Signal Processing 239 (2026) 110238

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Blind learning of the optimal fusion rule in wireless sensor networks ☆

J. Perez [a],*, I. Santamaria [a], A. Pagés-Zamora [b]

[a] Communications Engineering Department, University of Cantabria, Plaza de la Ciencia s/n, Santander, 39005, Spain
[b] SPCOM Group, Universitat Politècnica de Catalunya-Barcelona Tech (UPC), C/Jordi Girona 31, Barcelona, 08034, Spain

ARTICLE INFO

Keywords:
Decision fusion rule
Distributed detection and estimation
EM algorithm
Probabilistic mixture models
Wireless sensor networks

ABSTRACT

This work presents a general framework for blindly estimating the sensor parameters of decision-fusion systems over wireless sensor networks (WSNs). The sensors report their binary decisions to a fusion center (FC) through parallel binary symmetric channels. Then, the FC makes the final decision by combining the noisy sensor decisions according to a certain fusion rule.

We present an algorithm for the FC to blindly estimate the sensor parameters from the noisy sensor decisions received after a number of sensing periods. The algorithm covers a wide variety of situations that may arise in WSNs. For example, the algorithm is applicable when the FC knows in advance some of the parameters of some sensors, when it knows the true hypothesis for a subset of sensing periods, or when only a subset of sensors communicates their decisions in each sensing period.

Based on the estimates of the system parameters, optimal channel-aware fusion rules are derived considering the minimum Bayes risk criterion. Simulation results show that, after sufficient sensing periods, the estimates of the WSN parameters are accurate enough for the fusion rule to exhibit near-optimal detection performance.

1. Introduction

This paper addresses the so-called canonical distributed detection problem [1] in wireless sensor networks (DD-WSN) composed of a set of spatially distributed sensors and a fusion center (FC). Sensors report their binary decisions about the presence or absence of a given event of interest to the FC through dedicated wireless channels. The FC then fuses the noisy binary decisions from the sensors to make the final decision according to a given fusion rule. The sensors do not communicate with each other, and there is no feedback from the FC to the sensors. Communication is assumed only between each sensor and the FC through the corresponding reporting channel. Therefore, the only information at the FC is the noisy sensor decisions at the output of the reporting channels. Dedicated reporting channels are quite common in DD-WSNs [2–11]. They contrast with multi-access channels where multiple sensors transmit simultaneously over a shared channel. Using dedicated channels can provide more reliable and predictable communication, but may require more bandwidth and energy than multi-access channels.

In general, the design of DD-WSN systems requires determining the local decision rules at the sensors and the fusion rule at the FC [1]. Variations of this formulation include the optimization of only the local decision rules for a given fusion rule [12–14], and the optimization of the fusion rule for given sensor decision rules [15–19]. In this work, we focus on the last case, where the reliabilities of the sensor decision rules are unknown to the FC.

In homogeneous WSNs, all sensors exhibit identical probabilities of detection and false-alarm [5,20,21]. Some authors consider semi-homogeneous WSNs where the sensors can have different detection probabilities but identical false-alarm probabilities [4,8]. This work considers fully heterogeneous networks where the sensors may operate at different probabilities of detection and false-alarm [17,19].

The reliability of the sensor decisions observed by the FC depends not only on the sensors themselves but also on the reporting channels. Therefore, the design of fusion rules requires considering the effect of the reporting channels as well. Those rules are called channel-aware fusion rules [9]. In the WSN literature, the two most common reporting channel models are the so-called noise fading channel (NFC) [17–19, 22] and the binary symmetric channel (BSC) [4,5,8,11,12,23,24]. In the first case, the FC receives a faded and noisy version of each sensor decision. Therefore, the output of a NFC is a complex random value characterized by its channel response and the noise power. The BSCs model the relationship between the binary decisions transmitted by the sensors and the noisy binary decisions at the FC. The reliability of a

☆ This work has been funded by MCIN/AEI/10.13039/501100011033 under grants PID2022-137099NB-C41 and PID2022-137099NB-C43.
* Corresponding author.
E-mail address: jesus.perez@unican.es (J. Perez).

https://doi.org/10.1016/j.sigpro.2025.110238
Received 4 April 2025; Received in revised form 24 July 2025; Accepted 4 August 2025
Available online 13 August 2025

BSC is characterized by its bit error probability (often called crossover probability). In this work, we consider the general case in which BSCs can have different bit error probabilities [5]. This usually occurs when the sensors are placed at different locations with respect to the FC. As usual in the related literature, we assume that the FC knows the error probabilities of the BSCs [2,3,6,8,10,23]. Note that the coherent demodulation of the sensor signals requires the FC to estimate the reporting channels. From the channel estimates, the FC can obtain the error probabilities of the BSCs.

The design of the optimal fusion rule is conceptually straightforward when the FC knows the detection and false-alarm probabilities of the sensors. In this case, the detection problem is a simple binary hypothesis testing problem, so the optimal fusion rule is the likelihood ratio test (LRT) [4]. To make the final decision, the FC compares the likelihood ratio (LR) with a decision threshold whose value depends on the detection criterion used [25,26]. The Neyman-Pearson theorem establishes that the detection rule that maximizes the probability of detection, for a given probability of false-alarm, is the likelihood-ratio test (LRT) with a decision threshold determined by the distribution of the FC observations under the null hypothesis $\mathcal{H}_0$. Therefore, it requires the FC to exactly know the sensors' probabilities of false alarm, which are unknown in our standard problem. Consequently, the Neyman-Pearson criterion is not applicable in our problem. In this work, we consider the Minimum Bayes Risk (MBR) detection criterion, which includes the minimum probability of error as a particular case. In this case, the optimal fusion rule is also the LRT, but the decision threshold is independent of the distribution of the FC observations. It only depends on the prior probability of occurrence of the event to be detected [15]. However, the fact that the LRT depends on the model parameters (probabilities of detection and false-alarm of the sensors, and prior probabilities) is a major drawback, as some or even all of them may be unknown to the FC in practice. In these cases, the LRT is inapplicable. Two approaches have been proposed to overcome this difficulty: 1) to use sub-optimal fusion rules independent of the unknown model parameters, and 2) first estimate the unknown model parameters and then design a fusion rule accordingly.

The counting rules (CR), also called voting rules, are the most popular blind fusion rules. The test statistic is simply the sum of the sensor decisions. Therefore, the fusion rule implicitly assigns the same weight to sensor decisions regardless of their sensing performance. Their main advantages are their simplicity and the fact that the FC does not need to know any system parameters. Ref. [27] has shown interesting properties of the CR in DD-WSNs. For example, in semi-homogeneous WSNs when the FC knows the probability of false-alarm of the sensors, the CR is statistically equivalent to the generalized LRT and to the Rao test under the mild assumption that the common probability of false alarm is lower than $1/2$. Also, in semi-homogeneous WSNs, the CR is statistically equivalent to the locally most-mean powerful test [28]. Moreover, the CR is the Uniform Most Powerful Invariant test in heterogeneous DD-WSNs when the FC knows the sensor probabilities of false-alarm [27]. Typically, the decision threshold of the CR is an integer value, which represents the minimum number of sensor decisions for the FC to decide the alternative hypothesis. Several methods have been proposed to select the optimum decision threshold in homogeneous networks [1,21] and in heterogeneous networks [11], but all of them require knowing the reliability of the sensors. In any case, the optimal decision threshold is highly dependent on the model parameters.

Another fusion rule independent of the unknown model parameters is the so-called Ideal Sensor Rule (ISR), where the LR is approximated by assuming ideal sensors [8]. Therefore, if the detection probabilities are unknown, they are set to one. Similarly, false alarm probabilities are set to zero when they are unknown. Consequently, the resulting test statistic only depends on the BSC's error probabilities.

The fusion rule proposed in [5] (known as "Wu rule" [8]) is an example of the second approach. Its test statistic is based on approximated maximum likelihood (ML) estimates of the sensor detection probabilities. This fusion rule is limited to homogeneous networks where the FC knows the false-alarm probability of the sensors. The fusion rules in [29] also belong to the second approach. These are described later.

Contribution

This paper presents a general method for the FC to learn the model parameters (the sensors' probabilities of detection and false-alarm and the prior probability of the event to be detected) in heterogeneous DD-WSNs from the decisions reported by the sensors through BSCs with known error probabilities. These decisions may be incomplete in the sense that only a subset of sensors reports their decisions in each sensing period.

We first derive the learning algorithm for what we call the standard case, where we assume that all model parameters are unknown to the FC, and the FC does not know the true hypothesis for any of the sensing periods. We then specialize the standard algorithm to the following interesting practical cases:

- The FC knows some of the parameters (probability of false-alarm and/or probability of detection) of a subset of sensors.
- The FC knows the true hypothesis for a subset of sensing periods. This case is particularly interesting when there is an initial calibration stage of the DD-WSN under controlled conditions, so the FC knows the actual hypothesis for the corresponding sensor decisions.

To derive the learning algorithm, we consider a probabilistic mixture model [30] of the sensor decisions with two components associated with the two hypotheses (presence or absence of the event to be detected). The noisy sensor decisions at the FC are the observed variables, and the presence/absence of the event to be detected is a latent (hidden) variable. Then, we apply the Expectation–Maximization (EM) algorithm [30–32] to estimate the unknown model parameters, which leads to quite simple closed-form expressions for both the E-step and the M-step.

Learning the sensor parameters is valuable in its own right. But in DD-WSS, the ultimate goal is to have an accurate fusion rule. We rewrite the LRT so the test statistic is the posterior probability of the sensor decisions, and the decision threshold is independent of the model parameters. Then, we propose a fusion rule where the test statistic is the estimate of the posterior probability of the sensor decisions according to the EM estimates of the model parameters. Therefore, the performance of the fusion rule is determined by the accuracy of the estimates of the posterior probabilities rather than the accuracy of the individual sensor parameter estimates.

Extensive simulation experiments show that, after sufficient sensing periods, the posterior probability estimates are accurate enough for the fusion rule to exhibit near-optimal detection performance.

Related works

Ref. [33] establishes the analytical relationships between the DD-WSN parameters and the joint probabilities of the sensor decisions, assuming error-free reporting channels. Then, the authors propose estimating the DD-WSN parameters by substituting the joint probabilities with their sampling estimates. In the particular case of three sensors DD-WSN, the authors derive a batch and an adaptive algorithm to analytically estimate its parameters. In [29] the authors pose the Least-Squares (LS) and the ML estimation of the model parameters in M-ary distributed detection, where the sensors report M-ary decisions to the FC. The computation of the LS estimates requires solving a system of non-linear equations, whereas the ML estimates are the solutions of an optimization problem. In both cases, the parameter estimates must be obtained numerically. The above references do not take into account the constraints imposed by the reporting channels. In addition, they assume that all sensors report their decisions to the FC in each sensing period, and all model parameters are unknown to the FC.

The parameter estimation in DD-WSNs is closely related to the so-called unsupervised ensemble learning, where a meta-classifier blindly estimates the reliabilities of an ensemble of classifiers from their decisions. The meta-classifier then derives a fusion rule, taking into account the estimated classifiers' reliabilities. In DD-WSN, the FC and the sensors play the same role as the meta-classifier and the classifiers in unsupervised ensemble learning, respectively. In this context, focusing on binary classifiers, spectral-based methods have been proposed to rank the classifiers according to their accuracies [34] or to estimate their accuracy parameters [35]. More recently, [36] presents a learning scheme for multi-class ensemble classification (called M-ary distributed detection in the DD-WSN context) based on a moment matching method that leverages joint tensor and matrix factorization. Mathematically, the unsupervised ensemble learning problem is equivalent to estimating the parameters of a mixture of multinomial random variables, or Bernoulli random variables in the case of binary classifiers. Those parameters are related to the performance of the individual classifiers and the prior probabilities of the classes. Following this approach, the EM algorithm has been proposed for unsupervised ensemble learning [37]. But in unsupervised ensemble learning, the classifier parameters are usually unconstrained and all sensor decisions are available to the meta-classifier. The specific constraints of the DD-WSN problem imposed by the reporting channels and sensor operating regions, as well as the incompleteness of the observations, make all these ensemble learning algorithms inapplicable to DD-WSN.

