Distributed Estimation and Detection for Sensor Networks Using Hidden Markov Random Field Models[†]

Aleksandar Dogandžić and *Benhong Zhang*
ECpE Department, Iowa State University
3119 Coover Hall, Ames, IA 50011
Phone: (515) 294-0500, Fax: (515) 294-8432
email: {ald,zhangbh}@iastate.edu

Abstract

We develop a hidden Markov random field (HMRF) framework for distributed signal processing in sensor-network environments. Under this framework, spatially distributed observations collected at the sensors form a noisy realization of an underlying random field that has a simple structure with Markovian dependence. We derive iterated conditional modes (ICM) algorithms for distributed estimation of the hidden random field from the noisy measurements. We consider both parametric and nonparametric measurement-error models. The proposed distributed estimators are computationally simple, applicable to a wide range of sensing environments, and *localized*, implying that the nodes communicate only with their neighbors to obtain the desired results. We also develop a calibration method for estimating Markov random field (MRF) model parameters from training data and discuss initialization of the ICM algorithms. The HMRF framework and ICM algorithms are applied to event-region detection. Numerical simulations demonstrate the performance of the proposed approach.

I. Introduction

RECENT advances in integrated sensor and radio-frequency (RF) technologies, wireless communications, and signal processing allow development of sensor-network systems composed of low-cost sensor-processor elements (nodes) jointly working to collect and analyze *noisy* spatio-temporal measurements. Large-scale sensor networks that can monitor an environment at close range with high spatial and temporal resolutions are expected to play an important role in various applications, including assessing "health" of machines, aerospace vehicles, and civil-engineering structures; environmental, medical, food-safety, and habitat monitoring; energy management, inventory control, home and building automation, see also [1, Ch. 1.3] and [2]–[8]. Each node will have limited sensing, signal processing, and communication capabilities, but by cooperating with each other they will accomplish tasks that are difficult to perform with conventional centralized sensing systems [7], [8]. Sensor networks are expected to reveal previously unobservable phenomena in the physical world [8] and are currently attracting considerable attention.

[†]This work was supported by the National Science Foundation under Grant CCF-0545571 and the NSF Industry-University Cooperative Research Program, Center for Nondestructive Evaluation (CNDE), Iowa State University.

Markov random field (MRF) models have been widely used to describe spatially distributed random phenomena, see e.g. [9]–[11]. In this paper (see also [12]), we propose a hidden Markov random field (HMRF) framework for distributed signal processing in sensor-network environments. Under this framework, spatially distributed observations (collected at the sensors) form a noisy realization of a random field with Markovian dependence structure.[1] Previous work on distributed HMRF based signal processing for sensor networks focused on developing message passing algorithms for linear Gaussian measurement-error and MRF process models with known model parameters, see also the discussion in Section II-A. In contrast, our HMRF framework allows for general measurement and random-field models with *unknown* measurement-error model parameters. The unknown measurement-error model parameters vary from one node to another, thus taking into account imperfect calibration of the sensors at different nodes and permitting distributed localized processing and nonparametric measurement-error modeling. The nonparametric measurement-error models that we employ are important in practical applications where accurate parametric models are difficult to find, especially in large-scale sensor networks operating in time-varying environments [15]–[17].

We derive iterated conditional modes (ICM) algorithms for distributed estimation of a localized phenomenon (modeled as a *hidden random field*) from noisy measurements. In particular, the proposed ICM algorithms are designed to increase the *predictive likelihood* of the hidden field.[2] The underlying *distributed-processing paradigm* ensures robustness and reliability of the proposed approach. We demonstrate our approach by applying it to event-region detection, which is an important task in wireless sensor networks [14]. We consider parametric Gaussian and nonparametric (empirical likelihood and entropy) measurement-error models and utilize an autologistic MRF process model for event-region detection.

The HMRF framework is introduced in Section II and general ICM method is presented in Section III. We discuss the event-region detection problem in Section IV where we first propose suitable measurement-error and random-field models (Sections IV-A and IV-B) and then derive the corresponding ICM detection algorithms (Sections IV-C and IV-D). Initialization of the ICM iterations is discussed in Sections IV-C.1 and IV-D.1. In Section V, we develop a *pseudo predictive likelihood* (PPL) calibration method for estimating MRF model parameters from training data and specialize it to the event-region detection problem. This method is based on maximizing the product of the full conditional predictive probability density or mass functions (pdfs/pmfs) of the random-field values at all the nodes. In Section VI, we evaluate the performance of the proposed detection algorithms via numerical simulations. Concluding remarks are given in Section VII.

[1]Here, Markovian dependence implies that, given random-field values at all other locations, the conditional distribution of the random field at any location depends only on the field values at the neighboring locations, see also (2.7) in Section II.

[2]See [13, Ch. 16.3] for the definition of predictive likelihood and examples of its use.

II. Hidden Markov Random Field Framework

Assume that each node (sensor) $k \in \{1, 2, \ldots, K\}$ in the network collects a vector of measurements

$$\boldsymbol{y}_k = [y_k(1), y_k(2), \ldots, y_k(N)]^T \tag{2.1a}$$

where N denotes the number of observations collected by each node k and "^{T}" denotes a transpose. Define also the vector of all measurements:

$$\boldsymbol{y} = [\boldsymbol{y}_1^T, \boldsymbol{y}_2^T, \ldots, \boldsymbol{y}_K^T]^T. \tag{2.1b}$$

We assign a hidden random variable β_k to each node k and adopt the following hierarchical model for the collected observations:

- β_k, $k = 1, 2, \ldots, K$ form an MRF describing the *process model*:

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_K]^T. \tag{2.2}$$

- Given the MRF $\boldsymbol{\beta}$, $\boldsymbol{y}_k$ are conditionally independent random vectors with pdfs or pmfs $p_{\boldsymbol{y}_k|\beta_k}(\boldsymbol{y}_k|\beta_k; \boldsymbol{\upsilon}_k)$ that depend only on β_k

$$p_{\boldsymbol{y}|\boldsymbol{\beta}}(\boldsymbol{y} \,|\, \boldsymbol{\beta}; \boldsymbol{\upsilon}) = \prod_{k=1}^{K} p_{\boldsymbol{y}_k|\beta_k}(\boldsymbol{y}_k \,|\, \beta_k; \boldsymbol{\upsilon}_k) \tag{2.3}$$

  describing the *data (measurement-error) model*.

Here, $\boldsymbol{\upsilon}_k$ is the vector of *unknown* measurement-error model parameters at the kth node and

$$\boldsymbol{\upsilon} = [\boldsymbol{\upsilon}_1^T, \boldsymbol{\upsilon}_2^T, \ldots, \boldsymbol{\upsilon}_K^T]^T. \tag{2.4}$$

Note that the measurement-error model parameters $\boldsymbol{\upsilon}_k$ vary with the node index k, taking into account *imperfect calibration* of the sensors at different nodes. The above framework can account for both discrete and continuous measurements and random fields. The parameters $\boldsymbol{\upsilon}_k$ may be used to model the entire measurement-error probability distribution $p_{\boldsymbol{y}_k|\beta_k}(\boldsymbol{y}_k \,|\, \beta_k; \boldsymbol{\upsilon}_k)$ in a *nonparametric manner*, provided that the elements of $\boldsymbol{y}_k$ are conditionally independent, identically distributed (i.i.d.); see Section IV-A.2.

Our goal is to estimate the MRF $\boldsymbol{\beta}$ from the observations $\boldsymbol{y}_k$, $k = 1, 2, \ldots, K$. We define the probability distribution of $\boldsymbol{\beta}$ via a *conditionally-specified model* suitable for distributed neighborhood-based signal processing. Before formally defining an MRF, let us introduce some terminology and notation. Throughout this paper, we assume that the *neighborhood* of a node k [denoted by $\mathcal{N}(k)$] consists of all the nodes $l \in \{1, 2, \ldots, K\}$ that are within a *cutoff distance* d from that node, i.e.

$$\mathcal{N}(k) = \{l : \|\boldsymbol{r}_k - \boldsymbol{r}_l\| \leq d \text{ and } l \neq k\} \tag{2.5a}$$

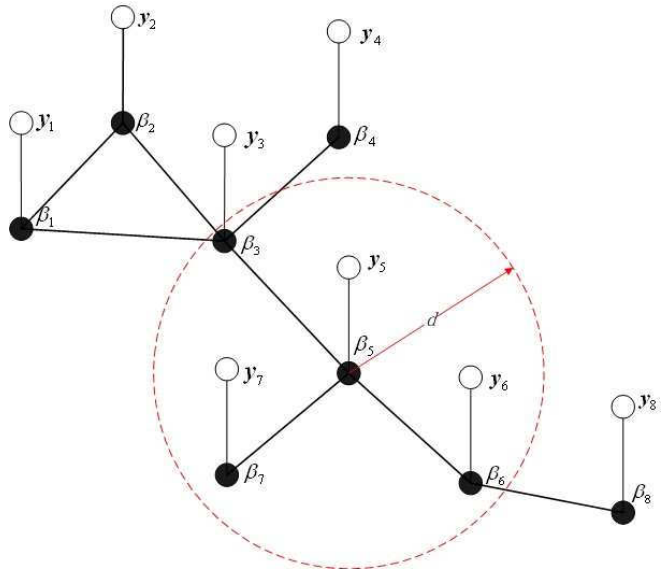

Fig. 1. A graphical representation of an HMRF model.

