\partial \mathbf{C}_z}{\partial c_{ij}} = \left\{ \begin{array}{ll} \boldsymbol{e}_i \boldsymbol{e}_i^T & \text{if } i = j \\ \boldsymbol{e}_i \boldsymbol{e}_j^T + \boldsymbol{e}_j \boldsymbol{e}_i^T & \text{if } i \neq j, \end{array} \right.$$

where $\boldsymbol{e}_i$ is the canonical vector, i.e., its i-th component is 1, and the remaining components are zero. Then, using the mapping $(i',j') \leftrightarrow k'$, the (k,k')-th component of $\boldsymbol{i}(\boldsymbol{\theta}^0)$ is, $k, k' = N+1, \ldots, N + \frac{1}{2}N(N+1)$:

$$\begin{aligned} \tfrac{1}{2}\text{tr}\left(\mathbf{C}_z^{-1} \tfrac{\partial \mathbf{C}_z}{\partial c_{ij}} \mathbf{C}_z^{-1} \tfrac{\partial \mathbf{C}_z}{\partial c_{i'j'}} \right) &= (\mathbf{C}_z^{-1})_{ii'} (\mathbf{C}_z^{-1})_{jj'} (1 - \tfrac{1}{2}\delta_{ij}\delta_{i'j'}) \\ &\quad + (\mathbf{C}_z^{-1})_{ij'} (\mathbf{C}_z^{-1})_{i'j} (1 - \delta_{ij})(1 - \delta_{i'j'}) \end{aligned}$$

Finally, λ_g can be readily computed using (3).

In the case of the $T_{\text{L-MP}}$ statistic, the off-diagonal components of the covariance matrix are non-observable parameters for local estimation at each SU and the parameter test is in this case:

$$\left\{ \begin{array}{ll} \mathcal{H}_0: & \boldsymbol{\theta}^{\text{loc}} = \boldsymbol{\theta}^{\text{loc},0} \\ \mathcal{H}_1: & \boldsymbol{\theta}^{\text{loc}} \neq \boldsymbol{\theta}^{\text{loc},0}, \end{array} \right. \tag{25}$$

where[4] $\boldsymbol{\theta}^{\text{loc}} = [\boldsymbol{\mu}_z^T, \bar{\text{diag}}(\mathbf{C}_z)^T]^T \in \mathbb{R}^P$, $P = 2N$. $\boldsymbol{\theta}^{\text{loc},0} = [(\boldsymbol{\mu}_z^0)^T, \bar{\text{diag}}(\mathbf{C}_z^0)^T]^T$ and $\boldsymbol{\theta}^{\text{loc},1} = [(\boldsymbol{\mu}_z^1)^T, \bar{\text{diag}}(\mathbf{C}_z^1)^T]^T$ are obtained evaluating $\boldsymbol{\theta}^{\text{loc}}$ at $\boldsymbol{s} = \mathrm{s}_0$ and $\boldsymbol{s} = \mathrm{s}_1$, respectively.

To compute the asymptotic performance of $T_{\text{L-MP}}$ we first evaluate $\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},m}) \in \mathbb{R}^{2N \times 2N}$, $m = 0, 1$, from (8). As $p_k(z_k; \boldsymbol{\theta}_k^{\text{loc}}) = \mathcal{N}(\mu_k, c_{kk})$, $\boldsymbol{\theta}_k^{\text{loc}} = [\mu_k, c_{kk}]^T$.

[4] Actually, $\boldsymbol{\theta}^{\text{loc}}$ defined here is a permutation of that one defined in Lem. 1. This allows us a more compact notation.

Note that $\frac{\partial \log p_k(z_k; \boldsymbol{\theta}_k^{\text{loc}})}{\partial \mu_k} = \frac{z_k - \mu_k}{c_{kk}}$, and $\frac{\partial \log p_k(z_k; \boldsymbol{\theta}_k^{\text{loc}})}{\partial c_{kk}} = \frac{1}{2c_{kk}}\left(\frac{(z_k - \mu_k)^2}{c_{kk}} - 1\right)$. The (i,j)-th component of the first $N \times N$ block-matrix of $\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},m})$ is $\mathbb{E}_{\boldsymbol{\theta}^m}\left(\frac{z_i - \mu_i}{c_{ii}} \frac{z_j - \mu_j}{c_{jj}}\right) = \frac{c_{ij}}{c_{ii} c_{jj}}$. In matrix form, it is $\boldsymbol{i}_1 \equiv \text{diag}(\mathbf{C}_z)^{-1} \mathbf{C}_z \text{diag}(\mathbf{C}_z)^{-1}$. The (i,j)-th component of the second $N \times N$ block-matrix is $\mathbb{E}_{\boldsymbol{\theta}^m}\left(\frac{1}{2c_{ii}}\left(\frac{(z_i - \mu_i)^2}{c_{ii}} - 1\right) \frac{1}{2c_{jj}}\left(\frac{(z_j - \mu_j)^2}{c_{jj}} - 1\right)\right) = \frac{c_{ij}^2}{2c_{ii}^2 c_{jj}^2}$. In matrix form, it is $\frac{1}{2}\boldsymbol{i}_1 \circ \boldsymbol{i}_1$. Then, as $\mathbb{E}_{\boldsymbol{\theta}^m}\left(\frac{z_i - \mu_i}{c_{ii}}\left(\frac{(z_j - \mu_j)^2}{c_{jj}} - 1\right)\right) = 0$ for $i, j \in [1:N]$, it results that $\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},m}) = \text{blkdiag}(\boldsymbol{i}_1, \frac{1}{2}\boldsymbol{i}_1 \circ \boldsymbol{i}_1)$, where blkdiag generates a block-diagonal matrix with its arguments in the diagonal.