Paper organization

The remainder of this paper is organized as follows. Section 2 describes the system model and derives the probabilistic distribution of sensor decisions at the FC as a function of the model parameters. In Section 3 we propose the fusion rule, under the MBR detection criterion, based on the estimates of the DD-WSN parameters. Afterwards, Section 4 derives the EM estimators of the parameters for the standard case. In Section 5, we consider the scenario where the FC has partial knowledge of some model parameters. Next, we derive the EM estimators for the remaining unknown parameters. Section 6 shows a set of comprehensive simulation results for a wide variety of cases. It is divided into two subsections. The first one presents results that show the performance of the proposed estimators. Then, the second subsection shows simulation results that illustrate the detection performance of the proposed fusion rule. Finally, Section 7 concludes the paper.

Notation

Throughout this paper, we use light-face lowercase letters for scalar quantities, bold-face lowercase letters for vectors, and bold-face capital letters for matrices. We employ $\mathbf{a} = (a_i)_{1\le i\le I}$ to denote a real vector, of length I, whose ith entry is a_i, and $\mathbf{A} = (a_{i,j})_{1\le i\le I, 1\le j\le J}$ denotes a $I \times J$ real matrix whose (i,j)-th entry is $a_{i,j}$. The all-zeros vector and all-ones vector are denoted by $\mathbf{0}$ and $\mathbf{1}$, respectively. $\mathbb{E}[\cdot]$ denotes the expectation operator and $P(A)$ refers to the probability of the event A. $\overline{\mathcal{S}}$ denotes the complement of the set $\mathcal{S}$, and $|\mathcal{S}|$ is its cardinality. Finally, $\hat{\theta}$ refers to an estimate of parameter θ.

2. System model

Fig. 1 shows a DD-WSN comprising K sensors and a fusion center (FC).

The binary variable z denotes the presence ($z = 1$) or absence ($z = 0$) of the event to detect. We denote the above hypotheses by $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. Each sensor makes its own binary decision, $y_k \in \{0,1\}$, about the absence or presence of the phenomenon. The decision y_k is subject to a probability of detection $P^{(s)}_{d,k} = P(y_k = 1|\mathcal{H}_1)$ and a probability of false-alarm $P^{(s)}_{f,k} = P(y_k = 1|\mathcal{H}_0)$. Each sensor reports its decision to the FC through a dedicated BSC, so the FC receives a noisy binary value, $x_k \in \{0,1\}$. Let $e_k = P(x_k = 0|y_k = 1) = P(x_k = 1|y_k = 0)$

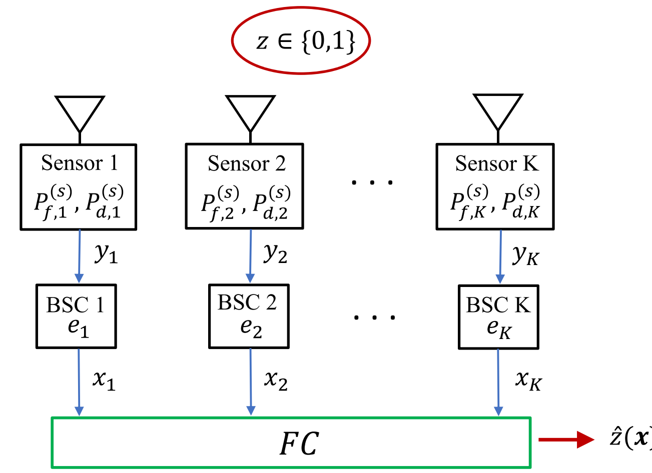

Fig. 1. Distributed detection in a wireless sensor network (DD-WSN).

be the error probability of the BSC of sensor k [10,38]. From the noisy decision vector $\mathbf{x} = (x_k)_{1\le k\le K}$, the FC makes the global decision according to a certain fusion rule $\hat{z}(\mathbf{x}) \in \{0,1\}$.

We assume that the sensor decisions at the FC, x_k, are conditionally independent. This requires both the BSC errors, e_k, and the sensor decisions, y_k, to be conditionally independent. This is a common assumption in the WSN literature [4,5,7,8,11,15,16,19,39] because it leads to tractable analyses and useful fusion rules.

Let $P^{(FC)}_{f,k} = P(x_k = 1|\mathcal{H}_0)$ and $P^{(FC)}_{d,k} = P(x_k = 1|\mathcal{H}_1)$ be the probabilities of false-alarm and detection of sensor k at the FC, respectively. They depend on the sensors' probabilities and on the BSC error probability as follows [4,11]

$$P^{(FC)}_{f,k} = e_k(1 - P^{(s)}_{f,k}) + (1 - e_k)P^{(s)}_{f,k} = e_k + P^{(s)}_{f,k}(1 - 2e_k), \tag{1}$$

$$P^{(FC)}_{d,k} = e_k(1 - P^{(s)}_{d,k}) + (1 - e_k)P^{(s)}_{d,k} = e_k + P^{(s)}_{d,k}(1 - 2e_k). \tag{2}$$

Note that, when e_k is known, $P^{(s)}_{f,k}$ and $P^{(s)}_{d,k}$ are determined by $P^{(FC)}_{f,k}$ and $P^{(FC)}_{d,k}$, respectively. Without loss of generality, we assume that $0 < e_k < 1/2$. The limit value $e_k = 0$ would correspond to an error-free channel, whereas $e_k = 1/2$ would be the case of a fully random channel. We also assume that $P^{(s)}_{f,k} \le P^{(s)}_{d,k}$. This assumption comes from the informativeness of the sensors' decisions. In other words, we assume that sensor decisions are not worse than random guessing. Then, from (1) and (2), it is straightforward to show that the sensor error probabilities at the FC are constrained as follows,

$$e_k \le P^{(FC)}_{f,k} \le P^{(FC)}_{d,k} \le 1 - e_k. \tag{3}$$

To simplify notation, hereafter we will denote $P^{(FC)}_{f,k}$ and $P^{(FC)}_{d,k}$ by f_k and d_k, respectively. The kth sensor output is governed by Bernoulli distributions with parameters f_k and d_k for each hypothesis. Since we assume that the x_k are conditionally independent, the conditional probability mass function (pmf) of the decision vector under each hypothesis will be

$$p(\mathbf{x}|\mathcal{H}_0;\mathbf{f}) = \prod_{k=1}^{K} f_k^{x_k}\,(1-f_k)^{1-x_k}, \qquad p(\mathbf{x}|\mathcal{H}_1;\mathbf{d}) = \prod_{k=1}^{K} d_k^{x_k}\,(1-d_k)^{1-x_k}, \tag{4}$$

where $\mathbf{f} = (f_k)_{1\le k\le K}$ and $\mathbf{d} = (d_k)_{1\le k\le K}$ contain the sensors' probabilities of false-alarm and detection observed by the FC, respectively. The marginal distribution of $\mathbf{x}$ will be the mixture of the conditional pmfs (4) with mixing coefficients equal to the prior probabilities of $\mathcal{H}_0$ and $\mathcal{H}_1$,

$$p(\mathbf{x};\Theta) = (1-u)\; p(\mathbf{x}|\mathcal{H}_0;\mathbf{f}) + u\; p(\mathbf{x}|\mathcal{H}_1;\mathbf{d}), \tag{5}$$

where $\Theta = \{\mathbf{f}, \mathbf{d}, u\}$ denotes the set of model parameters, $u = P(\mathcal{H}_1)$ and $1-u = P(\mathcal{H}_0)$. Let z be a binary variable denoting the true hypothesis. Then, the joint pmf of $\mathbf{x}$ and z can be written as follows

$$p(\mathbf{x}, z; \Theta) = \left[(1-u)\ p(\mathbf{x}|\mathcal{H}_0; \mathbf{f})\right]^{1-z} \left[u\ p(\mathbf{x}|\mathcal{H}_1; \mathbf{d})\right]^{z}. \tag{6}$$

3. Fusion rules

In DD-WSN, the FC makes the final decision from the decisions reported by the individual sensors, so the ultimate goal is to design an accurate fusion rule $\hat{z}(\mathbf{x}) \in \{0,1\}$ (see Fig. 1).

3.1. Performance metrics

As $\mathbf{x}$ is a binary random vector, the probabilities of false-alarm and detection of $\hat{z}(\mathbf{x})$ are

$$P_f = P(\hat{z}(\mathbf{x}) = 1|\mathcal{H}_0) = \sum_{\mathbf{x}\in\Omega_{\mathbf{x}}} \hat{z}(\mathbf{x})\ p(\mathbf{x}|\mathcal{H}_0; \mathbf{f}),$$

$$P_d = P(\hat{z}(\mathbf{x}) = 1|\mathcal{H}_1) = \sum_{\mathbf{x}\in\Omega_{\mathbf{x}}} \hat{z}(\mathbf{x})\ p(\mathbf{x}|\mathcal{H}_1; \mathbf{d}), \tag{7}$$

where the sample space of $\mathbf{x}$, denoted by $\Omega_{\mathbf{x}}$, is the set of binary K-tuples with cardinality $|\Omega_{\mathbf{x}}| = 2^K$. We consider the Bayesian risk [25,26] as the performance metric of the fusion rules,

$$B = P_f\ (1-u)\ C_{1,0} + (1-P_d)\ u\ C_{0,1} + (1-P_f)\ (1-u)\ C_{0,0} + P_d\ u\ C_{1,1}, \tag{8}$$

where $C_{i,j}$ is the cost of deciding $\hat{z}(\mathbf{x}) = i$ when the true value is $z = j$. Usually, no cost is assigned to the correct decisions so that $C_{0,0} = C_{1,1} = 0$. In addition, if the two types of errors (false-alarms and miss-detections) have unit cost, $C_{1,0} = C_{0,1} = 1$, the Bayes risk reduces to the probability of error:

$$P_e = P(\hat{z}(\mathbf{x}) \neq z) = P_f\ (1-u) + (1-P_d)\ u. \tag{9}$$

3.2. Optimal and proposed fusion rule

The Minimum Bayes Risk (MBR) criterion aims to minimize the Bayesian risk (8). Assuming the FC knows the model parameters, the MBR fusion rule is the LRT [1,15]

$$\hat{z}_{MBR}(\mathbf{x}) = \begin{cases} 1, & \text{if} \quad \Lambda(\mathbf{x}; \mathbf{f}, \mathbf{d}) > \lambda_{MBR}(u), \\ 0, & \text{if} \quad \Lambda(\mathbf{x}; \mathbf{f}, \mathbf{d}) \le \lambda_{MBR}(u), \end{cases} \tag{10}$$

where the test statistic (the LR) and the decision threshold are given by

$$\Lambda(\mathbf{x}; \mathbf{f}, \mathbf{d}) = \frac{p(\mathbf{x}|\mathcal{H}_1; \mathbf{d})}{p(\mathbf{x}|\mathcal{H}_0; \mathbf{f})}, \qquad \lambda_{MBR}(u) = \frac{1-u}{u}\ \frac{C_{1,0} - C_{0,0}}{C_{0,1} - C_{1,1}}.$$

It is easy to show that the MBR fusion rule can be rewritten as follows

$$\hat{z}_{MBR}(\mathbf{x}) = \begin{cases} 1, & \text{if} \quad t(\mathbf{x}; \Theta) > \lambda'_{MBR}, \\ 0, & \text{if} \quad t(\mathbf{x}; \Theta) \le \lambda'_{MBR}, \end{cases} \tag{11}$$

where the test statistic is now the posterior probability of $\mathcal{H}_1$,

$$t(\mathbf{x}; \Theta) = P\left(z = 1 \mid \mathbf{x}; \Theta\right) = \frac{u\ p(\mathbf{x}|\mathcal{H}_1; \mathbf{d})}{(1-u)\ p(\mathbf{x}|\mathcal{H}_0; \mathbf{f}) + u\ p(\mathbf{x}|\mathcal{H}_1; \mathbf{d})}, \tag{12}$$

and the new decision threshold is

$$\lambda'_{MBR} = \frac{C_{1,0} - C_{0,0}}{C_{0,1} - C_{1,1} + C_{1,0} - C_{0,0}}.$$

The test statistic in (11) depends on the model parameters Θ, while the decision threshold is independent of them. When all or some model parameters are unknown, we propose a fusion rule as (11), but replacing the unknown parameters in the test statistic expression (12) by their estimates provided by the EM algorithm: $t(\mathbf{x}; \hat{\Theta})$. Notice that the detection performance is determined by the accuracy of the posterior probability estimation, rather than by the accuracy of the individual model parameter estimates.