where

$$\|\boldsymbol{r}_k - \boldsymbol{r}_l\| = \sqrt{(\boldsymbol{r}_k - \boldsymbol{r}_l)^T(\boldsymbol{r}_k - \boldsymbol{r}_l)}$$

and $\boldsymbol{r}_k$ and $\boldsymbol{r}_l$ are the kth and lth node locations in Cartesian coordinates. Define the set of random-field values in this neighborhood:

$$\mathcal{N}_\beta(k) = \{\beta_l,\, l \in \mathcal{N}(k)\} \tag{2.5b}$$

and the conditional pdfs or pmfs of β_k given the neighboring MRF values:

$$p_{\beta_k|\mathcal{N}_\beta(k)}\big(\beta_k \,|\, \mathcal{N}_\beta(k)\big), \quad k = 1, 2, \ldots, K. \tag{2.6}$$

Then, the Markov property of an MRF $\boldsymbol{\beta}$ implies that, for all $k = 1, 2, \ldots, K$, the conditional pdfs/pmfs of β_k given the random-field values at all other nodes satisfy

$$p_{\beta_k|\{\beta_l,\, l\neq k\}}\big(\beta_k \,|\, \{\beta_l,\, l \neq k\}\big) = p_{\beta_k|\mathcal{N}_\beta(k)}\big(\beta_k \,|\, \mathcal{N}_\beta(k)\big). \tag{2.7}$$

A. HMRFs as Probabilistic Graphical Models

MRF and HMRF models belong to the (broader) class of *probabilistic graphical models* (see e.g. [11], [18]–[22] and references therein) and can be formulated using an undirected mathematical graph whose nodes correspond to the random variables in the field and its edges define the underlying neighborhood structure. In [18] and [20], graphical-model based extended *message passing*[3] algorithms are developed for inference on HMRF models with linear Gaussian measurement-error and MRF process models and known model

[3]See [22] for a detailed exposition on message passing algorithms for graphical models.

parameters, *embedded-trees* and *embedded-triangles* algorithms are developed for this scenario in [19], [20], and [23]. A *belief propagation* approach is proposed in [24] for multi-hypothesis testing of global phenomena in sensor-network environments. Fig. 1 shows a graphical representation of an HMRF model, where the filled circles depict the hidden random field (and their locations correspond to the node locations) and hollow circles the observed data. The edges in Fig. 1 describe the (conditional) statistical dependence between the nodes in the graph, as inferred from the specifications in (2.5a) and (2.3).

In the following, we present a *distributed algorithm* for computing maximum predictive likelihood estimates of the random field $\boldsymbol{\beta}$.

III. ICM Random-Field Estimation

We propose an ICM algorithm for estimating the MRF $\boldsymbol{\beta}$ where each node $k \in \{1, 2, \ldots, K\}$ performs the following steps:

(ICM1) collects the current estimates of β_l from its neighborhood $\mathcal{N}(k)$;

(ICM2) updates its estimate of β_k by maximizing the conditional predictive log likelihood:

$$L_k\big(\beta_k \,|\, \mathcal{N}_\beta(k)\big) = \max_{\boldsymbol{\upsilon}_k} \{\ln p_{\boldsymbol{y}_k|\beta_k}(\boldsymbol{y}_k|\beta_k; \boldsymbol{\upsilon}_k)\} + \ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(\beta_k \,|\, \mathcal{N}_\beta(k)\big) \tag{3.1}$$

with respect to β_k;

(ICM3) *broadcasts* the obtained estimate of β_k to the nodes in the neighborhood $\mathcal{N}(k)$.

When applied to each node k in turn, this procedure defines a single cycle of the ICM algorithm. The cycling is performed until convergence, i.e. until the estimates of β_k do not change significantly for all $k \in \{1, 2, \ldots, K\}$. The ICM approach is computationally simple and applicable to a wide range of sensing environments. It *does not* require careful treatments of loops in the inference graphs, constructing junction trees etc. It is also *localized,* implying that the nodes communicate only with their neighbors to obtain the desired results. Localized algorithms are *robust* to node failures and the communication overhead *scales well* with increasing network size, see [2] and [3]. Distributed localized algorithms and architectures also facilitate rapid data processing and information collection, and are well-suited for systems that utilize sleep modes to conserve energy [25].

Denote by $p_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ the joint pdf/pmf of $\beta_1, \beta_2, \ldots, \beta_K$. Then, applying (ICM1)–(ICM3) at each node k increases the joint predictive log-likelihood function of $\boldsymbol{\beta}$ (see also [13, Ch. 16.3]):[4]

$$L(\boldsymbol{\beta}) = \max_{\boldsymbol{\upsilon}} \{\ln[p_{\boldsymbol{y}|\boldsymbol{\beta}}(\boldsymbol{y}|\boldsymbol{\beta}; \boldsymbol{\upsilon})]\} + \ln[p_{\boldsymbol{\beta}}(\boldsymbol{\beta})\,] \tag{3.2}$$

[4]Note that the conditional predictive log likelihood $L_k\big(\beta_k \,|\, \mathcal{N}_\beta(k)\big)$ in (3.1) follows from the joint predictive log-likelihood $L(\boldsymbol{\beta})$ by substituting the identity $p_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = p_{\beta_k|\mathcal{N}_\beta(k)}\big(\beta_k \,|\, \mathcal{N}_\beta(k)\big) \cdot p_{\mathcal{N}_\beta(k)}\big(\mathcal{N}_\beta(k)\big)$ into (3.2) and keeping the terms that depend on β_k.

in a *stepwise-ascent manner*. In particular, combining the stepwise-ascent maximization approach with the Markovian property of $p_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ leads to the distributed localized iteration (ICM1)–(ICM3). In general, this iteration converges to a *local maximum* of $L(\boldsymbol{\beta})$. However, if the conditional predictive log likelihoods in (3.1) are *unimodal* in β_k (as in the HMRFs with linear Gaussian measurement-error and MRF process models studied in [18]–[20] and [23]), then the ICM algorithm converges to the global maximum of $L(\boldsymbol{\beta})$. Interestingly, its convergence to a local maximum of $L(\boldsymbol{\beta})$ [when initialized with the local maximum likelihood (ML) estimates of the β_k's] may be preferred compared with finding the global maximum because MRFs often have undesirable large-scale properties [10].

The predictive log likelihood in (3.2) has a Bayesian interpretation. Here, we view $p_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ as the *prior distribution* of the hidden field $\boldsymbol{\beta}$ and assign a flat prior distribution: $p_{\boldsymbol{\upsilon}}(\boldsymbol{\upsilon}) \propto 1$ to the measurement-error model parameters $\boldsymbol{\upsilon}$. Then, maximizing $L(\boldsymbol{\beta})$ in (3.2) yields a mode of the joint posterior pdf/pmf of the unknown parameters. We emphasize that the purpose of the proposed method is to resolve ambiguous measurements. Otherwise, if the data provides strong evidence about the hidden field $\boldsymbol{\beta}$, the influence of the prior $p_{\boldsymbol{\beta}}(\boldsymbol{\beta})$ disappears, which is true for Bayesian methods in general. The ICM approach to finding modes of joint posterior distributions dates back to the seminal paper by Lindley and Smith [26, Sect. 4], see also [27, Ch. 10.2.1] and [28, Ch. 6.2.2]. The iteration (ICM1)–(ICM3) generalizes the ICM algorithm for image analysis in [9, Ch. 7.4.3] and [10] and adapts it to the sensor-network scenario by allowing for more general neighborhood models and unknown measurement-error model parameters that vary from node to node. The latter extension is key for sensor-network applications where the nodes are not perfectly calibrated and data processing should be performed *locally* as much as possible. It also allows nonparametric measurement-error modeling, as discussed in Section IV-A.2.

In the following section, we demonstrate the proposed approach by applying it to event-region detection.

IV. Event-Region Detection Using the HMRF Framework and ICM Method

We utilize the proposed HMRF framework and ICM method to efficiently remove false alarms in event-region detection tasks. Here, our goal is to detect a region in the environment in which an event of interest has occurred. For example, if the network is capable of sensing concentration of chemical X, then it is of interest to answer the following question [14]: "In which regions in the environment is the concentration of chemical X greater than a specified level?"

We first describe measurement-error and process models suitable for event-region detection (Sections IV-A and IV-B) and then derive the corresponding ICM algorithms for event-region detection (Sections IV-C and IV-D).

A. Measurement-Error Model

In this section, we consider hidden fields that take two discrete values

- $\beta_k = 0$ (signal absent) and
- $\beta_k = 1$ (signal present)

and utilize a simple signal-plus-noise measurement-error model for the measurements $y_k(1), y_k(2), \ldots, y_k(N)$ collected at node $k \in \{1, 2, \ldots, K\}$:

$$y_k(t) = \mu_k(\beta_k) + e_k(t), \quad t = 1, 2, \ldots, N \tag{4.1a}$$

where

- $\mu_k(0) = 0$ (signal absent),
- $\mu_k(1)$ is the (unknown) nonzero signal, and
- $e_k(t)$, $t = 1, 2, \ldots, N$ is zero-mean i.i.d. noise.