Now we compute the matrices $\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},m})$ and $\boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},m})$. As $\boldsymbol{\theta}_k^{\text{loc}}$ and $\boldsymbol{\theta}_j^{\text{loc}}$ do not have parameters in common for $k \neq j$, and we have the identity $-\mathbb{E}_{\boldsymbol{\theta}_k^{\text{loc},m}}\left(\frac{\partial^2 \log p_k(z_k; \boldsymbol{\theta}_k^{\text{loc}})}{\partial \boldsymbol{\theta}_k^{\text{loc}} \partial \boldsymbol{\theta}_k^{\text{loc}}}\right) = \mathbb{E}_{\boldsymbol{\theta}_k^{\text{loc},m}}\left(\frac{\partial \log p_k(z_k; \boldsymbol{\theta}_k^{\text{loc}})}{\partial \boldsymbol{\theta}_k^{\text{loc}}} \frac{\partial \log p_k(z_k; \boldsymbol{\theta}_k^{\text{loc}})^T}{\partial \boldsymbol{\theta}_k^{\text{loc}}}\right)$, we have that $\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc},m}) = \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc},m}) = \text{diag}(\tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc},m}))$. Finally, the parameters of the asymptotic distribution can be readily computed as stated in Lem. 1, in order to evaluate the miss-detection and the false alarm probability next.

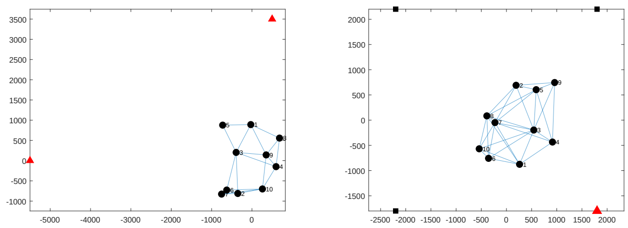
C. Numerical evaluation

We consider a system regulated by the wireless regional area network (WRAN) standard IEEE 802.22, where the consumer premise equipment (CPE) are the SUs and the TV stations are the PUs. This system employs OFDM for the TV stations, where each sub-carrier uses a given modulation scheme (QAM,PSK, among others). Here we will consider that the sample $x_m(i)$ are 64-QAM symbols independent through the indexes i and m. If R is the order of the modulation, the kurtosis of a R-QAM is $\kappa_{x_m} = \frac{7R-13}{5(R-1)}$. We fix $R = 64$ for all PUs which gives $\kappa_{x_m} \approx 1.381$.

We analyze two scenarios represented in Fig. 4. In both scenarios the SUs positions were uniformly randomly generated in an square area of 2000 m$\times$2000 m. The distance from each SU to the PUs was chosen such that the SU network is near to the edge of each PU cell, as illustrated in Fig. 3, where the problem is more challenging due to the low signal to interference and noise ratio (SINR) at each receiver. The common parameters for both scenarios are: $N = 10$, $L = 30$, $N_0 = -174$ dBm/MHz, $\alpha = 4$, $w = 5$ MHz, and $\tau = 10\mu$s. Then $w\tau = 50$ justifies the use of the CLT in (22). We assume $\eta_{mn} = 1$ for all m, n. The graph is generated as previously mentioned in Sec. V, with $|\mathcal{E}| = 20$ edges and $N_{it} = 20$ iterations are considered for the distributed implementation of $T_{\text{L-MP}}$.

1) Scenario I: In Fig. 4a we represent the first scenario, where the SUs (black circles) have to detect if any of the PUs (red triangles) is active. If the PUs are active, their transmit power is 100mW. In this scenario, the SUs want to detect if any of both PUs is transmitting ($\mathcal{H}_1$) or if both PUs are silent ($\mathcal{H}_0$). Thus, the state vector under $\mathcal{H}_0$ is $\mathrm{s}_0 = [0,0]^T$, and under $\mathcal{H}_1$, s_1 could be any vector in the set $\{[0,1]^T, [1,0]^T, [1,1]^T\}$.

In Figs. 5a and 5b we show two selected scatter plots of a realization of the network data under Scenario I, plotting z_2 vs



(a) Scenario I. (b) Scenario II.

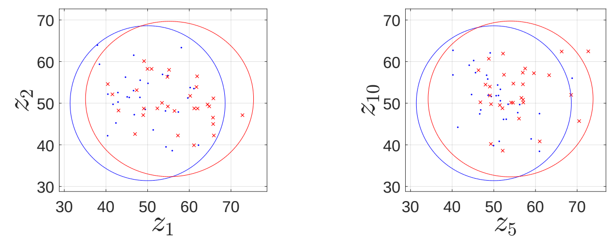
Figure 4: Location of the SUs and the PUs for each considered scenario. The blue lines represent the edges of the graph that models the sensor network.

z_1 and z_{10} vs z_5. The samples in blue are the realizations under $\mathcal{H}_0$, and the red ones are the realizations under $\mathcal{H}_1$ of the model (21). We also show the level curves of the limiting Gaussian distribution (22) under each hypothesis. Under $\mathcal{H}_0$, the sensor measurements are uncorrelated, and therefore the level curves are circumferences. Under $\mathcal{H}_1$, the sensor measurements are correlated. However, this is a second order effect for this scenario as shown by the covariance matrix $\mathbf{C}_z$ evaluated for $\boldsymbol{s} = [1,1]^T$ below (we only show its lower diagonal elements given that it is a symmetric matrix), and also visualized by the level curves, given that they are approximately circumferences. The vector mean for $\boldsymbol{s} = [1,1]^T$ is $\boldsymbol{\mu}_z = [55.3, 51.0, 52.2, 51.6, 54.0, 51.1, 51.1, 53.5, 52.2, 51.0]^T$.