4. EM estimates of the model parameters

Consider a sequence of N consecutive sensing periods, and let $\mathbf{z} = (z_n)_{1\le n\le N}$ be the unknown true hypothesis at them. In each sensing period, a subset of sensors reports their decisions to the FC. Let $l_{k,n}$ be a binary indicator variable such that $l_{k,n} = 1$ if sensor k reported its decision at sensing period n, and $l_{k,n} = 0$ otherwise. Therefore, $M_k = \sum_{n=1}^{N} l_{k,n}$ is the number of decisions reported by the sensor k to the FC, and $K_n = \sum_{k=1}^{K} l_{k,n}$ is the number of sensors that report decisions to the FC at sensing period n. We assume that each sensor reports its decisions at least once, so $M_k > 0$ for all k. We also assume that at least one sensor reports its decision in each sensing period, so $K_n > 0$ for all n.

Let $\mathbf{x}_n = (x_{k,n})_{1\le k\le K}$ be the decision vector (possibly incomplete) at sensing period n. From (4), the conditional pmfs at $\mathbf{x}_n$ can be written as follows:

$$p(\mathbf{x}_n|\mathcal{H}_0; \mathbf{f}) = \prod_{k=1}^{K} \left[f_k^{x_{k,n}} (1-f_k)^{1-x_{k,n}}\right]^{l_{k,n}},$$

$$p(\mathbf{x}_n|\mathcal{H}_1; \mathbf{d}) = \prod_{k=1}^{K} \left[d_k^{x_{k,n}} (1-d_k)^{1-x_{k,n}}\right]^{l_{k,n}}. \tag{13}$$

The sensor decisions at the FC can be arranged in an $K \times N$ incomplete matrix $\mathbf{X} = (x_{k,n})_{1\le k\le K,\ 1\le n\le N}$, whose nth column is $\mathbf{x}_n$. Given $\mathbf{X}$, the ML estimates of the model parameters are

$$\hat{\Theta} = \{\hat{\mathbf{f}}, \hat{\mathbf{d}}, \hat{u}\} = \underset{\Theta\in\mathcal{S}_\Theta}{\operatorname{argmax}}\ \log L(\Theta), \tag{14}$$

where the log-likelihood function is given by

$$\log L(\Theta) = \sum_{n=1}^{N} \log p(\mathbf{x}_n; \Theta). \tag{15}$$

The feasible set is $\mathcal{S}_\Theta = \{\mathbf{f}, \mathbf{d}, u \mid u \in (0,1), (f_k, d_k) \in \mathcal{S}_k, \forall k\}$, where $\mathcal{S}_k$ denotes the set of feasible operational points of sensor k given by (3),

$$\mathcal{S}_k = \{(f_k, d_k) \mid 0 < e_k \le f_k \le d_k \le 1 - e_k < 1\} \tag{16}$$

The optimization problem (14) has no closed-form solution, so we resort to the EM algorithm [31,32] to solve it, where $\mathbf{X}$ contains the observed variables and $\mathbf{z}$ the latent variables. The EM is an iterative algorithm whose iterations comprise two steps, termed the E-step and the M-step:

- E-step:

$$Q(\Theta; \hat{\Theta}^{(i-1)}) = \mathop{\mathbb{E}}_{\mathbf{z}}\left[\log L_c(\Theta) \mid \mathbf{X}; \hat{\Theta}^{(i-1)}\right], \tag{17}$$

- M-step:

$$\hat{\Theta}^{(i)} = \underset{\Theta\in\mathcal{S}_\Theta}{\operatorname{argmax}}\ Q(\Theta; \hat{\Theta}^{(i-1)}), \tag{18}$$

where i is the index to iterations and $\log L_c(\Theta)$ denotes the log-likelihood function if the true hypotheses $\mathbf{z}$ were known, which is given by

$$\log L_c(\Theta) = \sum_{n=1}^{N} \log p(\mathbf{x}_n, z_n; \Theta). \tag{19}$$

In [30,40] the EM algorithm is applied to mixtures of Bernoulli random variables. In [32,37] it is applied to mixtures of multinomial random variables, which generalize mixtures of Bernoulli random variables. But in all the above cases, the parameters are unconstrained and the observations are complete. The specific constraints (16) of the DD-WSN problem and the incomplete nature of the observations $\mathbf{x}_n$ make the application of the EM algorithm to (14) challenging and different from those references.

E-step:

Substituting (6) into (19),

$$\log L_c(\Theta) = \sum_{n=1}^{N} (1 - z_n) \log \left[(1-u)\ p(\mathbf{x}_n|\mathcal{H}_0;\mathbf{f})\right] + z_n \log \left[u\ p(\mathbf{x}_n|\mathcal{H}_1;\mathbf{d})\right]. \tag{20}$$

Then, the expectation function (17) will be

$$Q(\Theta;\hat{\Theta}^{(i-1)}) = \sum_{n=1}^{N} \left(1 - t_n^{(i)}\right) \log \left[(1-u)\ p(\mathbf{x}_n|\mathcal{H}_0;\mathbf{f})\right] + t_n^{(i)} \log \left[u\ p(\mathbf{x}_n|\mathcal{H}_1;\mathbf{d})\right], \tag{21}$$

where

$$t_n^{(i)} = \mathbb{E}_{\mathbf{z}}\left[z_n|\mathbf{x}_n;\hat{\Theta}^{(i-1)}\right] = P\left(z_n = 1|\mathbf{x}_n;\hat{\Theta}^{(i-1)}\right) \tag{22}$$

$$= \frac{\hat{u}^{(i-1)}\ p\left(\mathbf{x}_n|\mathcal{H}_1;\hat{\mathbf{d}}^{(i-1)}\right)}{\left(1-\hat{u}^{(i-1)}\right)\ p\left(\mathbf{x}_n|\mathcal{H}_0;\hat{\mathbf{f}}^{(i-1)}\right) + \hat{u}^{(i-1)}\ p\left(\mathbf{x}_n|\mathcal{H}_1;\hat{\mathbf{d}}^{(i-1)}\right)}.$$

Indeed, the posterior probability $t_n^{(i)}$ can be interpreted as a soft estimate of z_n at iteration i once $\mathbf{x}_n$ is observed and given the parameter estimation at the previous iteration. For notation convenience, we group these soft estimates in vector $\mathbf{t}^{(i)} = (t_n^{(i)})_{1\le n\le N}$.

M-step:

Substituting (21) into (18),

$$\hat{\Theta}^{(i)} = \{\hat{\mathbf{f}}^{(i)}, \hat{\mathbf{d}}^{(i)}, \hat{u}^{(i)}\} = \underset{\Theta\in S_\Theta}{\operatorname{argmax}} \sum_{n=1}^{N} (1 - t_n^{(i)})\log(1-u) + t_n^{(i)}\log u + (1 - t_n^{(i)})\log p(\mathbf{x}_n|\mathcal{H}_0;\mathbf{f}) + t_n^{(i)}\log p(\mathbf{x}_n|\mathcal{H}_1;\mathbf{d}). \tag{23}$$

Since the terms depending on u are decoupled from the rest, (23) can be broken down into two independent problems:

$$\hat{u}^{(i)} = \underset{0<u<1}{\operatorname{argmax}} \sum_{n=1}^{N} (1 - t_n^{(i)})\log(1-u) + t_n^{(i)}\log u, \tag{24}$$

$$\{\hat{\mathbf{f}}^{(i)}, \hat{\mathbf{d}}^{(i)}\} = \underset{(f_k,d_k)\in S_k,\forall k}{\operatorname{argmax}} \sum_{k=1}^{K}\sum_{n=1}^{N} \left(1 - t_n^{(i)}\right)\ l_{k,n}\left[x_{k,n}\log f_k + (1-x_{k,n})\log(1-f_k)\right] + t_n^{(i)}\ l_{k,n}\left[x_{k,n}\log d_k + (1-x_{k,n})\log(1-d_k)\right], \tag{25}$$

where we have considered the expressions of the conditional probabilities (13). Moreover, in (25) the terms of the objective function associated with each sensor are decoupled from each other, so it can be divided into K decoupled problems of the form

$$\{\hat{f}_k^{(i)}, \hat{d}_k^{(i)}\} = \underset{(f_k,d_k)\in S_k}{\operatorname{argmax}} \sum_{n=1}^{N} \left(1 - t_n^{(i)}\right)\ l_{k,n}\left[x_{k,n}\log f_k + (1-x_{k,n})\log(1-f_k)\right] + t_n^{(i)}\ l_{k,n}\left[x_{k,n}\log d_k + (1-x_{k,n})\log(1-d_k)\right]. \tag{26}$$

Optimization problem (24) can be expressed as

$$\hat{u}^{(i)} = \underset{0<u<1}{\operatorname{argmax}}\ N_0^{(i)}\log(1-u) + N_1^{(i)}\log u, \tag{27}$$

where

$$N_1^{(i)} = \sum_{n=1}^{N} t_n^{(i)}, \quad N_0^{(i)} = \sum_{n=1}^{N}(1 - t_n^{(i)}), \tag{28}$$

can be interpreted as soft estimates of the number of sensing periods under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively, according to $\mathbf{t}^{(i)}$. Setting the derivative of the objective function in (27) to zero, after simple algebraic manipulations, yields

$$\hat{u}^{(i)} = \frac{N_1^{(i)}}{N}, \tag{29}$$

where we have considered the fact that $N_0^{(i)} + N_1^{(i)} = N$.

Before solving (26), we first introduce the following notation

$$M_{k,1|1}^{(i)} = \sum_{n=1}^{N} l_{k,n}x_{k,n}t_n^{(i)}, \quad M_{k,0|0}^{(i)} = \sum_{n=1}^{N} l_{k,n}(1-x_{k,n})(1-t_n^{(i)}),$$
$$M_{k,1|0}^{(i)} = \sum_{n=1}^{N} l_{k,n}x_{k,n}(1-t_n^{(i)}), \quad M_{k,0|1}^{(i)} = \sum_{n=1}^{N} l_{k,n}(1-x_{k,n})t_n^{(i)}, \tag{30}$$

where the term $M_{k,j|s}^{(i)}$ can be interpreted as the soft estimate of the number of decisions $\mathcal{H}_j$ of sensor k under $\mathcal{H}_s$, according to $\mathbf{t}^{(i)}$. Notice that

$$M_{k,0}^{(i)} = M_{k,0|0}^{(i)} + M_{k,1|0}^{(i)} = \sum_{n=1}^{N} l_{k,n}(1-t_n^{(i)}), \qquad M_{k,1}^{(i)} = M_{k,0|1}^{(i)} + M_{k,1|1}^{(i)} = \sum_{n=1}^{N} l_{k,n}t_n^{(i)} \tag{31}$$

are the estimated number of decisions of sensor k under $\mathcal{H}_0$ and $\mathcal{H}_1$ respectively, and

$$M_{k,0}^{(i)} + M_{k,1}^{(i)} = \sum_{n=1}^{N} l_{k,n} = M_k \tag{32}$$

is the total number of decisions reported by sensor k.