We denote the pdf/pmf of the noise $e_k(t)$ by $p_{\mathrm{noise}_k}(e_k(t))$. Consequently, given β_k, $y_k(1), y_k(2), \ldots, y_k(N)$ are conditionally i.i.d. random variables with the joint pdf/pmf

$$p_{\boldsymbol{y}_k|\beta_k}(\boldsymbol{y}_k \,|\, \beta_k) = \prod_{t=1}^{N} p_{\mathrm{noise}_k}\Big(y_k(t) - \mu_k(\beta_k)\Big) \tag{4.1b}$$

see also (2.3) for a full measurement-error model specification.

1) Gaussian Measurement-Error Model: Under the Gaussian measurement-error model, we assume that the noise pdf at node k is zero-mean Gaussian:

$$p_{\mathrm{noise}_k}(e_k(t); \sigma_k^2) = \frac{1}{\sigma_k\sqrt{2\pi}} \cdot \exp\big[-e_k^2(t)/(2\sigma_k^2)\big] \tag{4.2}$$

where σ_k^2 is the unknown noise variance at the kth sensor. Here, the measurement-error model parameters are $\boldsymbol{\upsilon}_k = \sigma_k^2$ for $\beta_k = 0$ and $\boldsymbol{\upsilon}_k = [\mu_k(1), \sigma_k^2]^T$ for $\beta_k = 1$.

2) Nonparametric Measurement-Error Models: We now consider a scenario where the noise probability distribution $p_{\mathrm{noise}_k}(\cdot)$ at node k is *unknown* and utilize a class of *nonparametric* measurement-error models. This scenario is important in practical applications where accurate parametric measurement-error models are difficult to find, as is often the case in large-scale sensor networks operating in time-varying environments (see e.g. [15]–[17]). To simplify the notation, we omit the dependence of the mean value μ_k on β_k throughout this section. Clearly, the discussion on *unknown* μ_k corresponds to the case where $\mu_k = \mu_k(1) \neq 0$.

Assume that, given β_k, $y_k(1), y_k(2), \ldots, y_k(N)$ are conditionally i.i.d. random variables with mean $\mu_k = \mu_k(\beta_k)$ where each $y_k(t)$ is assigned a multinomial probability $p_{k,t}$. We then construct the following non-

parametric log-likelihood function of the mean μ_k at node k:

$$l_k(\mu_k) = \sum_{t=1}^{N} \ln p_{k,t}(\mu_k) \tag{4.3}$$

where $p_{k,t}(\mu_k)$, $t = 1, 2, \ldots, N$ are estimates of the probabilities $p_{k,t}$, $t = 1, 2, \ldots, N$ computed by minimizing the Cressie-Read power divergence[5] between the discrete distribution defined by $p_{k,t}$, $t = 1, 2, \ldots, N$ and the discrete uniform distribution on $t = 1, 2, \ldots, N$ (see [30], [31], and [32, Ch. 3.16]):

$$\min_{p_{k,t}\, t=1,2,\ldots,N} \frac{\sum_{t=1}^{N}[(N p_{k,t})^{-\kappa} - 1]}{N\,\kappa\,(1+\kappa)} \tag{4.4a}$$

subject to the constraints

$$\sum_{t=1}^{N} p_{k,t}\,[y_k(t) - \mu_k] = 0, \quad p_{k,t} \geq 0, \quad \sum_{t=1}^{N} p_{k,t} = 1. \tag{4.4b}$$

Here, $-\infty < \kappa < \infty$ is a *known* constant [defining a particular choice of the discrepancy measure in (4.4a)] and the degenerate cases $\kappa \in \{0, -1\}$ are handled by taking limits.

In the following, we focus on the non-trivial case where[6]

$$y_{k,\text{MIN}} = \min_{t\in\{1,2,\ldots,N\}} y_k(t) < \mu_k < \max_{t\in\{1,2,\ldots,N\}} y_k(t) = y_{k,\text{MAX}} \tag{4.5}$$

and on the limiting values of κ (i.e. $\kappa = 0$ and $\kappa = -1$), which correspond to commonly used *least favorable* distribution families [32, Chs. 9.6 and 9.11] and lead to the *empirical likelihood* and *empirical entropy* measurement-error models discussed below. (The concept of a least favorable family was introduced by Stein in [33].)

Empirical Likelihood: If $\kappa = 0$ in (4.4a), (4.3) simplifies to the following *concentrated empirical log-likelihood function*[7] of the mean μ_k at node k:

$$l_k(\mu_k) = \Big\{ \max_{p_{k,t}\, t=1,2,\ldots,N} \Big(\sum_{t=1}^{N} \ln p_{k,t} \Big) \;\Big|\; \sum_{t=1}^{N} p_{k,t}\,[y_k(t) - \mu_k] = 0,\, p_{k,t} \geq 0,\, \sum_{t=1}^{N} p_{k,t} = 1 \Big\} \tag{4.6}$$

which can be viewed as a *multinomial concentrated log likelihood* [34]. In this case, the measurement-error model parameters are $\boldsymbol{\upsilon}_k = [p_{k,1}, p_{k,2}, \ldots, p_{k,N}]^T$ for $\beta_k = 0$ and $\boldsymbol{\upsilon}_k = [\mu_k, p_{k,1}, p_{k,2}, \ldots, p_{k,N}]^T$ for $\beta_k = 1$, where the multinomial probabilities $p_{k,1}, p_{k,2}, \ldots, p_{k,N}$ are constrained to satisfy the conditions in (4.4b), see also (4.6).

Maximizing $l_k(\mu_k)$ with respect to μ_k yields

$$\max_{\mu_k} [l_k(\mu_k)] = -N \ln N \tag{4.7a}$$

[5]The Cressie-Read divergence is closely related to the Rényi divergence [29], see also [30] and [31].

[6]Note that the optimization problem in (4.4) does not have a solution if $\mu_k < y_{k,\text{MIN}}$ or $\mu_k > y_{k,\text{MAX}}$. In such cases, we set $l_k(\mu_k) = -\infty$ by convention. If $y_{k,\text{MIN}} = y_{k,\text{MAX}} = \mu_k$, we take $l_k(\mu_k) = -N \ln N$ and if $\mu_k = y_{k,\text{MIN}} < y_{k,\text{MAX}}$ or $\mu_k = y_{k,\text{MAX}} > y_{k,\text{MIN}}$, we set $l_k(\mu_k) = -\infty$.

[7]See also [32] and [34] for the definition and properties of the empirical likelihood.

which follows by noting that, subject to $\sum_{t=1}^{N} p_{k,t} = 1$, the log-likelihood function $\sum_{t=1}^{N} \ln p_{k,t}$ is maximized by choosing the discrete uniform distribution of the observations (i.e. $p_{k,t} = 1/N$, $t = 1, 2, \ldots, N$). This choice yields the *nonparametric maximum likelihood estimate* (NPMLE) of μ_k:

$$\overline{y}_k = \frac{1}{N} \cdot \sum_{t=1}^{N} y_k(t) = \arg\max_{\mu_k} l_k(\mu_k) \tag{4.7b}$$

also known as the *bootstrap estimate* of μ_k [34]. In Appendix A, we compute (4.6) by solving a *one-dimensional convex dual problem*:

$$l_k(\mu_k) = -N \ln N + \min_{\lambda_k} \Xi_k(\lambda_k; \mu_k) \tag{4.8a}$$

where

$$\Xi_k(\lambda_k; \mu_k) = -\sum_{t=1}^{N} \ln\{1 + \lambda_k \left[y_k(t) - \mu_k\right]\} \tag{4.8b}$$

is a convex function of λ_k. To ensure that the estimates of the multinomial probabilities remain in the allowed parameter space, the search for λ_k that minimizes (4.8b) should be constrained to the interval (see Appendix A):

$$\frac{1 - N^{-1}}{\mu_k - y_{k,\mathrm{MAX}}} < \lambda_k < \frac{1 - N^{-1}}{\mu_k - y_{k,\mathrm{MIN}}} \tag{4.9}$$

and can be efficiently performed using the damped Newton-Raphson iteration[8]:

$$\lambda_k^{(i+1)} = \lambda_k^{(i)} + \delta_k^{(i)} \cdot \Big(\sum_{t=1}^{N} \frac{\left[y_k(t) - \mu_k\right]^2}{\{1 + \lambda_k^{(i)} \left[y_k(t) - \mu_k\right]\}^2} \Big)^{-1} \cdot \sum_{t=1}^{N} \frac{y_k(t) - \mu_k}{1 + \lambda_k^{(i)} \left[y_k(t) - \mu_k\right]} \tag{4.10}$$

where the *damping factor* $0 < \delta_k^{(i)} \leq 1$ is chosen (at every step i) to ensure that (4.8b) decreases and $\lambda_k^{(i+1)}$ remains within the interval specified by (4.9).[9] The above iteration converges to the unique solution $\lambda_k = \lambda_k(\mu_k)$.