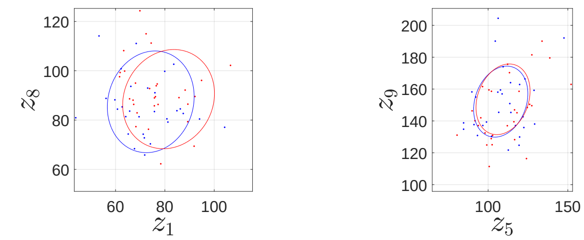
$$\mathbf{C}_z^1 = \begin{bmatrix} 60.77 & & & & & & & & & \\ 0.05 & 52.04 & & & & & & & & \\ 0.11 & 0.03 & 54.48 & & & & & & & \\ 0.08 & 0.02 & 0.04 & 53.22 & & & & & & \\ 0.20 & 0.04 & 0.09 & 0.06 & 58.22 & & & & & \\ 0.05 & 0.01 & 0.03 & 0.02 & 0.05 & 52.24 & & & & \\ 0.05 & 0.01 & 0.03 & 0.02 & 0.05 & 0.02 & 52.17 & & & \\ 0.16 & 0.03 & 0.07 & 0.05 & 0.13 & 0.03 & 0.03 & 57.09 & & \\ 0.11 & 0.02 & 0.05 & 0.03 & 0.09 & 0.02 & 0.02 & 0.07 & 54.40 & \\ 0.05 & 0.01 & 0.03 & 0.02 & 0.04 & 0.01 & 0.01 & 0.03 & 0.02 & 52.06 \end{bmatrix}.$$

In Fig. 6a, we plot the CROC curves for Scenario I for both statistics T_{G} and $T_{\text{L-MP}}$. As shown, the latter shows better performance than the GLR statistic. This can be interpreted as follows: as the correlation is weak under $\mathcal{H}_1$, the model mismatch when considering only the marginal distributions to build $T_{\text{L-MP}}$ is overcame by essentially the direction in which points $\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0$ respect to the matrix $\text{diag}(\boldsymbol{i}_1) = \text{diag}(\mathbf{C}_z^1)$, similar to what happens in the example in Sec. V (c.f. Figs. 1 and 2). This can be justified if we analyze the contribution of each term to the deflection coefficient $D_{\text{L-MP}}$ in (12). Its numerator is expressed as the sum of two positive terms, i.e., $\|\boldsymbol{\mu}_{\text{MP},1}\|^2$ and the trace, which is positive due to $\boldsymbol{\Sigma}_{\text{MP},1} - \boldsymbol{\Sigma}_{\text{MP},0}$ is positive definite. $\|\boldsymbol{\mu}_{\text{MP},1}\|^2$ can additionally be decomposed in two positive terms, one of which is $L(\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0)^T \text{diag}(\mathbf{C}_z^1)^{-1}(\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0)$. It results that for this scenario, this term contributes to the 96.5% of the total value of $D_{\text{L-MP}} = 6.18$, dominating the performance of the statistic.

In the case of the GLR, the non-central parameter λ_g also can be decomposed in two terms, one of which is



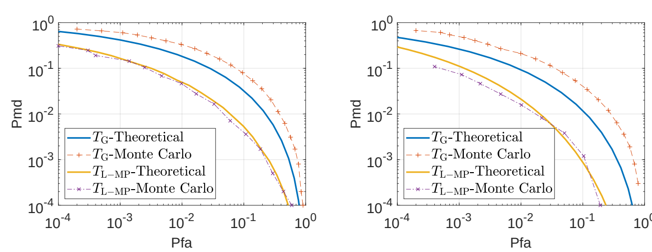
(a) Samples of SU #2 vs SU #1. (b) Samples of SU #10 vs SU #5.



(c) Samples of SU #8 vs SU #1. (d) Samples of SU #9 vs SU #5.

Figure 5: Selected scatter plots for the measurements of a sensor against other in the network. Samples in blue (red) are drawn under $\mathcal{H}_0$ ($\mathcal{H}_1$). (a) and (b) correspond to Scenario I: the state vector under $\mathcal{H}_1$ is $\boldsymbol{s} = [1,1]^T$ for the plotted data. (c) and (d) correspond to Scenario II: the state vector under $\mathcal{H}_0$ ($\mathcal{H}_1$) is $\boldsymbol{s}_0 = [1,1,1,0]^T$ ($\boldsymbol{s}_1 = [1,1,1,1]^T$). The x and y-axes are in proportion $1:1$.

$L/(w\tau)\|\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0\|^2$ (note that $\mathbf{C}_z^0 = w\tau\mathbf{I}_N$), and contributes to the 96.0% of D_{G}. However, $D_{\text{G}} = 3.98 < D_{\text{L-MP}}$ and $T_{\text{L-MP}}$ has better performance. See also that the asymptotic performance of $T_{\text{L-MP}}$ shows good agreement with the Monte Carlo estimation the CROC for finite data size ($L = 30$). In the case of T_{G}, the asymptotic performance predicts a better result than the obtained for finite data size.



(a) CROC for Scenario I. (b) CROC for Scenario II.

Figure 6: CROCs for both scenarios.

2) Scenario II: In Fig. 4b we represent the second scenario, where the SUs (black circles) have to detect if the PU (red triangle) is active or not, in presence of interferer transmitters (black squares), which are active PUs assumed to be further away than the PU to be detected, and therefore, the SUs are out of their corresponding protected region. Thus, the SUs could transmit in the same spectrum band without producing harmful interference for their intended receivers. We consider that if the PU to be detected is active, its transmit power is 30mW, and that the power of the interferer transmitters is 300mW. We assume that the distribution parameters under $\mathcal{H}_0$ are known. This could be possible in a scenario with wide-sense stationary interference, where the mean and covariance matrix under $\mathcal{H}_0$ are precomputed. In Figs. 5c and 5d we show the scatter plots for samples generated in Scenario II and the level curves for the corresponding limiting Gaussian distribution. The vector mean and the covariance matrix under $\mathcal{H}_1$ are, respectively, $\boldsymbol{\mu}_z = [81.5, 99.8, 80.1, 81.7, 109.4, 98.9, 85.4, 88.5, 153.6, 104.1]^T$ and

$$\mathbf{C}_z^1 = \begin{bmatrix} 128.6 \\ 15.4 & 184.7 \\ 13.6 & 20.0 & 124.6 \\ 14.3 & 17.8 & 14.6 & 129.5 \\ 15.9 & 37.5 & 21.6 & 19.9 & 211.2 \\ 19.5 & 19.4 & 16.1 & 14.8 & 18.6 & 177.7 \\ 15.4 & 22.4 & 15.4 & 13.8 & 22.1 & 22.1 & 140.6 \\ 16.0 & 24.2 & 16.1 & 14.1 & 23.4 & 23.5 & 21.5 & 150.5 \\ 19.3 & 53.2 & 29.3 & 27.3 & 64.1 & 20.9 & 27.2 & 28.4 & 361.4 \\ 20.6 & 21.6 & 17.2 & 15.3 & 20.3 & 32.9 & 24.6 & 26.5 & 22.6 & 194.7 \end{bmatrix}.$$