Considering (30), the optimization problem (26) can be written as follows

$$\{\hat{f}_k^{(i)}, \hat{d}_k^{(i)}\} = \underset{(f_k,d_k)\in S_k}{\operatorname{argmax}}\ M_{k,1|0}^{(i)}\log f_k + M_{k,0|0}^{(i)}\log(1-f_k) + M_{k,1|1}^{(i)}\log d_k + M_{k,0|1}^{(i)}\log(1-d_k). \tag{33}$$

Appendix A shows that (33) has only one solution that can be expressed as follows:

$$\left(\hat{f}_k^{(i)}, \hat{d}_k^{(i)}\right) = \begin{cases} \left(a_k^{(i)}, b_k^{(i)}\right), & \text{if} \quad e_k < a_k^{(i)} < b_k^{(i)} < 1 - e_k \\ \left(c_k, c_k\right), & \text{if} \quad e_k < c_k < 1 - e_k \quad \wedge \quad a_k^{(i)} \ge b_k^{(i)} \\ \left(e_k, b_k^{(i)}\right), & \text{if} \quad a_k^{(i)} \le e_k < b_k^{(i)} < 1 - e_k \\ \left(a_k^{(i)}, 1 - e_k\right), & \text{if} \quad e_k < a_k^{(i)} < 1 - e_k \le b_k^{(i)} \\ \left(e_k, e_k\right), & \text{if} \quad b_k^{(i)} \le e_k \quad \wedge \quad c_k \le e_k \\ \left(1 - e_k, 1 - e_k\right), & \text{if} \quad a_k^{(i)} \ge 1 - e_k \quad \wedge \quad c_k \ge 1 - e_k \\ \left(e_k, 1 - e_k\right), & \text{if} \quad a_k^{(i)} \le e_k \quad \wedge \quad 1 - e_k \le b_k^{(i)} \end{cases} \tag{34}$$

where

$$a_k^{(i)} = \frac{M_{k,1|0}^{(i)}}{M_{k,0}^{(i)}}, \quad b_k^{(i)} = \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}}, \qquad c_k = \frac{M_{k,1|0}^{(i)} + M_{k,1|1}^{(i)}}{M_k} = \frac{1}{M_k}\sum_{n=1}^{N} l_{k,n}x_{k,n}. \tag{35}$$

The terms $a_k^{(i)}$ and $b_k^{(i)}$ can be interpreted as the soft sample probabilities of false-alarm and detection of sensor k, according to $\mathbf{t}^{(i)}$. For $a_k^{(i)}$ and $b_k^{(i)}$ to be well-defined (take finite values), $M_{k,0}^{(i)}$ and $M_{k,1}^{(i)}$ must be greater than zero, which is fulfilled when $t_n^{(i)} \in (0,1)$. The term c_k is the fraction of $\mathcal{H}_1$ decisions of the sensor k, which does not change with iterations, so it can be calculated in advance.

The seven cases of the expression (34) are mutually exclusive. Each corresponds to a region of the $\left(a_k^{(i)}, b_k^{(i)}\right)$-coordinate plane, as depicted in Fig. 2.

4.1. Initialization and convergence

The EM algorithm requires an initial estimate of the model parameters: $\hat{\Theta}^{(0)}$. For this, we first set an initial vector of soft estimates, $\mathbf{t}^{(0)}$, and then apply the M-step to obtain $\hat{\Theta}^{(0)}$. We propose the following

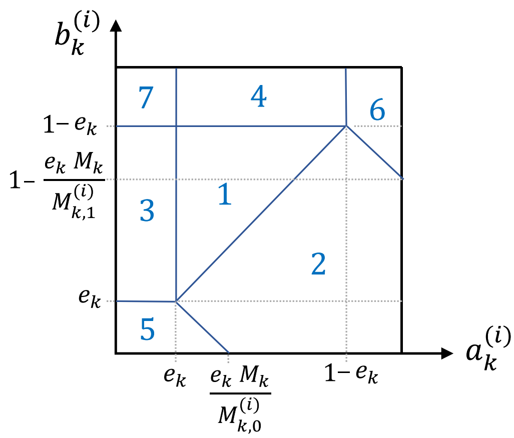

Fig. 2. Mutually exclusive regions associated with the cases of Eq. (34).

Algorithm 1 : ML estimation of the model parameters $\Theta = \{\mathbf{f}, \mathbf{d}, u\}$

input: $\mathbf{X}$, $\mathbf{e}$, ϵ_t, ϵ_L
Set the initial soft estimates $\mathbf{t}^{(0)}$ from (36)
Apply the M-step to $\mathbf{t}^{(0)}$ to obtain the initial estimates $\hat{\Theta}^{(0)}$
Compute the log-likelihood function, $\log L(\hat{\Theta}^{(0)})$, from (15)
Compute the fraction of $\mathcal{H}_1$ decisions of the sensors, $\{c_k\}_{k=1}^{K}$, from (35)
Initialize iterations index: $i = 0$
repeat
    $i = i + 1$
    E-step:
        Compute the soft estimates, $\mathbf{t}^{(i)}$, from (22) and (4)
    M-step:
        Compute $N_0^{(i)}$ and $N_1^{(i)}$ from (28)
        Compute $\hat{u}^{(i)}$ from (29)
        for $k = 1$ to K
            Compute $a_k^{(i)}$ and $b_k^{(i)}$ from (35)
            Compute $\hat{f}_k^{(i)}$ and $\hat{d}_k^{(i)}$ from (34)
        end for
    Compute the log-likelihood function $\log L(\hat{\Theta}^{(i)})$ from (15)
until convergence: $\frac{|\log L(\hat{\Theta}^{(i)}) - \log L(\hat{\Theta}^{(i-1)})|}{|\log L(\hat{\Theta}^{(i)})|} < \epsilon_L$
return: Final parameter estimates, $\hat{\Theta} = \hat{\Theta}^{(i)}$

heuristic soft estimates:

$$t_n^{(0)} = \begin{cases} \epsilon_t, & \text{if} \quad \sum_{k=1}^{K} l_{k,n} x_{k,n} < K_n/2 \\ \frac{1}{2}, & \text{if} \quad \sum_{k=1}^{K} l_{k,n} x_{k,n} = K_n/2 \\ 1-\epsilon_t, & \text{if} \quad \sum_{k=1}^{K} l_{k,n} x_{k,n} > K_n/2, \end{cases} \qquad 0 < \epsilon_t \ll 1. \tag{36}$$

Put into words, given a decision vector $\mathbf{x}_n$, its initial soft estimate $t_n^{(0)}$ is set to ϵ_t when the majority of the active sensors decide $\mathcal{H}_0$. Similarly, $t_n^{(0)}$ is set to $1 - \epsilon_t$ when the majority of the active sensors decide $\mathcal{H}_1$. In case of a tie, $t_n^{(0)}$ is set to $1/2$. The parameter ϵ_t ensures that $t_n^{(0)} \in (0, 1)$ for all n, so $a_k^{(0)}$ and $b_k^{(0)}$ will be well-defined and the initial estimates will belong to the feasible set: $\hat{\Theta}^{(0)} \in \mathcal{S}_\Theta$. Then, according to (13) and (22), $t_n^{(1)} \in (0, 1)$ for all n, so $\hat{\Theta}^{(1)} \in \mathcal{S}_\Theta$. Consequently, ϵ_t ensures that the expressions of the M-step will be consistent across all iterations, and $\hat{\Theta}^{(i)} \in \mathcal{S}_\Theta$ for all i. The EM algorithm always increases the value of the log-likelihood function in each iteration [30–32]. Therefore, since the log-likelihood function (15) is bounded for $\Theta \in \mathcal{S}_\Theta$, the algorithm converges monotonically. Then, convergence may be determined by observing when the log-likelihood function stops increasing with the iterations. Moreover, since $\mathcal{S}_\Theta$ is a convex set and $Q(\Theta; \Theta')$ is a continuous function in both arguments (see (21)), the limit point of the EM sequence $\hat{\Theta}$ is either a local maximum of $L(\Theta)$ in the interior of $\mathcal{S}_\Theta$ or a boundary point of $\mathcal{S}_\Theta$ where $L(\hat{\Theta}) \geq L(\Theta)$ for all $\Theta \in \mathcal{S}_\Theta \cap B_{\hat{\Theta}}$, $B_{\hat{\Theta}}$ being an open ball centered on $\hat{\Theta}$ [41].

In general, the log-likelihood function (15) can exhibit multiple local maxima. Moreover, there can be multiple local maxima attaining the ML. This usually occurs in situations with few decision vectors and/or a large number of parameters to estimate. In our experience, the initial soft estimates given by (36) are particularly useful for locating a local maximum that leads to good detection performance.

5. Partial knowledge at the FC

It might be the case that some of the model parameters are known by the FC. For example, it might know the probabilities of false-alarm and/or detection for a subset of sensors, or it might know the prior probability u. Those cases can occur simultaneously. The following Sections 5.1 and 5.2 particularize the EM algorithm to these cases.

In some cases, there is a training or calibration stage, under controlled conditions, in which the FC knows the actual hypothesis for a subset of sensing periods. We refer to the corresponding sensor decisions as labeled or supervised. Section 5.3 shows how to apply the EM algorithm in this case.

5.1. Known probabilities of false-alarm and/or detection of some sensors

Let $\mathcal{S}_\mathbf{f}$ and $\mathcal{S}_\mathbf{d}$ be the subsets of sensors for which the FC knows their probabilities of false-alarm and detection, respectively. Then, f_k is known for $k \in \mathcal{S}_\mathbf{f}$, and unknown for $k \in \overline{\mathcal{S}}_\mathbf{f}$. Similarly, d_k is known for $k \in \mathcal{S}_\mathbf{d}$, and unknown for $k \in \overline{\mathcal{S}}_\mathbf{d}$.

In the E-step, the soft estimates are calculated as in the standard case (22) but with the appropriate pmfs:

$$p(\mathbf{x}_n|\mathcal{H}_0; \hat{\mathbf{f}}^{(i-1)}) = \prod_{s \in \mathcal{S}_\mathbf{f}} \left[f_s^{x_{s,n}} (1-f_s)^{1-x_{s,n}} \right]^{l_{s,n}} \prod_{k \in \overline{\mathcal{S}}_\mathbf{f}} \left[\left(\hat{f}_k^{(i-1)} \right)^{x_{k,n}} \left(1 - \hat{f}_k^{(i-1)} \right)^{1-x_{k,n}} \right]^{l_{k,n}} \tag{37}$$

$$p(\mathbf{x}_n|\mathcal{H}_1; \hat{\mathbf{d}}^{(i-1)}) = \prod_{s \in \mathcal{S}_\mathbf{d}} \left[d_s^{x_{s,n}} (1-d_s)^{1-x_{s,n}} \right]^{l_{s,n}} \prod_{k \in \overline{\mathcal{S}}_\mathbf{d}} \left[\left(\hat{d}_k^{(i-1)} \right)^{x_{k,n}} \left(1 - \hat{d}_k^{(i-1)} \right)^{1-x_{k,n}} \right]^{l_{k,n}} \tag{38}$$

In the M-step, there are four possibilities regarding the optimization problem (33):

(1) $k \in \mathcal{S}_\mathbf{f} \cap \mathcal{S}_\mathbf{d}$: In this case, all parameters of sensor k are known by the FC, so there is nothing to estimate.
(2) $k \in \overline{\mathcal{S}}_\mathbf{f} \cap \overline{\mathcal{S}}_\mathbf{d}$: Since both f_k and d_k are unknown, their estimates $\hat{f}_k^{(i)}$ and $\hat{d}_k^{(i)}$ are given by (34).
(3) $k \in \mathcal{S}_\mathbf{f} \cap \overline{\mathcal{S}}_\mathbf{d}$: Now, f_k is known and d_k is unknown. Then, (33) reduces to

$$\hat{d}_k^{(i)} = \underset{d_k \in \mathcal{S}_k}{\operatorname{argmax}} \; M_{k,1|1}^{(i)} \log d_k + M_{k,0|1}^{(i)} \log(1-d_k), \tag{39}$$

where $\mathcal{S}_k = \{d_k \mid f_k \leq d_k \leq 1 - e_k\}$. Appendix B shows that in this case (39) has a unique solution given by

$$\hat{d}_k^{(i)} = \begin{cases} b_k^{(i)}, & \text{if} \quad f_k < b_k^{(i)} < 1 - e_k \\ f_k, & \text{if} \quad f_k \geq b_k^{(i)} \\ 1 - e_k, & \text{otherwise.} \end{cases} \tag{40}$$