In Appendix A, we sketch a proof that the empirical-likelihood approach employs a *least favorable nonparametric distribution family* for estimating μ_k and derive the Cramér-Rao bound (CRB) for estimating μ_k under the empirical likelihood measurement-error model. Assuming the discrete uniform distribution of the observations, this CRB simplifies to:

$$-\Big[\sum_{t=1}^{N} \frac{1}{N} \cdot \frac{d^2 l_k(\overline{y}_k)}{d\overline{y}_k^2} \Big]^{-1} = \frac{s_k^2}{N} \tag{4.11}$$

[8]See e.g. [35, Ch. 9.7] for an introduction to the Newton-Raphson algorithms. To simplify the notation in (4.10) and later in (4.16), we omit the dependence of $\lambda_k^{(i)}$ and $\delta_k^{(i)}$ on μ_k.

[9]In particular, we start with $\delta_k^{(i)} = 1$ and check if (4.8b) decreases and $\lambda_k^{(i+1)}$ remains within the interval (4.9). If these tests fail, we keep halving $\delta_k^{(i)}$ until they are satisfied.

where

$$s_k^2 = s_{0,k}^2 - \overline{y}_k^2 \tag{4.12a}$$

$$s_{0,k}^2 = \frac{1}{N} \cdot \sum_{t=1}^{N} y_k^2(t) \tag{4.12b}$$

and $\overline{y}_k$ has been defined in (4.7b).

Empirical Entropy: For $\kappa = -1$, (4.4a) reduces to

$$\min_{p_{k,t}\ t=1,2,\ldots,N} \sum_{t=1}^{N} p_{k,t} \ln(N p_{k,t}) \tag{4.13}$$

subject to the constraints in (4.4b). In (4.13), we minimize the *relative entropy*[10] between the multinomial distribution defined by the probabilities $p_{k,t}$, $t = 1, 2, \ldots, N$ and the discrete uniform distribution on $t = 1, 2, \ldots, N$, yielding the *empirical entropy* estimates $p_{k,t}(\mu_k)$, $t = 1, 2, \ldots, N$ of the multinomial probabilities. It can be shown that $p_{k,t}(\mu_k)$ have the following form (see Appendix B):

$$p_{k,t}(\mu_k) = \frac{\exp[\lambda_k(\mu_k)\, y_k(t)]}{\sum_{\tau=1}^{N} \exp[\lambda_k(\mu_k)\, y_k(\tau)]}, \quad t = 1, 2, \ldots, N \tag{4.14}$$

where $\lambda_k(\mu_k)$ is obtained by minimizing

$$\zeta_k(\lambda_k; \mu_k) = \sum_{t=1}^{N} \exp\{\lambda_k\, [y_k(t) - \mu_k]\} \tag{4.15}$$

with respect to λ_k. Note that $\zeta_k(\lambda_k; \mu_k)$ is a convex function of λ_k and can be efficiently minimized using the damped Newton-Raphson iteration:

$$\lambda_k^{(i+1)} = \lambda_k^{(i)} - \delta_k^{(i)} \cdot \Big\{ \sum_{t=1}^{N} \exp[\lambda_k^{(i)}\, y_k(t)] \cdot [y_k(t) - \mu_k]^2 \Big\}^{-1} \cdot \sum_{t=1}^{N} \exp[\lambda_k^{(i)}\, y_k(t)] \cdot [y_k(t) - \mu_k]. \tag{4.16}$$

Here, the damping factor $0 < \delta_k^{(i)} \leq 1$ is chosen to ensure that (4.15) decreases. Finally, we compute the nonparametric log-likelihood function of μ_k by substituting (4.14) into (4.3):

$$l_k(\mu_k) = N \cdot \lambda_k(\mu_k)\, \overline{y}_k - N \cdot \ln \Big\{ \sum_{t=1}^{N} \exp[\lambda_k(\mu_k)\, y_k(t)] \Big\}. \tag{4.17}$$

The above empirical-entropy approach is closely related to the *nonparametric tilting* in [37, Sect. 11] and [38, Ch. 10.10]. It is also known as the *empirical exponential family likelihood* [39, Ch. 10] because it can be derived by constraining the probability distribution of $y_k(1), y_k(2), \ldots, y_k(N)$ to belong to the exponential family of distributions.

In [37, Sect. 11], Efron presented the CRB for μ_k under the empirical entropy measurement-error model and used it to argue that the empirical-entropy approach employs a least favorable family for estimating μ_k. Assuming the discrete uniform distribution of the observations, the expression for this CRB simplifies to (4.11), see Appendix B and [37, eq. (11.10)].

[10]Relative entropy, also known as Kullback-Leibler distance, is defined in e.g. [36, Ch. 2.3].

B. Autologistic MRF Process Model

Assume that each node k makes a binary decision about its current status, i.e. it decides between the hypothesis

$\mathrm{H}_{0,k}$: (signal absent, $\beta_k = 0$), corresponding to $\mu_k = 0$

versus the one-sided alternative

$\mathrm{H}_{1,k}$: (signal present, $\beta_k = 1$), corresponding to $\mu_k > 0$.

This formulation is suitable for detecting event regions with elevated concentrations of chemicals, see the example at the beginning of Section IV. In this example, we restrict the parameter space of the mean signal μ_k to the set of non-negative real numbers. To describe the binary MRF for event-region detection problems, we adopt the *autologistic MRF process model* specified by the conditional pmfs (see [9, Ch. 6.5.1]):

$$p_{\beta_k|\mathcal{N}_\beta(k)}(\beta_k \,|\, \mathcal{N}_\beta(k)) = \frac{\exp\left(a_k\beta_k + \beta_k \cdot \sum_{l\in\mathcal{N}(k)} c_{k,l}\,\beta_l\right)}{1 + \exp\left(a_k + \sum_{l\in\mathcal{N}(k)} c_{k,l}\,\beta_l\right)} \tag{4.18a}$$

for $k = 1, 2, \ldots, K$, where a_k and $c_{k,l}$ are *spatial-trend* and *spatial-dependence* MRF model parameters. Furthermore, we utilize the following simple spatial trend and dependence models:

- *constant spatial trend* (independent of k):

$$a_k = a \tag{4.18b}$$

- *homogeneous spatial dependence* with equal evidence from each neighbor:

$$c_{k,l} = \begin{cases} \eta, & \|\boldsymbol{r}_k - \boldsymbol{r}_l\| \le d \\ 0, & \|\boldsymbol{r}_k - \boldsymbol{r}_l\| > d \end{cases} \tag{4.18c}$$

  where d is the cutoff distance, see also Section II.

In event-region detection problems, η is a positive constant describing the *field strength*. This spatial-dependence model quantifies the notion that the random-field values at the nodes that are close (in terms of the spatial distance) should be more similar than the values at the nodes that are far apart. More complex spatial dependence models can be developed along the lines of [9] (for isotropic dependence) and [40] (for anisotropic dependence).

In applications where the cutoff distance d is approximately equal to the radio-transmission range of the sensor elements, the neighborhood $\mathcal{N}(k)$ consists of those nodes with which k can communicate directly. Then, we can determine the neighborhoods without utilizing the node location information. However, the assumption that the cutoff distance coincides with the communication range may be impractical. In addition, the effective cutoff distance may vary slightly from one node to another.

In the following, we specialize the general ICM algorithm in Section III to the event-region detection problem using the measurement-error model in Section IV-A and process model in (4.18a)–(4.18c).

C. ICM Detection for Gaussian Measurement-Error Model

We first define the indicator function

$$i_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{otherwise} \end{cases} . \tag{4.19}$$

Under the Gaussian measurement-error and autologistic process models, Step (ICM2) in the ICM algorithm simplifies to selecting $\beta_k = 1$ if

$$\begin{aligned} L_k\big(1 \,|\, \mathcal{N}_\beta(k)\big) - L_k\big(0 \,|\, \mathcal{N}_\beta(k)\big) &= \frac{N}{2} \cdot \ln\Big(\frac{s_{0,k}^2}{s_k^2}\Big) \cdot i_{[0,\infty)}(\overline{y}_k) + a_k + \sum_{l \in \mathcal{N}(k)} c_{k,l}\beta_l & (4.20a) \\ &= \frac{N}{2} \cdot \ln\Big(\frac{s_{0,k}^2}{s_k^2}\Big) \cdot i_{[0,\infty)}(\overline{y}_k) + a + \eta\, u_k & (4.20b) \\ &\geq 0 & (4.20c) \end{aligned}$$

and selecting $\beta_k = 0$ otherwise; see Appendix C for details of the derivation. Here,

$$u_k = \sum_{l \in \mathcal{N}(k)} \beta_l \tag{4.21}$$

is the number of neighbors of k reporting the presence of signal and $\mathcal{N}(k), \overline{y}_k, s_k^2$ and $s_{0,k}^2$ have been defined in (2.5a), (4.7b), (4.12a), and (4.12b). Equation (4.20b) follows by substituting (4.18b) and (4.18c) into (4.20a). The first term in (4.20b) is the one-sided t-test statistic for the mean μ_k (based on the "local" measurements collected at node k), whereas the second and third terms account for the spatial trend and spatial dependence effects introduced by the MRF model.