From both, the covariance matrix and the level curves (ellipses with their major and minor-axis rotated in the Cartesian coordinates), it is evident that the correlation for this scenario is not negligible. In Fig. 6b, it can be seen that the performance of $T_{\text{L-MP}}$ is better than T_{G}. Similar to the previous scenario, the better performance of $T_{\text{L-MP}}$ can be explained in terms of the direction of $\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0$ with respect to the corresponding Fisher sub-matrix that favors $T_{\text{L-MP}}$ more than T_{G}. For this scenario, the term $L(\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0)^T \text{diag}(\mathbf{C}_z^1)^{-1}(\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0)$ of the numerator of $D_{\text{L-MP}}$ contributes to the 95.7% of its total value (6.41). In the case of D_{G}, the term $L(\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0)^T(\mathbf{C}_z^0)^{-1}(\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0)$ of λ_g in the numerator of D_{G} contributes to the 95.3% of its total value (4.66). In both statistics, the term that includes $\boldsymbol{\mu}_z^1 - \boldsymbol{\mu}_z^0$ dominates their performances. As in the previous curves, the asymptotic results show fair agreement with the Monte Carlo estimations for finite data size ($L = 30$) for both statistics.

VII. Conclusions

We have proposed and characterized the asymptotic performance of a fully distributed detection algorithm for a WSN with statistically dependent observations. The proposed statistic successfully reduce the amount of local transmissions to build the statistic estimating the unknown parameters locally, and thus, saving valuable resources as energy, bandwidth or delay. The analytical and simulated results show us that in some scenarios, the advantages of using $T_{\text{L-MP}}$ is twofold: it performs better and its implementation is much more efficient in terms of network resources. We also derived a statistic model for the energy detector in the context of spectrum sensing for cognitive radio networks. This model shows that the observations are in general spatially correlated and with different levels of correlation, depending on the specific situation. We evaluated the proposed statistic for two relevant scenarios in practice and found that it outperforms the classical GLR test. This model results show that implementing distributed detection algorithms as if the measurements were spatially statistically independent, even when they are not, not only do not introduce penalties but can also be beneficial. Moreover, it provides theoretical support for many WSN algorithms found in the literature that discard the statistical spatial dependence of the measurements when designing a distributed detection algorithm. In general, the tools developed in this paper allow us to quantitatively characterize the improvement/penalty introduced by discarding the statistical dependence of the observations in both, estimating the parameter locally and implementing a statistic using only the marginal pdfs.

Appendix A
Proof of Lemma 1

The proof of the lemma can be splitted in two parts: the asymptotic distribution of the local MLE, and the asymptotic distribution of the statistic $T_{\text{L-MP}}$. We will refer to $\boldsymbol{\theta}^{\text{loc}*}$ as the true parameter of the distribution ($\boldsymbol{\theta}^{\text{loc}*} = \boldsymbol{\theta}^{\text{loc},0}$ under $\mathcal{H}_0$ and $\boldsymbol{\theta}^{\text{loc}*} = \boldsymbol{\theta}^{\text{loc},1}$ under $\mathcal{H}_1$). Before starting with the proof we need to present the following result:

Theorem 1 (Mean value theorem)**.** [35, Th. 12.9] Let B be an open subset of $\mathbb{R}^P$ and assume that $\boldsymbol{f} : B \to \mathbb{R}^P$ is differentiable at each point of B. Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two points in B such that the segment $S(\boldsymbol{x}, \boldsymbol{y}) = \{t\boldsymbol{x} + (1-t)\boldsymbol{y} : t \in [0,1]\} \in B$. Then for every vector $\boldsymbol{a}$ in $\mathbb{R}^P$ there is a point $\boldsymbol{w} \in S(\boldsymbol{x}, \boldsymbol{y})$ such that

$$\boldsymbol{a}^T(\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})) = \boldsymbol{a}^T J(\boldsymbol{w})(\boldsymbol{y} - \boldsymbol{x}), \tag{26}$$

where $J(\boldsymbol{w})$ is the Jacobian matrix of $\boldsymbol{f}$ evaluated in $\boldsymbol{w}$, i.e., $[J]_{i,j} = \frac{\partial f_i}{\partial x_j}$, where $\{f_i\}$ are the components of $\boldsymbol{f}$. It is important to remark that $\boldsymbol{w}$ depends on $\boldsymbol{a}$.

A. Asymptotic distribution of the local MLE

Given that the hypothesis $\mathcal{H}_0$ is simple ($\boldsymbol{\theta}^{\text{loc},0}$ is a known vector), we have to consider the unrestricted MLE. Let $\boldsymbol{\psi}(\boldsymbol{z}_l; \boldsymbol{\theta}^{\text{loc}}) \in \mathbb{R}^P$ be

$$= \left[\frac{\partial \log p_1(z_1(l); \boldsymbol{\theta}_1^{\text{loc}})^T}{\partial \boldsymbol{\theta}_1^{\text{loc}}}, \ldots, \frac{\partial \log p_N(z_N(l); \boldsymbol{\theta}_N^{\text{loc}})^T}{\partial \boldsymbol{\theta}_N^{\text{loc}}}\right]^T.$$

By definition, the MLE must satisfy

$$\tfrac{1}{L}\textstyle\sum_{l=1}^{L} \boldsymbol{\psi}(\boldsymbol{z}_l; \hat{\boldsymbol{\theta}}^{\text{loc}}) = \mathbf{0}. \tag{27}$$