(4) $k \in \overline{S}_{\mathbf{f}} \cap S_{\mathbf{d}}$: Now, d_k is known and f_k is unknown, so (33) reduces to

$$\hat{f}_k^{(i)} = \underset{f_k \in S_k}{\operatorname{argmax}} \quad M_{k,1|0}^{(i)} \log f_k + M_{k,0|0}^{(i)} \log(1 - f_k), \tag{41}$$

where $S_k = \{f_k \mid e_k \leq f_k \leq d_k\}$. Appendix C shows that in this case (41) has a unique solution given by

$$\hat{f}_k^{(i)} = \begin{cases} a_k^{(i)}, & \text{if} \quad e_k < a_k^{(i)} < d_k \\ e_k, & \text{if} \quad e_k \geq a_k^{(i)} \\ d_k, & \text{otherwise.} \end{cases} \tag{42}$$

5.2. Known prior probability

When the FC knows the prior probability u, the set of unknown parameters reduces to $\Theta = \{\mathbf{f}, \mathbf{d}\}$ and the feasible set is now $S_\Theta = \{\mathbf{f}, \mathbf{d} \mid (f_k, d_k) \in S_k, \forall k\}$. In the E-step, the soft estimates will be given by

$$t_n^{(i)} = \underset{\mathbf{z}}{\mathbb{E}} \left[z_n \mid \mathbf{x}_n; \hat{\Theta}^{(i-1)}\right] = \frac{u \quad p\left(\mathbf{x}_n | \mathcal{H}_1; \hat{\mathbf{d}}^{(i-1)}\right)}{(1-u) \quad p\left(\mathbf{x}_n | \mathcal{H}_0; \hat{\mathbf{f}}^{(i-1)}\right) + u \quad p\left(\mathbf{x}_n | \mathcal{H}_1; \hat{\mathbf{d}}^{(i-1)}\right)}, \tag{43}$$

which has the same form as in (22), but substituting the estimates $\hat{u}^{(i-1)}$ by u. In the M-step, the computation of $\hat{\mathbf{f}}^{(i)}$ and $\hat{\mathbf{d}}^{(i)}$ remains the same as in the standard case.

5.3. Supervised decision vectors

In the standard case, we assumed that the FC infers the model parameters only from the decisions of the available sensors $\mathbf{X}$. Let us now consider the case in which the FC also knows the true hypothesis for a subset of decision vectors $S_{\mathbf{z}} \subseteq \{1, 2, \ldots, N\}$. We call them supervised decision vectors. Now, the available data at the FC is $\mathbf{X}$ and $\{z_n\}_{n \in S_{\mathbf{z}}}$. Then, the expectation function of the E-step will be

$$Q(\Theta; \hat{\Theta}^{(i-1)}) = \underset{\{z_n\}_{n \notin S_{\mathbf{z}}}}{\mathbb{E}} [\log L_c(\Theta) | \mathbf{X}, \{z_n\}_{n \in S_{\mathbf{z}}}; \hat{\Theta}^{(i-1)}].$$

Now, the terms $\{t_n^{(i)}\}_{n \in S_{\mathbf{z}}}$ are no longer soft estimates but the known values, so $t_n^{(i)} = z_n$, whereas the soft estimates for the unsupervised decision vectors will be as in (22). The M-step remains the same.

If all decision vectors were supervised, $S_{\mathbf{z}} = \{1, \ldots, N\}$, the resulting estimate is just the ML estimate for the complete data, $\{\mathbf{X}, \mathbf{z}\}$, which is given by

$$\hat{\Theta} = \{\hat{\mathbf{f}}, \hat{\mathbf{d}}, \hat{u}\} = \underset{\Theta \in S_\Theta}{\operatorname{argmax}} \quad \log L_c(\Theta). \tag{44}$$

Its solution can be viewed as a particular case of the EM algorithm where $\mathbf{t} = \mathbf{z}$. Note that now the terms N_j in (28), $M_{k,j|s}$ in (30) and $M_{k,s}$ in (31) are known in advance, so the solution to (44) is obtained after a single iteration of the M-step.

5.4. Computational complexity

The memory storage requirements of the EM algorithm are negligible, whereas its computational cost (time complexity) depends mainly on the number of iterations of the EM algorithm, which is quite unpredictable. When the detection and false-alarm probabilities of a subset of sensors are known (section 5.1, case 1), they do not need to be estimated in the M-step. When the prior probability is known in advance (section 5.2), Eqs. (28)–(30) do not apply in the M-step. However, this does not result in a significant reduction in computational cost. Finally, when supervised decision vectors exist (section 5.3), their posterior probabilities are known. Therefore, only the posterior probabilities of the unsupervised decision vectors need to be estimated in the E-step. Focusing on a single iteration of the EM algorithm, the computational cost of the E-steps and M-steps is summarized in the following table for the different cases considered in Sections 4 and 5 (see Table 1).

6. Simulation results

Following [5,7,8], we assume that the sensors employ on-off keying signaling to transmit their decisions to the FC through the reporting channels, which experience slow flat fading. Then, at time n, the discrete-time base-band signal at the FC coming from sensor k is given by $r_{k,n} = h_k\, y_{k,n} + w_{k,n}$. We make the standard assumption that the noise at the FC, $w_{k,n}$, is independent, identically distributed, and circularly symmetric complex Gaussian with variance σ_w^2. We assume that the FC knows the channel responses, h_k, so it employs coherent detection to obtain $x_{k,n}$ from $r_{k,n}$ with decision threshold $|h_k|/2$. Then, the bit-error rate is $e_k = Q\left(\frac{|h_k|}{2\sigma_w}\right)$. We also assume that the reporting channels experience independent Rayleigh fading, so $h_k \sim \mathcal{N}_{\mathbb{C}}(0,1)$. Notice that the above assumptions are just the ones we consider in the simulations. They are not required by the proposed method.

We analyze the estimation and detection performance of the EM algorithm using Monte Carlo simulations. Each experiment averages the performance of $R = 10^5$ independent runs. In each experiment, we fix the prior probability u, the number of sensors K, and the number of decision vectors N. We also choose σ_w^2 so that the average error probability of the reporting channels takes a given value $\bar{e} = E[e_k] = E\left[Q\left(\frac{|h_k|}{2\sigma_w}\right)\right]$. These parameters do not change in the R runs of each experiment. Unless otherwise stated, we consider a network comprising $K = 8$ sensors. The prior probability is $u = \frac{1}{3}$ and the average BER of the reporting channels is $E[e_k] = \bar{e} = 0.02$. In any case, the figures show the values of the simulation parameters in the experiments.

In each run, the local sensor probabilities are drawn independently from uniform distributions within the following intervals: $P_{f,k}^{(s)} \sim U(0, 1/2)$, $P_{d,k}^{(s)} \sim U(P_{f,k}^{(s)}, 1)$. We also draw independent channel responses, h_k, and the corresponding error rates e_k. Then, from (1) and (2), we obtain the probabilities $\mathbf{f}$ and $\mathbf{d}$ in the run. Finally, a sequence of states $\mathbf{z}$ is drawn according to u, and the matrix of decisions $\mathbf{X}$ is drawn from $\mathbf{z}$, $\mathbf{f}$ and $\mathbf{d}$. Finally, the EM algorithm is applied to obtain the estimates $\hat{\Theta} = \{\hat{\mathbf{f}}, \hat{\mathbf{d}}, \hat{u}\}$ from $\mathbf{X}$, and the estimation errors are computed. In each run, the EM algorithm is applied only once with the initialization (36) with $\epsilon_t = 10^{-3}$. From the EM estimates, we analytically calculate the detection performance of the resulting EM-based fusion rule using (7). After the R runs of the experiment, the estimation errors and the detection performance are averaged. These are the outputs of the experiment.

In the simulations, unless otherwise indicated, we consider the standard case where the FC does not know any model parameters $\Theta = \{\mathbf{f}, \mathbf{d}, u\}$, so it must estimate them from $\mathbf{X}$ exclusively. In addition, if not otherwise stated, we assume that the decision matrix is complete, so the FC receives the decisions from all sensors in each sensing period.

6.1. Estimation performance

To analyze the estimation performance of the EM algorithm, we compare it with the fully supervised estimator (see Section 5.3), which can be considered as an upper bound for the EM estimator. We use the root mean squared error (RMSE) as a performance metric:

$$\mathrm{RMSE}(\hat{\mathbf{f}}) = \sqrt{\frac{1}{K R} \sum_{r=1}^{R} \|\hat{\mathbf{f}}^{(r)} - \mathbf{f}^{(r)}\|^2},$$

where r is the index of the runs, $\mathbf{f}^{(r)}$ is the vector of the false-alarm probabilities in run r, and $\hat{\mathbf{f}}^{(r)}$ is the corresponding estimate. The RMSE of $\hat{\mathbf{d}}$ and $\hat{u}$ is defined analogously.

Fig. 3 compares the estimation performance of the EM estimator and the fully supervised estimator (labeled S) as a function of the number of decision vectors. As expected, the S estimator outperforms the EM, and the RMSEs decrease with N more rapidly when N is low.

Table 1
Computational cost per iteration of the EM algorithm.

	E-step	M-step
Section 4: Standard problem	$O(N)$	$O(K)$
Section 5.1: Known performance of some sensors	$O(N)$	$O(\|S_{\mathbf{f}} \cap S_{\mathbf{d}}\|)$
Section 5.2: Known prior probability	$O(N)$	$O(K)$
Section 5.3: Supervised decision vectors	$O(N - \|S_z\|)$	$O(K)$

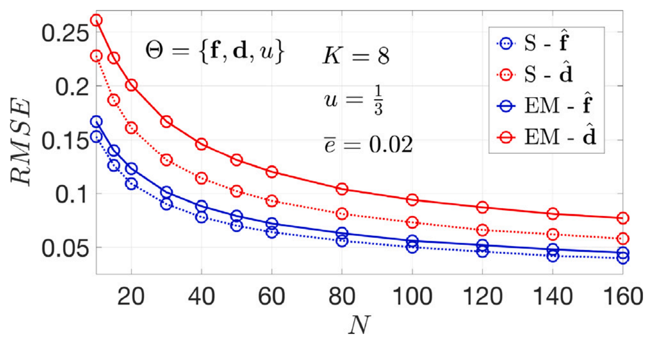

Fig. 3. RMSE of the EM and supervised estimates as a function of the number of decision vectors.

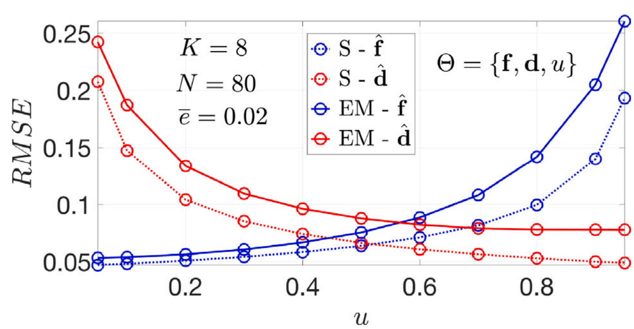

Fig. 4. RMSE of the EM and supervised estimates as a function of the prior probability of $\mathcal{H}_1$.

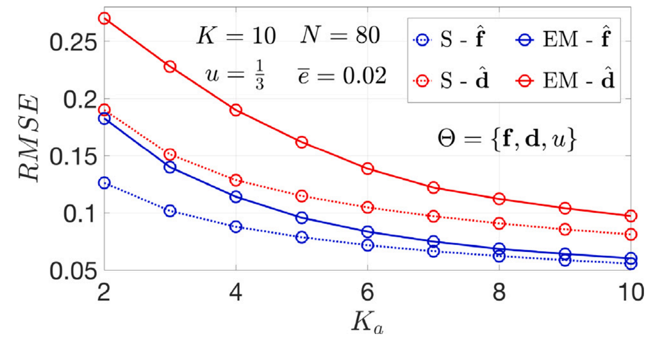

Fig. 5. RMSE of the EM estimates and the supervised estimates as function of the number of active sensors.