1) Initialization: To obtain initial decisions at each node k, we ignore the neighborhood dependence and apply the *local* one-sided t test for the mean μ_k: select $\beta_k = 1$ if

$$\frac{s_{0,k}^2}{s_k^2} \cdot i_{[0,\infty)}(\overline{y}_k) \geq \tau_{\mathrm{G}} \tag{4.22a}$$

and select $\beta_k = 0$ otherwise. This test is also the generalized likelihood ratio (GLR) test for the hypothesis testing problem described in Section IV-B. Denote by $\mathcal{B}(0.5(N-1), 0.5)$ the *central beta distribution* with parameters $0.5(N-1)$ and 0.5. We select the threshold

$$\tau_{\mathrm{G}} = b^{-1}_{0.5(N-1),0.5,2P_{\mathrm{FA}}} \tag{4.22b}$$

to guarantee a specified probability of false alarm P_{FA}. Here, $b_{0.5(N-1),0.5,p}$ is defined using

$$P[\beta \leq b_{0.5(N-1),0.5,p}] = p \tag{4.22c}$$

where β is a $\mathcal{B}(0.5(N-1), 0.5)$ random variable.

D. ICM Detection for Nonparametric Measurement-Error Models

Under the nonparametric measurement-error models in Section IV-A.2, the condition (4.5) implies that one-sided detection in Section IV-B will be meaningful only if

$$y_{k,\mathrm{MAX}} \geq 0 \tag{4.23}$$

with equality implying $y_{k,\mathrm{MIN}} = y_{k,\mathrm{MAX}} = 0$. For the empirical likelihood and entropy measurement-error models, Step (ICM2) simplifies to selecting $\beta_k = 1$ if

$$\begin{aligned} & L_k\big(1 \,|\, \beta_l \in \mathcal{N}_\beta(k)\big) - L_k\big(0 \,|\, \beta_l \in \mathcal{N}_\beta(k)\big) \\ & \quad = \max_{\mu_k > 0} [l_k(\mu_k)] - l_k(0) + \ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(1|\mathcal{N}_\beta(k)\big) - \ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(0|\mathcal{N}_\beta(k)\big) && \text{(4.24a)} \\ & \quad = [-N \ln N - l_k(0)] \cdot i_{[0,\infty)}(\overline{y}_k) + a + \eta\, u_k && \text{(4.24b)} \\ & \quad \geq 0 && \text{(4.24c)} \end{aligned}$$

and selecting $\beta_k = 0$ otherwise, see Appendix D. Here, the nonparametric log likelihoods $l_k(0)$ for the empirical likelihood and entropy models are computed using (4.8) and (4.17), respectively.

1) Initialization: We now discuss the initialization of the ICM iteration under the empirical likelihood and entropy measurement-error models. We propose the following local GLR tests that ignore the neighborhood dependence: select $\beta_k = 1$ if

$$\sqrt{2\,[-N \ln N - l_k(0)]} \cdot i_{[0,\infty)}(\overline{y}_k) \geq \tau_{\mathrm{NP}} \tag{4.25a}$$

and select $\beta_k = 0$ otherwise. The threshold τ_{NP} which guarantees a specified probability of false alarm P_{FA} can be approximately computed by solving (see Appendix E):

$$\Phi(\tau_{\mathrm{NP}}) = 1 - P_{\mathrm{FA}} \tag{4.25b}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal random variable. The above approximation is based on the Wilks' theorem for the empirical likelihood [32, Th. 2.2], [34, Sect. 2.3] and similar results for empirical entropy [41], [42], derived under the assumption that $N \to \infty$, see also Appendix E. Therefore, its accuracy improves as the number of observations per sensor increases.

In the above ICM algorithms, the nodes exchange binary messages ($\beta_k = 0$ or $\beta_k = 1$) to inform neighbors about their status; the communication cost of this exchange is low, which is important in most practical applications that require energy and bandwidth efficiency [7].

V. MRF CALIBRATION

Assume that *training data* is available containing both the observations $\boldsymbol{y}_k$, $k = 1, 2, \ldots, K$ and the true values of the MRF $\boldsymbol{\beta}$. We develop a calibration method for estimating the ***MRF model parameters*** from the training data. Denote the vector of MRF model parameters by $\boldsymbol{\omega}$. To emphasize the dependence of the local and global predictive log-likelihood functions in (3.1) and (3.2) on $\boldsymbol{\omega}$, we use $L_k\big(\beta_k \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big)$ and $L(\boldsymbol{\beta}; \boldsymbol{\omega})$ to denote these functions throughout this section. Similarly, we use $p_{\beta_k|\mathcal{N}_\beta(k)}\big(\beta_k \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big)$ to denote the conditional pdfs in (2.6).

We denote

$$\frac{\exp[L(\boldsymbol{\beta}; \boldsymbol{\omega})]}{\int \exp[L(\boldsymbol{b}; \boldsymbol{\omega})]\, d\boldsymbol{b}} \quad \text{or} \quad \frac{\exp[L(\boldsymbol{\beta}; \boldsymbol{\omega})]}{\sum_{\boldsymbol{b}} \exp[L(\boldsymbol{b}; \boldsymbol{\omega})]} \tag{5.1}$$

as the "predictive" pdfs or pmfs of $\boldsymbol{\beta}$, see [13, Ch. 16.3]. Then, we may compute maximum "predictive" likelihood estimates of $\boldsymbol{\omega}$ by maximizing the expressions in (5.1). However, the denominators in (5.1) are usually computationally intractable. Motivated by Besag's pseudolikelihood approach in [9, pp. 461–463] and [43], we construct a computationally tractable log pseudo predictive likelihood function:

$$L_{\text{PPL}}(\boldsymbol{\omega}) = \sum_{k=1}^{K} \ln \Big\{ \frac{\exp[L_k\big(\beta_k \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big)]}{\sum_{i=0}^{1} \exp[L_k\big(i \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big)]} \Big\} \tag{5.2}$$

and estimate the MRF model parameters $\boldsymbol{\omega}$ by maximizing $L_{\text{PPL}}(\boldsymbol{\omega})$ with respect to $\boldsymbol{\omega}$. Here,

$$\frac{\exp[L_k\big(\beta_k \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big)]}{\sum_{i=0}^{1} \exp[L_k\big(i \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big)]} \tag{5.3}$$

is the full conditional predictive pdf/pmf of β_k. The above calibration method applies to the general measurement-error and MRF models described in Section II.

Event-Region Detection: We now specialize (5.2) to the event-region detection problem, leading to

$$\begin{aligned} L_{\text{PPL}}(\boldsymbol{\omega}) \;\; &= \;\; \text{const} + a \cdot \Big(\sum_{k=1}^{K} \beta_k \Big) + \eta \cdot \Big(\sum_{k=1}^{K} \beta_k \, u_k \Big) \\ &\qquad - \sum_{k=1}^{K} \ln \Big\{ 1 + \exp\big[L_k\big(1 \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big) - L_k\big(0 \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big) \big] \Big\} \end{aligned} \tag{5.4}$$

where const denotes terms that do not depend on the MRF model parameters $\boldsymbol{\omega}$. Here, (5.4) follows by substituting the autologistic MRF model (4.18a)–(4.18c) into (5.2) and neglecting constant terms. Under the Gaussian and nonparametric measurement-error models in Section IV-A, the expressions $L_k\big(\beta_k = 1 \,|\, \beta_l \in \mathcal{N}_\beta(k)\big) - L_k\big(\beta_k = 0 \,|\, \beta_l \in \mathcal{N}_\beta(k)\big)$ in (5.4) simplify to (4.20b) and (4.24b), respectively. To efficiently compute the last term in (5.4), we utilize the following approximation: for large positive x,

$$\ln[1 + \exp(x)] \approx x. \tag{5.5}$$

Interestingly, setting the data-dependent log-likelihood terms in (3.1) to zero and substituting the resulting expressions into (5.2) yields the Besag's *pseudo log-likelihood function*

$$L_{\mathrm{PL}}(\boldsymbol{\omega}) = \sum_{k=1}^{K} \ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(\beta_k \,|\, \mathcal{N}_\beta(k); \boldsymbol{\omega}\big) \tag{5.6}$$

for estimating the MRF model parameters, see [9, pp. 461–463] and [43]. Note that (5.6) utilizes only the MRF $\boldsymbol{\beta}$ and does not depend on the measurements $\boldsymbol{y}_k$, $k = 1, 2, \ldots, K$. Maximizing the pseudo log likelihood (5.6) would yield reasonable estimates of the MRF parameters if the measurement-error model parameters $\boldsymbol{\upsilon}_1, \boldsymbol{\upsilon}_2, \ldots, \boldsymbol{\upsilon}_K$ were *known* in the ICM estimation/detection stage. Note, however, that we assume that $\boldsymbol{\upsilon}_1, \boldsymbol{\upsilon}_2, \ldots, \boldsymbol{\upsilon}_K$ are *unknown* and estimate them *locally* at each node, which is taken into account by the PPL calibration method in (5.2).