Consider Theorem 1 with $\boldsymbol{f}(\boldsymbol{\theta}^{\text{loc}}) = \frac{1}{L}\sum_{l=1}^{L} \boldsymbol{\psi}(\boldsymbol{z}_l; \boldsymbol{\theta}^{\text{loc}})$, $\boldsymbol{x} = \boldsymbol{\theta}^{\text{loc}*}$ and $\boldsymbol{y} = \hat{\boldsymbol{\theta}}^{\text{loc}}$, then we have

$$\begin{aligned} &\boldsymbol{a}^T \left(\tfrac{1}{L}\textstyle\sum_{l=1}^{L} \boldsymbol{\psi}(\boldsymbol{z}_l; \hat{\boldsymbol{\theta}}^{\text{loc}}) - \tfrac{1}{L}\sum_{l=1}^{L} \boldsymbol{\psi}(\boldsymbol{z}_l; \boldsymbol{\theta}^{\text{loc}*})\right) \\ &\qquad = \boldsymbol{a}^T J(\boldsymbol{w}^L)(\hat{\boldsymbol{\theta}}^{\text{loc}} - \boldsymbol{\theta}^{\text{loc}*}), \quad \forall \boldsymbol{a} \in \mathbb{R}^N \end{aligned}$$

where $\boldsymbol{w}^L$ belongs to the segment $S(\hat{\boldsymbol{\theta}}^{\text{loc}}, \boldsymbol{\theta}^{\text{loc}*})$ and depends on $\boldsymbol{a}$. Using (27) in the previous equation produces

$$\begin{aligned} &\boldsymbol{a}^T \left\{ \tfrac{1}{\sqrt{L}} J(\boldsymbol{w}^L) \left(J(\boldsymbol{w}^L)^{-1} \tfrac{1}{\sqrt{L}} \textstyle\sum_{l=1}^{L} \boldsymbol{\psi}(\boldsymbol{z}_l; \boldsymbol{\theta}^{\text{loc}*}) \right.\right. \\ &\qquad \left.\left. + \sqrt{L}(\hat{\boldsymbol{\theta}}^{\text{loc}} - \boldsymbol{\theta}^{\text{loc}*}) \right) \right\} = 0, \quad \forall \boldsymbol{a}. \end{aligned} \tag{28}$$

By consistency of the estimator $\hat{\boldsymbol{\theta}}^{\text{loc}}$, the segment $S(\hat{\boldsymbol{\theta}}^{\text{loc}}, \boldsymbol{\theta}^{\text{loc}*})$ becomes the point $\boldsymbol{\theta}^{\text{loc}*}$ and $\boldsymbol{w}^L \xrightarrow{p} \boldsymbol{\theta}^{\text{loc}*}$ as $L \to \infty$. Thus, the expression inside the parenthesis in (28) becomes independent of $\boldsymbol{a}$ as $L \to \infty$, and therefore, it must converge in probability to $\mathbf{0}$. Then, using the continuity of the second-order partial derivatives of the log-likelihood function, we apply the Continuous Mapping Theorem (CMT) [36] to obtain $J(\boldsymbol{w}^L) \xrightarrow{p} -\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc}*})$, where $\boldsymbol{j}(\boldsymbol{\theta}^{\text{loc}*})$ is defined in (9). Additionally, by the multivariate central limit theorem (CLT)

$$\tfrac{1}{\sqrt{L}} \textstyle\sum_{l=1}^{L} \boldsymbol{\psi}(\boldsymbol{z}_l; \boldsymbol{\theta}^{\text{loc}*}) \overset{a}{\sim} \mathcal{N}(\mathbf{0}, \tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc}*})),$$

where the mean of the Gaussian distribution is $\mathbf{0}$ by the first assumption and its covariance matrix is defined in (8). Finally,

$$\begin{aligned}\sqrt{L}(\hat{\boldsymbol{\theta}}^{\text{loc}} - \boldsymbol{\theta}^{\text{loc}*}) &\overset{a}{\sim} -J(\boldsymbol{w}^L)^{-1} \tfrac{1}{\sqrt{L}} \textstyle\sum_{l=1}^{L} \boldsymbol{\psi}(\boldsymbol{z}_l; \boldsymbol{\theta}^{\text{loc}*}) \\ &\overset{a}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{j}(\boldsymbol{\theta}^{\text{loc}*})^{-1} \tilde{\boldsymbol{i}}(\boldsymbol{\theta}^{\text{loc}*}) \boldsymbol{j}(\boldsymbol{\theta}^{\text{loc}*})^{-1})\end{aligned}$$

Solving for $\hat{\boldsymbol{\theta}}^{\text{loc}}$ we obtain the first result of the lemma in (7).

B. Asymptotic distribution of $T_{\text{L-MP}}$

To prove this lemma we need fundamentally to show that the following factorization is valid:

$$\frac{\partial \log p_{\text{MP}}(\boldsymbol{z}_l; \boldsymbol{\theta}^{\text{loc}*})}{\partial \boldsymbol{\theta}^{\text{loc}}} = \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^{\text{loc}*})(\hat{\boldsymbol{\theta}}^{\text{loc}} - \boldsymbol{\theta}^{\text{loc}*}), \tag{29}$$

where $\hat{\boldsymbol{\theta}}^{\text{loc}}$ is the local MLE estimator. Note that this is a similar factorization to that one found for estimators attaining the Cramer-Rao bound but with the true joint pdf replaced by $p_{\text{MP}}(\cdot, \boldsymbol{\theta}^{\text{loc}*})$. We show next that this equation is valid even when $\hat{\boldsymbol{\theta}}^{\text{loc}}$ does not attain, in general, the Cramer-Rao bound, asymptotically achieved by the global MLE.