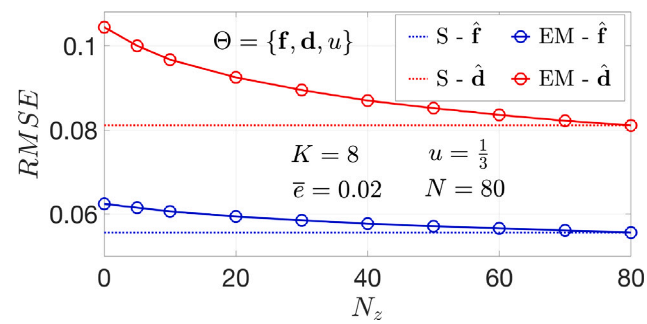

Fig. 6. RMSE as a function of the number of supervised decision vectors.

Fig. 4 shows the performance of the EM estimator and the fully supervised estimator as a function of the prior probability of $\mathcal{H}_1$. For low values of u, most of the decision vectors come from $\mathcal{H}_0$, so the RMSE of the $\mathbf{f}$ estimates is small, while the RMSE of the $\mathbf{d}$ estimates is high. The opposite occurs for high values of u. In any case, after $N = 80$ decision vectors, the performance gap between the S and the EM estimators is quite low for any value of u.

Fig. 5 shows the RMSE of the estimates when a subset of K_a sensors, out of $K = 10$, are active in each sensing period. In other words, $K_n = K_a$ for all n. The active sensors are randomly selected in each sensing period. Now, it is more difficult for the FC to learn the sensor parameters because it has fewer decisions from each sensor. It can be observed that the RMSE of the EM estimates converges quite fast to the supervised ones as the number of active sensors grows.

Fig. 6 shows the estimation performance when N_z of the $N = 80$ decision vectors are supervised (see Section 5.3). It is observed that the RMSEs of the EM estimates decrease rather slowly with N_z, so a high fraction of supervised decision vectors is required to obtain a significant improvement.

As it was mentioned, the detection performance of the fusion rule proposed in Section 3 is determined by the accuracy of the test statistic estimates $t(\mathbf{x};\hat{\Theta})$ which, in turn, is determined by the accuracy of the parameter estimates $\hat{\Theta}$. The RMSE of the test statistic estimates is given by

$$\mathrm{RMSE}(\hat{t}) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\sum_{\mathbf{x}\in\Omega_{\mathbf{x}}} |t(\mathbf{x};\hat{\Theta}^{(r)}) - t(\mathbf{x};\Theta^{(r)})|^2 \; p(\mathbf{x};\Theta^{(r)})}.$$

Fig. 7 shows the RMSE of $t(\mathbf{x};\hat{\Theta})$ as a function of the number of decision vectors. The simulation parameters are as in Fig. 3. As expected, according to Fig. 3, the RMSEs of the test statistic estimates decrease faster for low values of N. Interestingly, the performance gap between the supervised and the EM estimates remains quite constant in the entire range of N.

6.2. Detection performance

This subsection shows the detection performance of the EM-based detection rule (proposed in Section 3.2 and labeled EMR in the figures). It is compared with the fusion rule based on fully supervised estimates of the model parameters (labeled SR in the figures), and with the optimal fusion rule (10) (labeled LRT in the figures) when the FC knows the model parameters. The performance of the LRT and the SR are both upper bounds for the EMR, with SR being the tightest one. We also compare the EMR with the counting rules (CRs) with different values of decision threshold. If not otherwise stated, we use the probability of error (9) as a performance metric. The detection performance of the ISR is the worst in all experiments, so it is not shown in the figures.

Fig. 8 shows the average probability of error of the fusion rules as functions of the prior probability of $\mathcal{H}_1$. As expected, the optimal

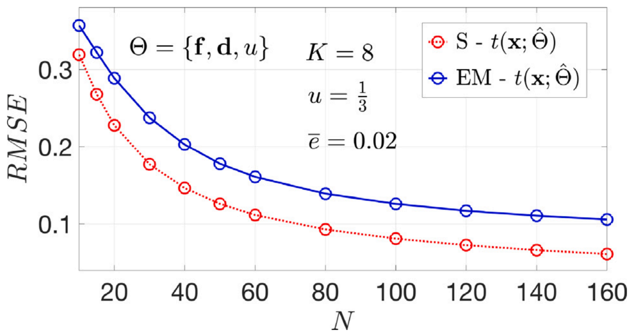

Fig. 7. RMSE of the test statistic estimates $t(\mathbf{x}; \hat{\Theta})$ as a function of the number of decision vectors.

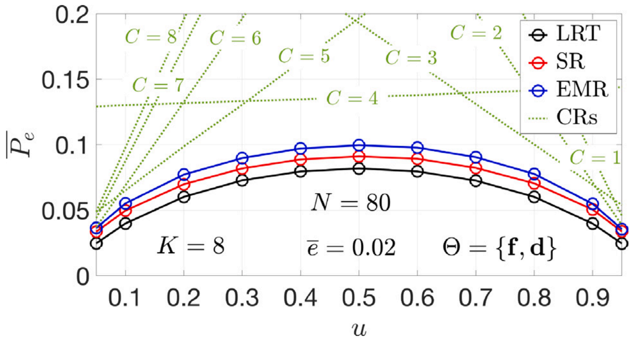

Fig. 8. Average probability of error of the fusion rules as a function of the prior probability of $\mathcal{H}_1$.

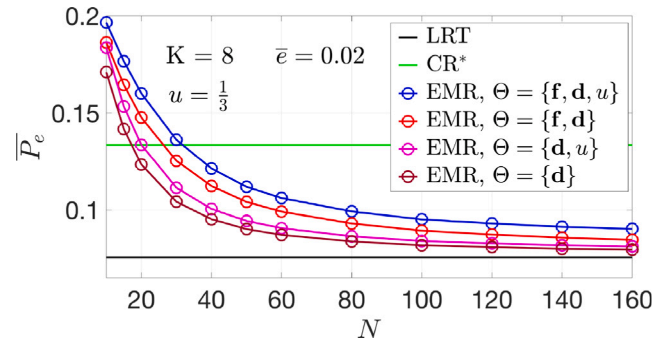

Fig. 9. Average probability of error of the EMR when some of the model parameters are known.

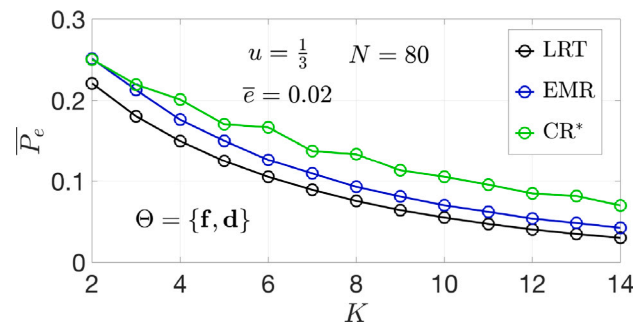

Fig. 10. Average probability of error for different number of sensors.

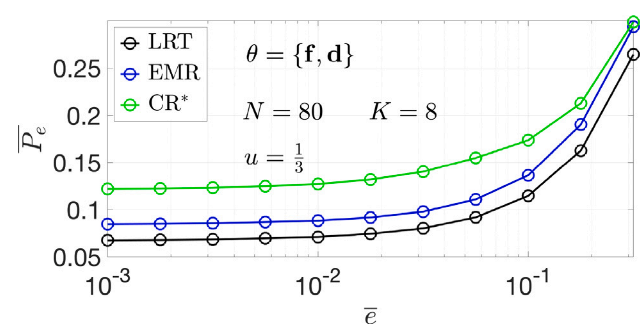

Fig. 11. Average probability of error as a function of the average crossover probability of the BSCs.

decision threshold of the CR (denoted by C) is highly dependent on the value of u. Note that, if the model parameters are unknown, the FC cannot know the optimal C. In any case, after $N = 80$ decision vectors, the EMR outperforms all CRs over the entire range of u values. The performance gaps between the EMR, the LRT and the SR remain quite constant for any value of u.

Fig. 9 shows the average probability of error of the EMR, as a function of the number of decision vectors, when different subsets of model parameters are known to the FC. As expected, the more parameters the FC knows, the better the EMR performs. The EMR is compared with the LRT and the optimal CR (labeled CR*). In this case, the optimal CR threshold is $C = 4$, as Fig. 8 shows. Notice that the EMR outperforms the CR* in all cases after $N = 30$ decision vectors. As N increases, the P_e of the EMR converges to the LRT more rapidly when more parameters are known by the FC.

Fig. 10 compares the average probability of error of the fusion rules, as a function of the number of sensors. It can be observed that, as K increases, the performance of the EMR detector converges quite fast to the LRT. In contrast, the performance gap between the CR* and the LRT remains fairly constant as K increases.

Fig. 11 shows the average probability of error of the fusion rules as a function of the average error probability of the BSCs. In this case, the best CR threshold is again $C = 4$. The EMR outperforms the CR* except for extremely large values of $\bar{e}$. Moreover, the P_e gap between the EMR and the SR and LRT remains almost constant with $\bar{e}$. In other words, the EMR fusion rule is quite robust against the errors in the BSCs.

Fig. 12 shows the average Bayes risk (8) as a function of the ratio of the Bayesian costs assigned to false alarms and miss-detections, $C_{1,0}/C_{0,1}$. The hits are not penalized ($C_{0,0} = C_{1,1} = 0$), so the Bayes risk when $C_{1,0} = C_{0,1}$ coincides with the probability of error. The figure shows that the optimal CR highly depends on the costs ratio. The EMR outperforms the optimal CR except for extreme values of $C_{1,0}/C_{0,1}$. Interestingly, the higher performance gap between the EMR and the LRT occurs when the costs ratio is close to 1 (minimum P_e criterion), and decreases as $C_{1,0}/C_{0,1}$ deviates from 1. The little asymmetry with respect $C_{1,0}/C_{0,1} = 1$ is due to the value of u.

Fig. 13 shows the detection performance of the EMR when the FC knows the operational point (f_k, d_k) of $K_{f,d}$ sensors out of K. In particular, it shows the average probability of error after learning the unknown model parameters for $N = 80$ and $N = 40$ decision vectors. In this case, u is unknown. As expected, knowing the operational point of some sensors produces a detection improvement and reduces the performance gap between the LRT and the EMR. Even when the FC knows the operating points of all sensors, $K_{f,d} = K$, the EMR does not achieve the LRT performance because the prior probability u is estimated.

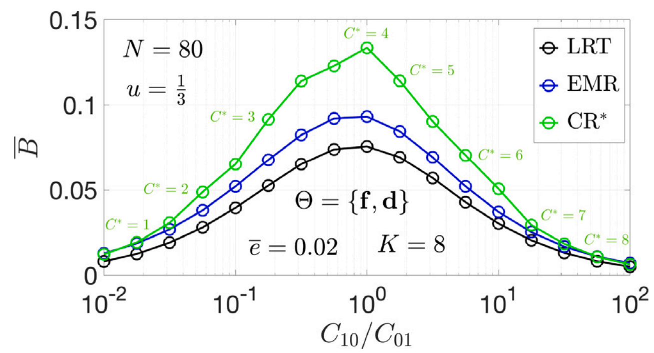

Fig. 12. Average Bayes Risk as a function of the ratio between the costs assigned to false alarms and miss-detections.