VI. Numerical Examples

To assess the performance of the proposed event-region detection methods, we consider sensor networks containing $K = 1000$ nodes randomly (uniformly) distributed on a 50 m $\times$ 50 m grid with 1 m spacing between the potential node locations. We assume that each sensor k collects $N = 5$ measurements corrupted by i.i.d. zero-mean additive noise, unless specified otherwise (see e.g. Section VI-C). The noiseless field containing two event regions is shown in Fig. 2 (left) and the sensor locations (with corresponding ideal decisions) are shown in Fig. 2 (right). Here, the filled circles correspond to the nodes in the event regions. The noisy measurements collected at the nodes k in the two event regions have means $\mu_k \in \{0.8, 1\}$ whereas the noise-only measurements collected at the nodes outside the event regions have zero means.

Throughout this section, we set the cutoff distance to $d = 3$ m and define neighborhoods according to (2.5a). In all simulation examples, we estimated the MRF model parameters a (spatial trend) and η (field strength) using the calibration procedure in Section V, where the calibration field and other details of our implementation are given in Section VI-D.

A. Gaussian Measurement Scenario

In the first set of simulations, we generated the simulated data using the Gaussian measurement-error model in Section IV-A.1 with constant noise variance $\sigma_k^2 = 0.5$ for all $k = 1, 2, \ldots, K$. In Fig. 3 (left), we show the averaged observations $\overline{y}_k$, $k = 1, 2, \ldots, K$ in (4.7b) as functions of the node locations for one realization of the noisy field. Applying the one-sided t test in (4.22) yields the results in Fig. 3 (right), where the threshold τ_{G} was chosen to satisfy the false-alarm probability $P_{\mathrm{FA}} = 5\%$. The filled circles correspond to the nodes declaring the presence of signal whereas hollow circles correspond to the nodes declaring the signal absence. The t-test decisions were used to initialize the Gaussian ICM detector (described in Section IV-C, see also

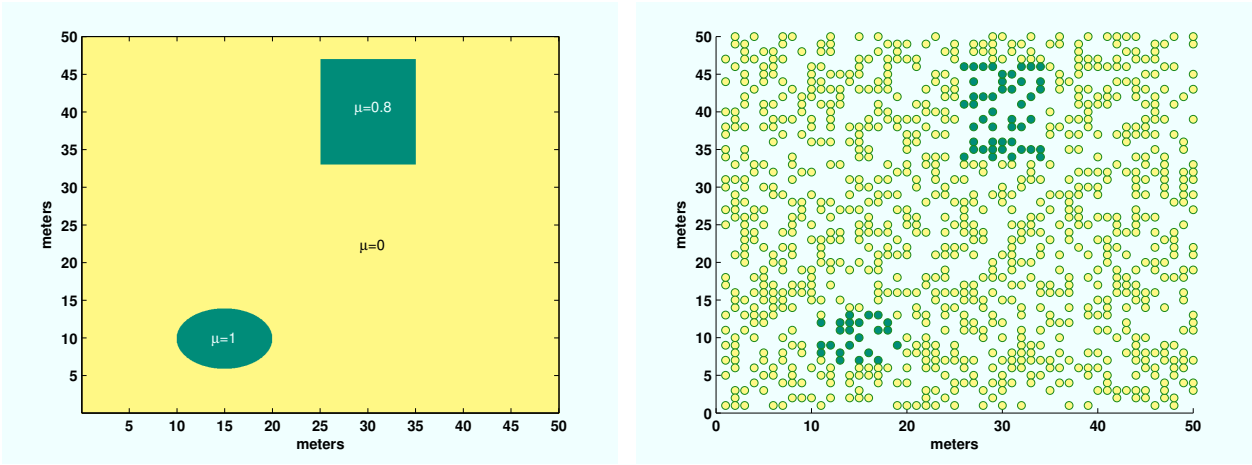

Fig. 2. (Left) Noiseless field and (right) a sensor network with $K = 1000$ nodes.

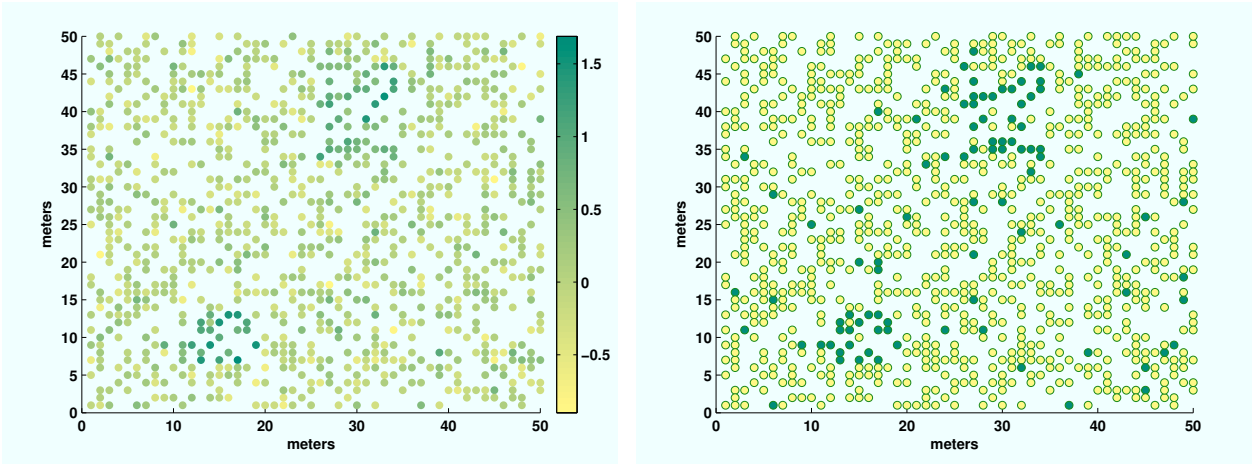

Fig. 3. Gaussian measurement scenario: (left) averaged observations $\overline{y}_k$, $k = 1, 2, \ldots, K$ as functions of the node locations and (right) one-sided t-test results for $P_{\mathrm{FA}} = 5\%$.

Section III); the decisions after one and two ICM cycles are shown in Fig. 4. In this example, all isolated nodes reporting the presence of signal were correctly recognized as false alarms already after two ICM cycles. The Gaussian ICM algorithm converged in four cycles yielding the results in Fig. 5.

Applying the nonparametric ICM detectors in Section IV-D yields (upon convergence) the results in Fig. 6. These detectors were initialized using the local GLR tests in (4.25) with the threshold τ_{NP} chosen to (approximately) satisfy the false-alarm probability $P_{\mathrm{FA}} = 5\%$. Both the empirical likelihood and entropy based ICM algorithms converged in four cycles and were successful in removing the false alarms.

B. Quantized Gaussian Measurement Scenario

We now study the performance of the proposed methods in the case where the Gaussian observations [generated as described in Section VI-A] have been coarsely quantized, leading to non-Gaussian measurements from a discrete probability distribution. Here, we quantized the measurements to the closest integer

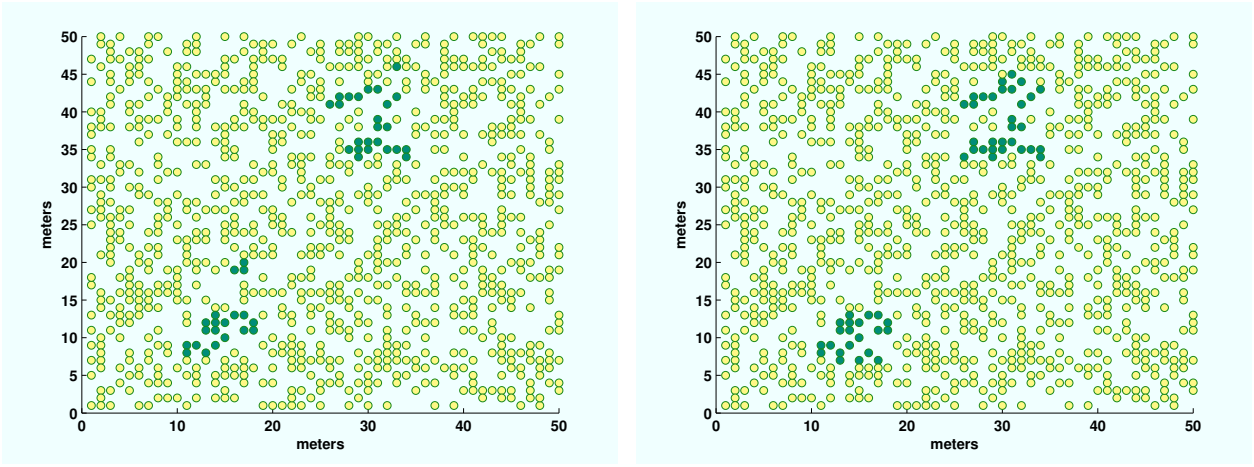

Fig. 4. Gaussian measurement scenario: Event-region detection results after (left) one cycle and (right) two cycles of the Gaussian ICM algorithm.