Hereafter, where there is no ambiguity, we will drop the supra index loc from $\hat{\boldsymbol{\theta}}^{\text{loc}}$ and $\boldsymbol{\theta}^{\text{loc}*}$ and call them $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$, respectively, in order to shrink the size of the equations. Given that the local MLE $\hat{\boldsymbol{\theta}}$ is consistent, $\boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}^{\text{loc}*}}(\hat{\boldsymbol{\theta}}) = \int \hat{\boldsymbol{\theta}} p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta}) d\boldsymbol{z}$ is asymptotically satisfied when $L \to \infty$. Then

$$\begin{aligned}\frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\theta}} = \mathbf{I}_P &= \int \hat{\boldsymbol{\theta}} \frac{\partial p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} d\boldsymbol{z} \\ &= \int \hat{\boldsymbol{\theta}} \frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta}) d\boldsymbol{z} \\ &= \int (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta}) d\boldsymbol{z}\end{aligned}$$

where in the last equality we used the second assumption. Let $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^P$ arbitrary vectors. After pre- and post-multiplication of the last equation by $\boldsymbol{a}^T$ and $\boldsymbol{b}$, respectively, we have:

$$\boldsymbol{a}^T \boldsymbol{b} = \int \boldsymbol{a}^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \boldsymbol{b} \; p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta}) d\boldsymbol{z}. \tag{30}$$

Then, we need the following Cauchy-Schwarz inequality

$$\left[\int w(\boldsymbol{z}) g(\boldsymbol{z}) h(\boldsymbol{z}) d\boldsymbol{z}\right]^2 \leq \int w(\boldsymbol{z}) g^2(\boldsymbol{z}) d\boldsymbol{z} \int w(\boldsymbol{z}) h^2(\boldsymbol{z}) d\boldsymbol{z},$$

where $g(\boldsymbol{z})$ and $h(\boldsymbol{z})$ are arbitrary scalar functions and $w(\boldsymbol{z}) \geq 0, \forall \boldsymbol{z}$, and the equality holds if and only if $g(\boldsymbol{z}) = c\, h(\boldsymbol{z})$. Let $w(\boldsymbol{z}) = p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})$, $g(\boldsymbol{z}) = \boldsymbol{a}^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ and $h(\boldsymbol{z}) = \frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \boldsymbol{b}$ and apply the Cauchy-Schwarz inequality to (30) to obtain

$$\begin{aligned}(\boldsymbol{a}^T \boldsymbol{b})^2 \leq& \int \boldsymbol{a}^T (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \boldsymbol{a} \; p_{\text{MP}}(\boldsymbol{z}, \boldsymbol{\theta}) d\boldsymbol{z} &(31)\\ &\times \int \boldsymbol{b} \frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \boldsymbol{b} \; p_{\text{MP}}(\boldsymbol{z}, \boldsymbol{\theta}) d\boldsymbol{z} \\ =& \boldsymbol{a}^T \mathbf{C}_{\text{MP}} \boldsymbol{a} \; \boldsymbol{b}^T \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}) \boldsymbol{b} &(32)\\ =& \boldsymbol{a}^T \mathbf{C}_{\text{MP}} \boldsymbol{a} \; \boldsymbol{a}^T \boldsymbol{b} &(33)\end{aligned}$$

where in (32) we defined $\mathbf{C}_{\text{MP}} = \mathbb{E}_{\boldsymbol{\theta}^{\text{loc}*}}((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T)$, and in (33) we select $\boldsymbol{b} = \boldsymbol{i}_{\text{MP}}^{-1}(\boldsymbol{\theta}) \boldsymbol{a}$. As a consequence of assumption iii), $\boldsymbol{i}_{\text{MP}}^{-1}(\boldsymbol{\theta})$ is positive-definite, $\boldsymbol{a}\boldsymbol{b}^T \geq 0$, and from (33) we have that $\boldsymbol{a}^T(\mathbf{C}_{\text{MP}} - \boldsymbol{i}_{\text{MP}}^{-1}(\boldsymbol{\theta}))\boldsymbol{a} \geq 0, \forall \boldsymbol{a}$. Now, the equality in (31) holds if and only if $\boldsymbol{a}^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = c\, \boldsymbol{a}^T \boldsymbol{i}_{\text{MP}}^{-1}(\boldsymbol{\theta}) \frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. As this is satisfied $\forall \boldsymbol{a}$, we finally obtain (29) given that the constant c is proved to be 1.

The second part of the proof is as follows: by consistency of $\hat{\boldsymbol{\theta}}$, (29) is also satisfied with $\hat{\boldsymbol{\theta}}$ instead of $\boldsymbol{\theta}$ when $L \to \infty$. Then, using a first-order Taylor expansion of $\boldsymbol{i}(\boldsymbol{\theta})$ around $\hat{\boldsymbol{\theta}}$ and discarding the second order terms as $L \to \infty$, we have

$$\frac{\partial \log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = L \boldsymbol{i}_{\text{MP}}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}). \tag{34}$$

Integrating this equation with respect to $\boldsymbol{\theta}$:

$$\log p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta}) = -\tfrac{L}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \boldsymbol{i}_{\text{MP}}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + c(\hat{\boldsymbol{\theta}}), \tag{35}$$

where the integration constant must be $c(\hat{\boldsymbol{\theta}}) = \log p_{\text{MP}}(\boldsymbol{z}; \hat{\boldsymbol{\theta}})$ given that (35) is satisfied asymptotically by the consistence of $\hat{\boldsymbol{\theta}}$ when $L \to \infty$. Therefore,

$$p_{\text{MP}}(\boldsymbol{z}; \boldsymbol{\theta}) = p_{\text{MP}}(\boldsymbol{z}; \hat{\boldsymbol{\theta}}) e^{-L\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \boldsymbol{i}_{\text{MP}}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}.$$

Using the previous equation in the expression of $T_{\text{L-MP}}$ and replacing $\boldsymbol{\theta}$ by $\boldsymbol{\theta}^0$, we obtain

$$T_{\text{L-MP}}(\boldsymbol{z}) = \frac{p_{\text{MP}}(\boldsymbol{z}; \hat{\boldsymbol{\theta}})}{p_{\text{MP}}\left(\boldsymbol{z}; \boldsymbol{\theta}^0\right)} = e^{L\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\mathbf{0}})^T \boldsymbol{i}_{\text{MP}}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\mathbf{0}})},$$

or

$$2 \log T_{\text{L-MP}}(\boldsymbol{z}) = L(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\mathbf{0}})^T \boldsymbol{i}_{\text{MP}}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\mathbf{0}})$$

Using again the CMT and the continuity of the second-order partial derivatives, the following is satisfied when $L \to \infty$:

$$\boldsymbol{i}_{\text{MP}}(\hat{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) = \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^i)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \text{ under } \mathcal{H}_i, i = 0, 1.$$