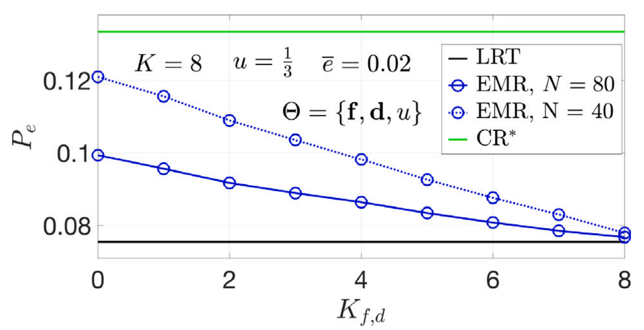

Fig. 13. Average probability of error of the EMR when the FC knows the operational point of $K_{f,d}$ of the sensors.

7. Conclusion

We have presented an algorithm for the FC to blindly estimate the sensor parameters in canonical DD-WSN with BSCs. It is flexible and generally applicable in the sense that it can cover a wide variety of situations that may arise in practical DD-WSN. It is applicable when all or some of the parameters of a subset of sensors are unknown to the FC. In addition, the FC may know or ignore the prior probabilities of the hypotheses in advance. The algorithm is also applicable when only a subset of sensors reports their decisions in each sensing period, and when the FC knows the true hypothesis for a subset of sensing periods. All these cases can occur simultaneously.

The algorithm is derived by applying the EM algorithm, where the observed variables are the noisy sensor decisions at the FC, and the latent variables are the unknown true hypotheses at each sensing period. The application of the EM algorithm leads to quite simple closed-form expressions for both the E-step and the M-step, so its computational cost is quite low. Numerical simulations show the accuracy of the estimates after sufficient sensing periods.

In DD-WSN, the ultimate goal is to have an accurate fusion rule of the sensor decisions. In this work, we have proposed a fusion rule considering the MBR criterion. It is based on the estimates of the posterior probability of the alternative hypothesis obtained from the EM estimates of the model parameters. Exhaustive numerical simulations show that, in all cases, the proposed fusion rules exhibit near-optimal detection performance when the estimates of the unknown model parameters are sufficiently accurate. And this is achieved after a rather small number of sensing periods.

In closing, we mention a possible extension of our work. We have considered a stationary environment in which sensors' performance does not change with time, so we use a batch EM algorithm to estimate them. But in many practical DD-WSN systems, the sensors' detection probabilities can vary with time. In these cases, some online version of the EM algorithm could be used.

CRediT authorship contribution statement

J. Perez: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **I. Santamaria:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **A. Pagés-Zamora:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. M-step of the em algorithm when f_k and d_k are unknown

The feasible set $\mathcal{S}_k$ is determined by the constraints $e_k \le f_k \le d_k \le 1 - e_k$. Therefore, (33) is a two-variable nonlinear optimization problem with three inequality constraints. The feasible set is compact (closed and bounded) and convex. The objective function of (33) is differentiable, bounded, and strictly concave in $\mathcal{S}_k$. Therefore, it has only one maximum in $\mathcal{S}_k$, which is the only solution to the optimization problem. Since there are no irregular points associated with the constraints, the solution to (33) is the only point (f_k, d_k) that satisfies the Karush–Kuhn–Tucker (KKT) conditions. The gradient of the objective function is

$$\left[\frac{M^{(i)}_{k,1|0}}{f_k} - \frac{M^{(i)}_{k,0|0}}{1-f_k}, \quad \frac{M^{(i)}_{k,1|1}}{d_k} - \frac{M^{(i)}_{k,0|1}}{1-d_k} \right]^T. \tag{45}$$

The inequality constraint functions are $f_k - d_k$, $e_k - f_k$ and $d_k - (1 - e_k)$, and their gradients are $[1, -1]^T$, $[-1, 0]^T$ and $[0, 1]^T$, respectively. Then, the KKT conditions are

$$-\frac{M^{(i)}_{k,1|0}}{f_k} + \frac{M^{(i)}_{k,0|0}}{1-f_k} + \lambda_1 - \lambda_2 = 0, \qquad -\frac{M^{(i)}_{k,1|1}}{d_k} + \frac{M^{(i)}_{k,0|1}}{1-d_k} - \lambda_1 + \lambda_3 = 0,$$
$$\lambda_1(f_k - d_k) = 0, \qquad \lambda_2(e_k - f_k) = 0, \qquad \lambda_3(d_k - (1 - e_k)) = 0,$$
$$f_k \le d_k, \qquad e_k \le f_k, \qquad d_k \le 1 - e_k, \qquad \lambda_1, \lambda_2, \lambda_3 \ge 0,$$

where λ_1, λ_2 and λ_3 are the KKT multipliers. In the following lines, we particularize the KKT conditions to the different types of points (f_k, d_k) on the feasible set (see Fig. 2):

(1) Interior points: $e_k < f_k < d_k < 1 - e_k \Rightarrow \lambda_1 = \lambda_2 = \lambda_3 = 0$. Then the KKT conditions reduce to

$$-\frac{M^{(i)}_{k,1|0}}{f_k} + \frac{M^{(i)}_{k,0|0}}{1-f_k} = 0, \quad -\frac{M^{(i)}_{k,1|1}}{d_k} + \frac{M^{(i)}_{k,0|1}}{1-d_k} = 0, \quad f_k < d_k, \quad e_k < f_k, \quad d_k < 1 - e_k.$$

From the two equations, we obtain $\hat{f}^{(i)}_k = \frac{M^{(i)}_{k,1|0}}{M^{(i)}_{k,0}}$, $\hat{d}^{(i)}_k = \frac{M^{(i)}_{k,1|1}}{M^{(i)}_{k,1}}$, where we have considered (31). The inequalities require $e_k < \frac{M^{(i)}_{k,1|0}}{M^{(i)}_{k,0}} < \frac{M^{(i)}_{k,1|1}}{M^{(i)}_{k,1}} < 1 - e_k$.

(2) Points on the boundary $e_k < f_k = d_k < 1 - e_k \Rightarrow \lambda_2 = \lambda_3 = 0$. Then the KKT conditions reduce to

$$-\frac{M^{(i)}_{k,1|0}}{f_k} + \frac{M^{(i)}_{k,0|0}}{1-f_k} + \lambda_1 = 0, \qquad -\frac{M^{(i)}_{k,1|1}}{d_k} + \frac{M^{(i)}_{k,0|1}}{1-d_k} - \lambda_1 = 0,$$

$$f_k = d_k,\quad e_k < f_k,\quad d_k < 1-e_k,\quad \lambda_1 \geq 0.$$

From the equations we have $\hat{f}_k^{(i)} = \hat{d}_k^{(i)} = \frac{M_{k,1|0}^{(i)}+M_{k,1|1}^{(i)}}{M_k} = \frac{1}{M_k}\sum_{n=1}^{N} l_{k,n}\, x_{k,n}$, where we have considered (30), (31) and (32). The inequalities require $e_k < \frac{1}{M_k}\sum_{n=1}^{N} l_{k,n}\, x_{k,n} < 1-e_k$ and $\frac{M_{k,1|0}^{(i)}}{M_{k,0}^{(i)}} \geq \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}}$.

(3) Points on the boundary $e_k = f_k < d_k < 1-e_k \Rightarrow \lambda_1 = \lambda_3 = 0$. Then the KKT conditions reduce to

$$-\frac{M_{k,1|0}^{(i)}}{f_k} + \frac{M_{k,0|0}^{(i)}}{1-f_k} - \lambda_2 = 0,\quad -\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} = 0,$$
$$f_k = e_k,\quad f_k < d_k,\quad d_k < 1-e_k,\quad \lambda_2 \geq 0,$$

From the equations we have $\hat{f}_k^{(i)} = e_k$, $\hat{d}_k^{(i)} = \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}}$, and the inequalities require $\frac{M_{k,1|0}^{(i)}}{M_{k,0}^{(i)}} \leq e_k < \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}} < 1-e_k$.

(4) Points on the boundary $e_k < f_k < d_k = 1-e_k \Rightarrow \lambda_1 = \lambda_2 = 0$. Then the KKT conditions reduce to

$$-\frac{M_{k,1|0}^{(i)}}{f_k} + \frac{M_{k,0|0}^{(i)}}{1-f_k} = 0,\quad -\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} + \lambda_3 = 0,$$
$$e_k < f_k,\quad f_k < d_k,\quad d_k = 1-e_k,\quad \lambda_3 \geq 0.$$

From the equations $\hat{f}_k^{(i)} = \frac{M_{k,1|0}^{(i)}}{M_{k,0}^{(i)}}$, $\hat{d}_k^{(i)} = 1-e_k$, and the inequalities require $e_k < \frac{M_{k,1|0}^{(i)}}{M_{k,0}^{(i)}} < 1-e_k \leq \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}}$.

(5) The vertex $(f_k, d_k) = (e_k, e_k) \Rightarrow \lambda_3 = 0$. Then the KKT conditions reduce to

$$-\frac{M_{k,1|0}^{(i)}}{f_k} + \frac{M_{k,0|0}^{(i)}}{1-f_k} + \lambda_1 - \lambda_2 = 0,\quad -\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} - \lambda_1 = 0,$$
$$e_k = f_k = d_k,\quad \lambda_1, \lambda_2 \geq 0.$$

The inequalities require $e_k \geq \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}}$, $e_k \geq \frac{1}{M_k}\sum_{n=1}^{N} l_{k,n}\, x_{k,n}$

(6) The vertex $(f_k, d_k) = (1-e_k, 1-e_k) \Rightarrow \lambda_2 = 0$. Then the KKT conditions reduce to

$$-\frac{M_{k,1|0}^{(i)}}{f_k} + \frac{M_{k,0|0}^{(i)}}{1-f_k} + \lambda_1 = 0,\quad -\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} - \lambda_1 + \lambda_3 = 0,$$
$$f_k = d_k = 1-e_k,\quad \lambda_1, \lambda_3 \geq 0.$$

The inequalities require $1-e_k \leq \frac{M_{k,1|0}^{(i)}}{M_{k,0}^{(i)}}$, $1-e_k \leq \frac{1}{M_k}\sum_{n=1}^{N} l_{k,n}\, x_{k,n}$.

(7) The vertex $(f_k, d_k) = (e_k, 1-e_k) \Rightarrow \lambda_1 = 0$. Then the KKT conditions reduce to

$$-\frac{M_{k,1|0}^{(i)}}{f_k} + \frac{M_{k,0|0}^{(i)}}{1-f_k} - \lambda_2 = 0,\quad -\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} + \lambda_3 = 0,$$
$$f_k = e_k,\quad d_k = 1-e_k,\quad \lambda_2, \lambda_3 \geq 0.$$

The inequalities require $e_k \geq \frac{M_{k,1|0}^{(i)}}{M_{k,0}^{(i)}}$, $1-e_k \leq \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}}$.

Appendix B. M-step of the em algorithm when f_k is known and d_k is unknown

The optimization problem (39) is a single-variable nonlinear problem with two inequality constraints: $f_k - d_k \leq 0$ and $d_k - (1-e_k) \leq 0$. The derivatives of the objective function and the inequality constraint functions are

$$\frac{M_{k,1|1}^{(i)}}{d_k} - \frac{M_{k,0|1}^{(i)}}{1-d_k},\quad -1,\quad 1,$$

respectively. The feasible set, $[f_k, 1-e_k]$, is compact and convex, and the objective function is differentiable, bounded, and strictly concave in the feasible set. Then (39) has only one maximum, which is the only solution. Since there are no irregular points associated with the constraints, the solution to (39) is the only value d_k that fulfills the Karush–Kuhn–Tucker (KKT) conditions:

$$-\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} - \lambda_1 + \lambda_2 = 0,\quad \lambda_1(f_k - d_k) = 0,\quad \lambda_2(d_k - (1-e_k)) = 0,$$
$$f_k \leq d_k,\quad d_k \leq 1-e_k,\quad \lambda_1, \lambda_2 \geq 0,$$

where λ_1 and λ_2 are the KKT multipliers associated with the inequality constraints. We particularize the KKT conditions for the possible values of d_k in the feasible set.

(1) Interior point: $f_k < d_k < 1-e_k \Rightarrow \lambda_1 = \lambda_2 = 0$. The KKT conditions reduce to

$$-\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} = 0,\qquad f_k < d_k,\qquad d_k < 1-e_k.$$

From the equation, we obtain $\hat{d}_k^{(i)} = \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}}$, and the inequalities require $f_k < \frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}} < 1-e_k$.