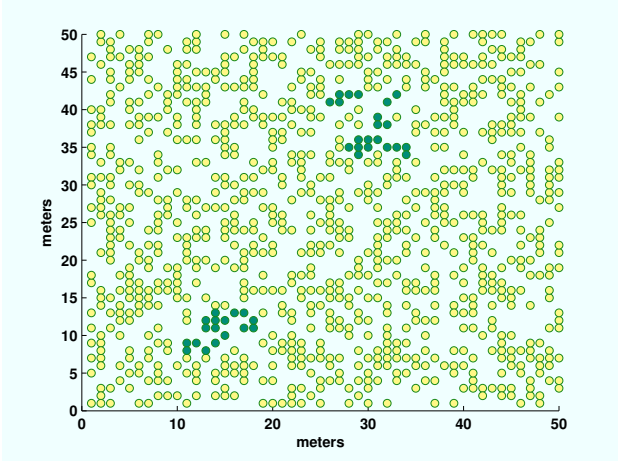

Fig. 5. Gaussian measurement scenario: Event-region detection results upon convergence of the Gaussian ICM algorithm.

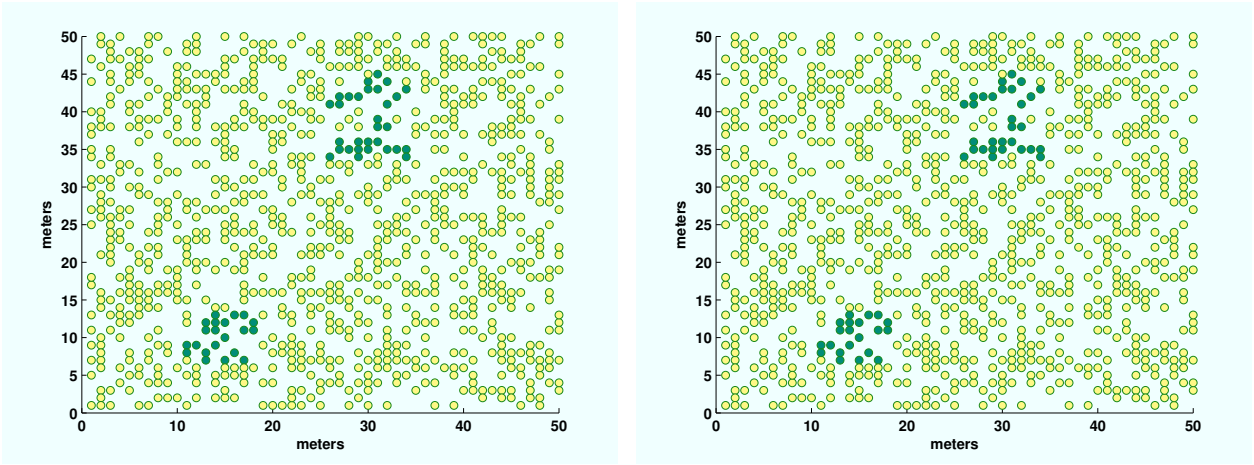

Fig. 6. Gaussian measurement scenario: Event-region detection results for (left) the empirical likelihood and (right) empirical entropy nonparametric ICM algorithms.

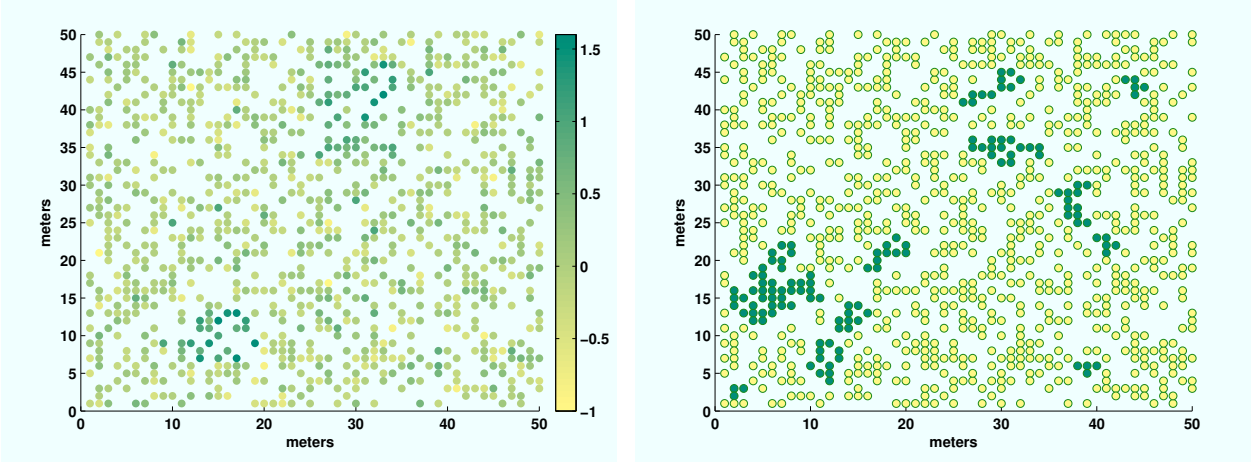

Fig. 7. Quantized Gaussian measurement scenario: (Left) averaged observations $\overline{y}_k$, $k = 1, 2, \ldots, K$ as functions of the node locations and (right) event-region detection results for the Gaussian ICM algorithm.

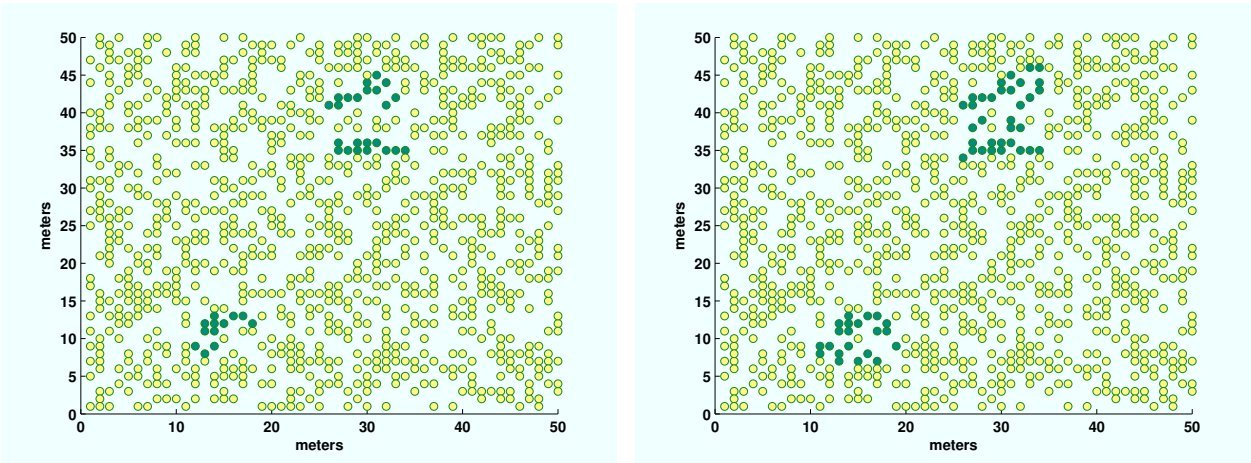

Fig. 8. Quantized Gaussian measurement scenario: Event-region detection results for (left) the empirical likelihood and (right) empirical entropy nonparametric ICM algorithms.

values in the interval $[-2, 3]$. In Fig. 7 (left), we show the averages $\overline{y}_k$, $k = 1, 2, \ldots, K$ of the quantized observations [see (4.7b)] as functions of the node locations. Applying the ICM detectors for Gaussian and nonparametric measurement-error models to the quantized measurements yields the results in Figs. 7 (right) and 8, respectively. The ICM algorithms were initialized using the local GLR tests in (4.22) and (4.25) with the thresholds τ_{G} and τ_{NP} chosen using (4.22b)–(4.22c) and (4.25b) to satisfy the false-alarm probability $P_{\mathrm{FA}} = 5\%$. The Gaussian ICM algorithm performs poorly under this scenario due to the mismatch between the quantized observations and assumed Gaussian measurement-error model, see Fig. 7 (right). The empirical likelihood and empirical entropy based ICM methods estimated the unknown probability distributions of the measurements and successfully removed the false alarms, see Fig. 8. Unlike the Gaussian and empirical likelihood approaches, the empirical entropy method provides a connected estimate of the event region in the upper right corner of the network.

C. Probabilities of False Alarm and Miss

We analyze the *average error performances* of the GLR and ICM methods under the Gaussian and quantized Gaussian measurement scenarios. Our performance metrics are the average probabilities of *false alarm* and *miss*, calculated using 100 independent trials[11] where averaging has been performed over the noisy random-field realizations, random node locations, and scheduling (in the ICM methods).

We first consider the Gaussian measurement scenario and present the average probabilities of false alarm and miss for different methods as functions of the number of observations per sensor N, see Fig. 9. In this case,

- the average false-alarm and miss error performances of all ICM methods improve as N increases;
- the average false-alarm probability of the one-sided t test is *constant* and equal to the specified value of 5%, verifying the validity of (4.22b)–(4.22c);
- the false-alarm probabilities of the local nonparametric GLR tests attain the specified level of 5% *asymptotically* (i.e. for large N, see also Section IV-D);
- the Gaussian ICM method achieves the smallest false-alarm probability for all N (compared with the other methods).