Finally, we have

$$\begin{aligned}2 \log T_{\text{L-MP}}(\boldsymbol{z}) &= L(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)^T \boldsymbol{i}_{\text{MP}}(\boldsymbol{\theta}^i)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) &(36)\\ &= \|\sqrt{L} \boldsymbol{i}_{\text{MP}}^{\frac{1}{2}}(\boldsymbol{\theta}^i)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)\|^2 &(37)\\ &\overset{a}{\sim} f_P(\boldsymbol{\mu}_{\text{MP},\boldsymbol{i}}, \boldsymbol{\Sigma}_{\text{MP},\boldsymbol{i}}) \text{ under } \mathcal{H}_i. &(38)\end{aligned}$$

where the parameters $\boldsymbol{\mu}_{\text{MP},i}$ and $\boldsymbol{\Sigma}_{\text{MP},i}$ of the asymptotic distribution f_P are the mean and the covariance matrix of the multivariate Gaussian vector inside the square norm. They are obtained using (7) and are presented in the lemma.

Appendix B
Moments of the SS model

Equations (20) and (21) can be succinctly written as

$$\boldsymbol{y}_n = \mathbf{X}\tilde{\boldsymbol{h}}_n + \boldsymbol{v}_n \tag{39}$$

$$z_n = \frac{1}{N_0}\|\boldsymbol{y}_n\|^2 \tag{40}$$

where $\boldsymbol{x}_m = [x_m(1), \ldots, x_m(w\tau)]^T$, $\mathbf{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]$, $\tilde{\boldsymbol{h}}_n = \boldsymbol{s} \circ \boldsymbol{h}_n$, and $\boldsymbol{v}_n = [v_n(1), \ldots, v_n(w\tau)]^T$. Then,

$$\begin{aligned}\mu_{z_n} = \mathbb{E}(z_n) &= \frac{1}{N_0}\mathbb{E}(\|\mathbf{X}\tilde{\boldsymbol{h}}_n\|^2 + \|\boldsymbol{v}_n\|^2 + 2\Re(\tilde{\boldsymbol{h}}_n^H \mathbf{X}^H \boldsymbol{v}_n)) \\ &= \frac{w\tau}{N_0}\left((\boldsymbol{s} \circ \boldsymbol{E}_x)^T \boldsymbol{g}_n + N_0\right), \end{aligned} \tag{41}$$

where we have used the fact that $\mathbb{E}(\mathbf{X}^H\mathbf{X}) = w\tau \mathrm{diag}(\boldsymbol{E}_x)$ given that the signals from different PUs are zero-mean, uncorrelated and independent of $\boldsymbol{v}_n$. On the other hand, the covariance of z_n and $z_{n'}$, with $n, n' = 1, \ldots, N$, is

$$\mathrm{Cov}(z_n, z_{n'}) = \frac{1}{N_0^2}\mathbb{E}\big[(\alpha_n + \beta_n + \gamma_n)(\alpha_{n'} + \beta_{n'} + \gamma_{n'})\big]$$

where $\alpha_n = \|\mathbf{X}\tilde{\boldsymbol{h}}_n\|^2 - w\tau(\boldsymbol{s} \circ \boldsymbol{E}_x)^T \boldsymbol{g}_n$, $\beta_n = \|\boldsymbol{v}_n\|^2 - w\tau N_0$ and $\gamma_n = 2\Re(\tilde{\boldsymbol{h}}_n^H \mathbf{X}^H \boldsymbol{v}_n)$ are all zero-mean random variables. Then,

$$\begin{aligned}\mathbb{E}(\alpha_n\alpha_{n'}) &= w\tau\Big(\textstyle\sum_{m=1}^K s_m |h_{mn}|^2 |h_{mn'}|^2 E_{x_m}^2(\kappa_{x_m} - 2) \\ &\quad + \textstyle\sum_{m,m'=1}^K s_m s_{m'} h_{mn}^* h_{m'n} h_{m'n'}^* h_{mn'} E_{x_m} E_{x_{m'}}\Big), \\ \mathbb{E}(\beta_n\beta_{n'}) &= w\tau N_0^2 \delta_{nn'}, \text{and} \\ \mathbb{E}(\gamma_n\gamma_{n'}) &= w\tau 2N_0(\boldsymbol{s} \circ \boldsymbol{E}_x)^T \boldsymbol{g}_n \delta_{nn'}.\end{aligned}$$

The rest of the terms are zero: $\mathbb{E}(\alpha_n\beta_{n'}) = 0$ because α_n and $\beta_{n'}$ are zero-mean and independent random variables; $\mathbb{E}(\alpha_n\gamma_{n'}) = 0$ because $\mathbf{X}$ and $\boldsymbol{v}_{n'}$ are independent and $\boldsymbol{v}_{n'}$ is a zero mean random vector; and $\mathbb{E}(\beta_n\gamma_{n'}) = 0$ because $\mathbf{X}$ and $\boldsymbol{v}_{n'}$ are independent and $\mathbf{X}$ is a zero-mean random matrix. Finally, using vector notation we arrive to (23).

References

[1] Faisal Karim Shaikh and Sherali Zeadally. Energy harvesting in wireless sensor networks: A comprehensive review. *Renewable and Sustainable Energy Reviews*, 55:1041–1054, 2016.

[2] Bushra Rashid and Mubashir Husain Rehmani. Applications of wireless sensor networks for urban areas: A survey. *Journal of Network and Computer Applications*, 60:192–219, 2016.

[3] F. Li, M. Valero, Y. Cheng, T. Bai, and W. Song. Distributed sensor networks based shallow subsurface imaging and infrastructure monitoring. *IEEE Transactions on Signal and Information Processing over Networks*, 6:241–250, 2020.

[4] V. Gupta and S. De. Collaborative multi-sensing in energy harvesting wireless sensor networks. *IEEE Transactions on Signal and Information Processing over Networks*, 6:426–441, 2020.

[5] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami. Internet of Things (IoT): A vision, architectural elements, and future directions. *Future generation computer systems*, 29(7):1645–1660, 2013.

[6] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, Mohammed Aledhari, and Moussa Ayyash. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Communications Surveys & Tutorials*, 17(4):2347–2376, 2015.