(2) Point on the boundary $d_k = f_k \Rightarrow \lambda_2 = 0$. The KKT conditions reduce to

$$-\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} - \lambda_1 = 0,\qquad f_k = d_k,\qquad \lambda_1 \geq 0.$$

Then, $\hat{d}_k^{(i)} = f_k$, and the inequality requires $\frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}} \leq f_k$.

(3) Point on the boundary $d_k = 1-e_k \Rightarrow \lambda_1 = 0$. The KKT conditions reduce to

$$-\frac{M_{k,1|1}^{(i)}}{d_k} + \frac{M_{k,0|1}^{(i)}}{1-d_k} + \lambda_2 = 0,\qquad d_k = 1-e_k,\qquad \lambda_2 \geq 0.$$

Then, $\hat{d}_k^{(i)} = 1-e_k$, and the inequality requires $\frac{M_{k,1|1}^{(i)}}{M_{k,1}^{(i)}} \geq 1-e_k$.

Appendix C. M-step of the em algorithm when f_k is unknown and d_k is known

The optimization problem (41) is a single-variable nonlinear problem with two inequality constraints: $e_k - f_k \leq 0$ and $f_k - d_k \leq 0$. The derivatives of the objective function and the inequality constraint functions are

$$\frac{M_{k,1|0}^{(i)}}{f_k} - \frac{M_{k,0|0}^{(i)}}{1-f_k},\quad -1,\quad 1,$$

respectively. The feasible set, $[e_k, d_k]$, is compact and convex, and the objective function is differentiable, bounded, and strictly concave in the feasible set. Then (41) has only one maximum, which is the only solution. Since there are no irregular points associated with the constraints, the solution to (41) is the only value d_k that fulfills the Karush–Kuhn–Tucker (KKT) conditions:

$$-\frac{M_{k,1|0}^{(i)}}{f_k} + \frac{M_{k,0|0}^{(i)}}{1-f_k} - \lambda_1 + \lambda_2 = 0,\quad \lambda_1(e_k - f_k) = 0,\quad \lambda_2(f_k - (d_k)) = 0,$$
$$e_k \leq f_k,\quad f_k \leq d_k,\quad \lambda_1, \lambda_2 \geq 0,$$

where λ_1 and λ_2 are the KKT multipliers associated with the inequality constraints. We particularize the KKT conditions for the possible values of f_k in the feasible set.

(1) Interior point: $e_k < f_k < d_k \Rightarrow \lambda_1 = \lambda_2 = 0$. Then the KKT conditions reduce to

$$-\frac{M^{(i)}_{k,1|0}}{f_k} + \frac{M^{(i)}_{k,0|0}}{1-f_k} = 0, \qquad e_k < f_k, \qquad f_k < d_k.$$

From the equation, we obtain $\hat{f}^{(i)}_k = \frac{M^{(i)}_{k,1|0}}{M^{(i)}_{k,0}}$, and the inequalities require $e_k < \frac{M^{(i)}_{k,1|0}}{M^{(i)}_{k,0}} < d_k$.

(2) Point on the boundary $f_k = e_k \Rightarrow \lambda_2 = 0$. The KKT conditions reduce to

$$-\frac{M^{(i)}_{k,1|0}}{f_k} + \frac{M^{(i)}_{k,0|0}}{1-f_k} - \lambda_1 = 0, \qquad f_k = e_k, \qquad \lambda_1 \geq 0.$$

Then, $\hat{f}^{(i)}_k = e_k$, and the inequality requires $\frac{M^{(i)}_{k,1|0}}{M^{(i)}_{k,0}} \leq e_k$.

(3) Point on the boundary $f_k = d_k \Rightarrow \lambda_1 = 0$. The KKT conditions reduce to

$$-\frac{M^{(i)}_{k,1|0}}{f_k} + \frac{M^{(i)}_{k,0|0}}{1-f_k} + \lambda_2 = 0, \qquad f_k = d_k, \qquad \lambda_2 \geq 0.$$

Then, $\hat{f}^{(i)}_k = d_k$, and the inequality requires $\frac{M^{(i)}_{k,1|0}}{M^{(i)}_{k,0}} \geq d_k$.

Data availability

No data was used for the research described in the article.

References

[1] P.K. Varshney, Distributed Detection and Data Fusion, Springer New York, 1997.
[2] D. Ciuonzo, P. Salvo Rossi, Distributed detection of a non-cooperative target via generalized locally-optimum approaches, Inf. Fusion 36 (2017) 261–274.
[3] D. Ciuonzo, P. Salvo Rossi, P.K. Varshney, Distributed detection in wireless sensor networks under multiplicative fading via generalized score tests, IEEE Internet Things J. 8 (11) (2021) 9059–9071.
[4] S. Chaudhari, J. Lunden, V. Koivunen, H.V. Poor, Cooperative sensing with imperfect reporting channels: Hard decisions or soft decisions? IEEE Trans. Signal Process. 60 (1) (2012) 18–28.
[5] J.Y. Wu, C.W. Wu, T.Y. Wang, T.S. Lee, Channel-aware decision fusion with unknown local sensor detection probability, IEEE Trans. Signal Process. 58 (3) (2010) 1457–1463.
[6] Q. Cheng, B. Chen, P.K. Varshney, Detection performance limits for distributed sensor networks in the presence of nonideal channels, IEEE Trans. Wirel. Commun. 5 (11) (2006) 3034–3038.
[7] F.R. Mohammad, D. Ciuonzo, Z.A.K. Mohammed, Mean-based blind hard decision fusion rules, IEEE Signal Process. Lett. 25 (5) (2018) 630–634.
[8] D. Ciuonzo, P. Salvo Rossi, Decision fusion with unknown sensor detection probability, IEEE Signal Process. Lett. 21 (2) (2014) 208–212.
[9] B. Chen, L. Tong, P.K. Varshney, Channel-aware distributed detection in wireless sensor networks, IEEE Signal Process. Mag. 23 (4) (2006) 16–26.
[10] Gaoyuan Zhang, Kai Chen, Congfang Ma, Sravan Kumar Reddy, Baofeng Ji, Yongen Li, Congzheng Han, Xiaohui Zhang, Zhumu Fu, Decision fusion for multi-route and multi-hop wireless sensor networks over the binary symmetric channel, Comput. Commun. 196 (2022) 167–183.
[11] N.R. Banavathu, M.Z.A. Ali, Optimization of n-out-of-k rule for heterogeneous cognitive radio networks, IEEE Signal Process. Lett. 26 (3) (2019) 445–449.
[12] B. Chen, P.K. Willett, On the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels, IEEE Trans. Inform. Theory 51 (2) (2005) 693–699.
[13] A. Kashyap, Comments on on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels, IEEE Trans. Inform. Theory 52 (3) (2006) 1274–1275.
[14] H. Chen, B. Chen, P.K. Varshney, Further results on the optimality of the likelihood-ratio test for local sensor decision rules in the presence of nonideal channels, IEEE Trans. Inform. Theory 55 (2) (2009) 828–832.
[15] Z. Chair, P.K. Varshney, Optimal data fusion in multiple sensor detection systems, IEEE Trans. Aerosp. Electron. Syst. AES-22 (1) (1986) 98–101.
[16] S.C.A. Thomopoulos, R. Viswanathan, D.C. Bougoulias, Optimal decision fusion in multiple sensor systems, IEEE Trans. Aerosp. Electron. Syst. AES-23 (5) (1987) 644–653.
[17] B. Chen, R. Jiang, T. Kasetkasem, P.K. Varshney, Channel aware decision fusion in wireless sensor networks, IEEE Trans. Signal Process. 52 (12) (2004) 3454–3458.
[18] R. Jiang, B. Chen, Fusion of censored decisions in wireless sensor networks, IEEE Trans. Wirel. Commun. 4 (6) (2005) 2668–2673.
[19] R. Niu, B. Chen, P.K. Varshney, Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks, IEEE Trans. Signal Process. 54 (3) (2006) 1018–1027.
[20] K.B. Letaief, W. Zhang, Cooperative communications for cognitive radio networks, Proc. IEEE 97 (5) (2009) 878–893.
[21] W. Zhang, R.K. Mallik, K.B. Letaief, Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks, IEEE Trans. Wirel. Commun. 8 (12) (2009) 5761–5766.
[22] A. Lei, R. Schober, Coherent max-log decision fusion in wireless sensor networks, IEEE Trans. Commun. 58 (5) (2010) 1327–1332.
[23] S.C.A. Thomopoulos, L. Zhang, Distributed decision fusion in the presence of networking delays and channel errors, Inform. Sci. 66 (1992) 91–118.
[24] D. Ciuonzo, G. Romano, P. Salvo Rossi, Channel-aware decision fusion in distributed MIMO wireless sensor networks: Decode-and-fuse vs. decode-then-fuse, IEEE Trans. Wirel. Commun. 11 (8) (2012) 2976–2985.
[25] S.M. Kay, Fundamentals of Statistical Signal Processing. Detection Theory, in: Signal Processing Series, Prentice Hall, 1998.
[26] H.V. Poor, An Introduction to Signal Detection and Estimation, Springer-Verlag, 1994.
[27] D. Ciuonzo, A. De Maio, P. Salvo Rossi, A systematic framework for composite hypothesis testing of independent bernoulli trials, IEEE Signal Process. Lett. 22 (9) (2015) 1249–1253.
[28] A.S. Gupta, L. Vermeire, Locally optimal tests for multiparameter hypotheses, J. Amer. Statist. Assoc. 81 (395) (1986) 819–825.
[29] B. Liu, J. Aleksandar, K.M. Wong, Optimal distributed detection of multiple hypotheses using blind algorithm, IEEE Trans. Aerosp. Electron. Syst. 47 (1) (2011) 317–331.
[30] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[31] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood for incomplete data via the EM algorithm, J. R. Stat. Soc. Ser. B (Methodological) 39 (1) (1977) 1–38.
[32] G.J. McLachlan, T. Krishnan, The EM Algorithm and Extensions, John Wiley and Sons, 2008.
[33] G. Mirjalily, L. Zhi-Quan, T.N. Davidson, E. Bosse, Blind adaptive decision fusion for distributed detection, IEEE Trans. Aerosp. Electron. Syst. 39 (1) (2003) 34–52.
[34] F. Parisi, F. Strino, B. Nadler, Y. Kluger, Ranking and combining multiple predictors without labeled data, Proc. Natl. Acad. Sci. 111 (4) (2014) 1253–1258.
[35] A. Jaffe, B. Nadler, Y. Kluger, Estimating the accuracies of multiple classifiers without labeled data, in: Proceedings of the 18th International Conference on Artificial Intelligence and Statistics, vol. 2, 2015, pp. 407–415.
[36] Panagiotis A. Traganitis, Alba Pagès-Zamora, Georgios B. Giannakis, Blind multiclass ensemble classification, IEEE Trans. Signal Process. 66 (18) (2018) 4737–4752.
[37] A.P. Dawid, A.M. Skene, Maximum likelihood estimation of observed error-rates using the EM algorithm, Appl. Stat. 28 (1) (1979) 20–28.
[38] Gaoyuan Zhang, Yongen Li, Baofeng Ji, Kai Chen, Yu Mu, Weiguang Wang, M-ary distributed decision fusion for multihop relay wireless sensor networks: Decision fusion rule, implementation framework, and performance analysis, IEEE Internet Things J. 11 (18) (2024) 29569–29587.
[39] R. Viswanathan, P.K. Varshney, Distributed detection with multiple sensors, Proc. IEEE 85 (1) (1997) 54–63.
[40] G.J. McLachlan, D. Peel, Finite Mixture Models, John Wiley & Sons, 2000.
[41] D. Nettleton, Convergence properties of the EM algorithm in constrained parameter spaces, Can. J. Stat./Rev. Can. Stat. 27 (3) (1999) 639–648.