[7] Sundeep Prabhakar Chepuri and Geert Leus. Sparse sensing for distributed detection. *IEEE Transactions on Signal Processing*, 64(6):1446–1460, 2016.

[8] D. Ciuonzo and P. S. Rossi. Distributed detection of a non-cooperative target via generalized locally-optimum approaches. *Information Fusion*, 36:261–274, 2017.

[9] Sami Ahmed Aldalahmeh, Saleh O Al-Jazzar, Des McLernon, Syed Ali Raza Zaidi, and Mounir Ghogho. Fusion rules for distributed detection in clustered wireless sensor networks with imperfect channels. *IEEE Transactions on Signal and Information Processing over Networks*, 5(3):585–597, 2019.

[10] Ali H Sayed et al. Adaptation, learning, and optimization over networks. *Foundations and Trends® in Machine Learning*, 7(4-5):311–801, 2014.

[11] S. Al-Sayed, J. Plata-Chaves, M. Muma, M. Moonen, and A.M. Zoubir. Node-specific diffusion lms-based distributed detection over adaptive networks. *IEEE Trans. on Sig. Proc.*, 66(3):682–697, 2018.

[12] Ali H Sayed, Sheng-Yuan Tu, Jianshu Chen, Xiaochuan Zhao, and Zaid J Towfic. Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior. *IEEE Signal Processing Magazine*, 30(3):155–171, 2013.

[13] D. Ciuonzo and P.S. Rossi. Decision fusion with unknown sensor detection probability. *IEEE Signal Processing Letters*, 21(2):208–212, 2014.

[14] S Hamed Hamed and Ali Peiravi. Reliable distributed detection in multi-hop clustered wireless sensor networks. *IET Signal Processing*, 6(8):743–750, 2012.

[15] Soummya Kar, Ravi Tandon, H Vincent Poor, and Shuguang Cui. Distributed detection in noisy sensor networks. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 2856–2860. IEEE, 2011.

[16] Shang Li and Xiaodong Wang. Fully distributed sequential hypothesis testing: Algorithms and asymptotic analyses. *IEEE Transactions on Information Theory*, 64(4):2742–2758, 2018.

[17] Elias Drakopoulos and C-C Lee. Optimum multisensor fusion of correlated local decisions. *IEEE Transactions on Aerospace and Electronic Systems*, 27(4):593–606, 1991.

[18] Leandro A Villas, Azzedine Boukerche, Horacio ABF De Oliveira, Regina B De Araujo, and Antonio AF Loureiro. A spatial correlation aware algorithm to perform efficient data collection in wireless sensor networks. *Ad Hoc Networks*, 12:69–85, 2014.

[19] L. Wang, J. D. Reiss, and A. Cavallaro. Over-determined source separation and localization using distributed microphones. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(9):1573–1588, Sep 2016.

[20] Siguang Chen, Shujun Zhang, Xiaoyao Zheng, and Xiukai Ruan. Layered adaptive compression design for efficient data collection in industrial wireless sensor networks. *Journal of Network and Computer Applications*, 129:37–45, 2019.

[21] Jarmo Lunden, Visa Koivunen, and H Vincent Poor. Spectrum exploration and exploitation for cognitive radio: Recent advances. *IEEE signal processing magazine*, 32(3):123–140, 2015.

[22] S. Kay. *Fundamentals of Statistical Signal Processing, Volume II: Detection Theory*. Prentice-Hall, 1st ed. edition, 1998.

[23] Bernard C Levy. *Principles of signal detection and parameter estimation*. Springer, 2008.

[24] Steven M Kay. *Fundamentals of statistical signal processing, Volume I: Estimation theory*. Prentice-Hall, Inc., 1993.

[25] Lin Xiao and Stephen Boyd. Fast linear iterations for distributed averaging. *Systems & Control Letters*, 53(1):65–78, 2004.

[26] B. Ying, K. Yuan, and A. H. Sayed. Dynamic average diffusion with

randomized coordinate updates. *IEEE Transactions on Signal and Information Processing over Networks*, 5(4):753–767, 2019.

[27] Robert Lugannani and Stephen Rice. Saddle point approximation for the distribution of the sum of independent random variables. *Advances in applied probability*, pages 475–490, 1980.

[28] H.Y. Zhou, D.Y. Luo, Y. Gao, and D.C. Zuo. Modeling of node energy consumption for wireless sensor networks. *Wireless Sensor Network*, 3(1):18, 2011.

[29] S. Sengupta, S. Das, MD Nasir, and B.K. Panigrahi. Multi-objective node deployment in wsns: In search of an optimal trade-off among coverage, lifetime, energy consumption, and connectivity. *Engineering Applications of Artificial Intelligence*, 26(1):405–416, 2013.

[30] Ami Wiesel and Alfred O Hero. Distributed covariance estimation in gaussian graphical models. *IEEE Transactions on Signal Processing*, 60(1):211–220, 2011.

[31] J. Lunden, V. Koivunen, and H.V. Poor. Spectrum exploration and exploitation for cognitive radio: Recent advances. *Signal Processing Magazine, IEEE*, 32(3):123–140, May 2015.

[32] Tevfik Yücek and Hüseyin Arslan. A survey of spectrum sensing algorithms for cognitive radio applications. *Communications Surveys & Tutorials, IEEE*, 11(1):116–130, 2009.

[33] Ruilong Deng, Jiming Chen, Chau Yuen, Peng Cheng, and Youxian Sun. Energy-efficient cooperative spectrum sensing by optimal scheduling in sensor-aided cognitive radio networks. *Vehicular Technology, IEEE Transactions on*, 61(2):716–725, 2012.

[34] Karaputugala Madushan Thilina, Kae Won Choi, Nazmus Saquib, and Ekram Hossain. Machine learning techniques for cooperative spectrum sensing in cognitive radio networks. *IEEE Journal on selected areas in communications*, 31(11):2209–2221, 2013.

[35] Tom M. Apostol. *Mathematical analysis*, volume 2. Addison-Wesley Reading, MA, 1974.

[36] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, Jun 2000.