Consider now the quantized Gaussian measurement scenario. In Fig. 10, we show the average probabilities of false alarm and miss for different methods as functions of N. Observe that

- as in the Gaussian scenario, the average false-alarm and miss error probabilities of all ICM methods decrease with N;
- the average false-alarm probabilities of the local t and nonparametric GLR tests attain the specified level of 5% for large N;
- for small N, the nonparametric ICM methods achieve smaller average false-alarm and miss error probabilities than the Gaussian ICM method;
- due to the averaging effect, the Gaussian ICM method performs well when N is large.

Note that the error-probability results presented in Figs. 9 and 10 do not show if the obtained event-region estimates were connected or not, which may be of interest in practical applications.

D. MRF Calibration

We utilize the calibration method in Section V to estimate the MRF model parameters a and η. The training data were generated by randomly placing $K = 1000$ nodes on a 50 m $\times$ 50 m grid and simulating noisy

[11] Here, the two error probabilities were estimated using the ideal decisions in Fig. 2 (right).

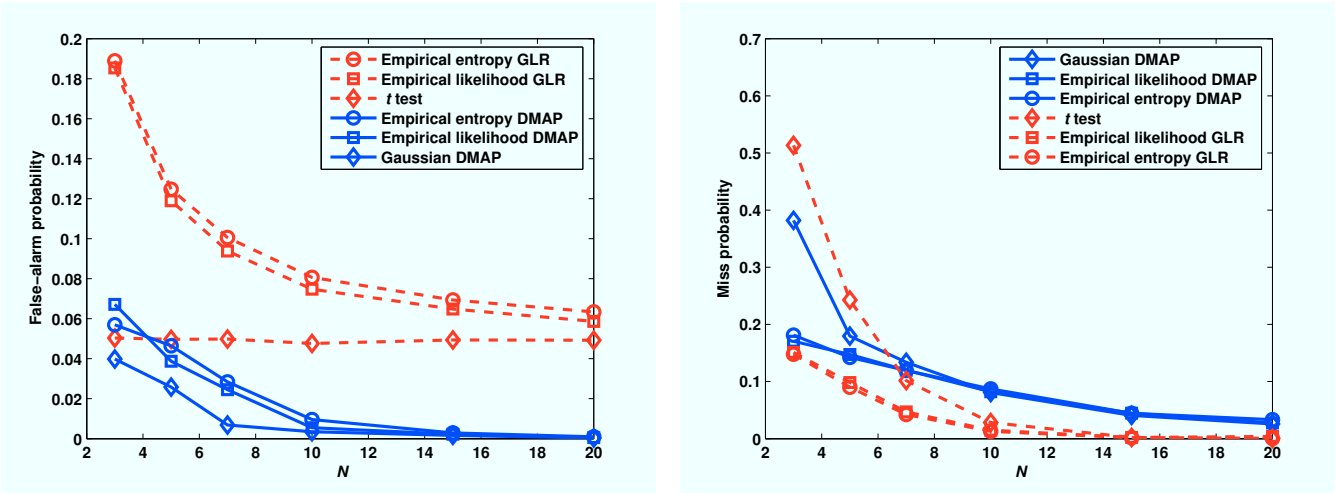

Fig. 9. Gaussian measurement scenario: Average probabilities of (left) false alarm and (right) miss, as functions of the number of observations per sensor N.

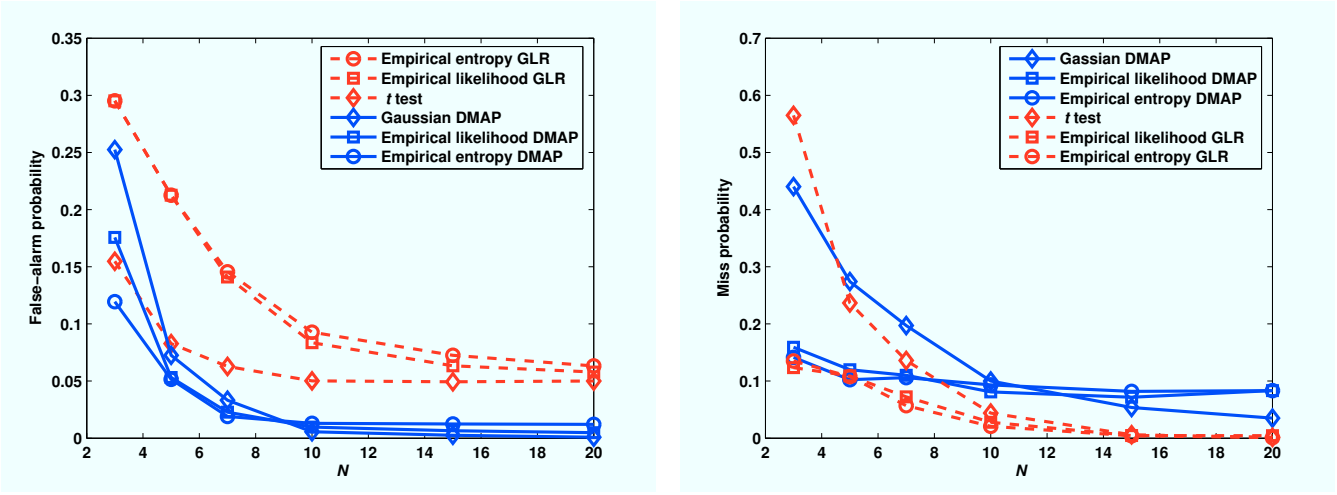

Fig. 10. Quantized Gaussian measurement scenario: Average probabilities of (left) false alarm and (right) miss, as functions of N.

realizations of a calibration field having constant mean $\mu > 0$ within a circular event region with radius 8 m, see Fig. 11. Twenty training data sets were generated by varying the noise realizations, node locations, and values of the event-region mean μ. We applied the calibration method proposed in Section V to fit each training data set and then averaged the obtained estimates, yielding the final calibration results. To obtain the average error probabilities in Figs. 9 and 10, the values of μ in the twenty training data sets were generated by sampling from the uniform$(0.4, 1.4)$ distribution. To calibrate the ICM algorithms whose results are shown in Figs. 4–8, we sampled μ from a wider range of values [following the uniform$(0.4, 3.4)$ distribution]; the resulting calibration provided smaller false-alarm probabilities and larger miss probabilities [compared with the results obtained by sampling μ from uniform$(0.4, 1.4)$].

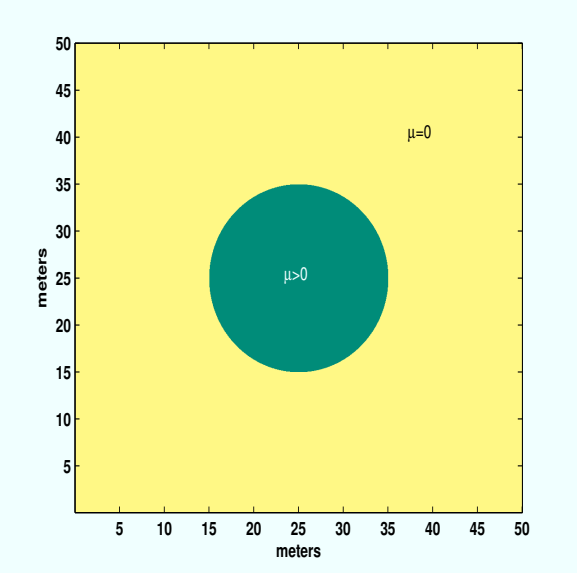

Fig. 11. Noiseless field used for calibration.

VII. Concluding Remarks

We presented an HMRF framework for distributed localized estimation and detection in sensor-network environments. We developed a calibration method for estimating the MRF model parameters from the training data and discussed initialization of the proposed algorithms. The proposed framework was applied to event-region detection.

Further research will include: extending the HMRF framework and ICM method to allow tracking of the field changes over time, analyzing the impact of communication errors (among the nodes) on the performance of the ICM method, comparing the ICM and message passing approaches, relaxing the conditional independence assumption in (2.3), developing data aggregation algorithms and energy-aware sensor-network design strategies for HMRFs (e.g., deciding which nodes will be in "alert" or "sleeping" modes), and studying asymptotic properties of the proposed methods as the number of measurements per node grows.

It is also of interest to relate the proposed ICM and *distributed consensus* approaches recently proposed in [24], [25], and [44]. If we select a Gaussian MRF model structure and modify the ICM iteration by replacing the measurements y_k with the estimates of the hidden field β_k from the previous ICM cycle, the resulting algorithm closely resembles the average-consensus scheme in [44, eq. (3)]. Note that the consensus methods estimate *global phenomena* (e.g., the mean field) whereas the ICM methods estimate *localized features,* which is an important distinction between the two approaches.

Since the autologistic MRF model may be too simplistic for many applications, it is important to develop more general process models that will allow utilizing multiple information bits to describe the hidden field

of interest. Here, it is of particular interest to derive physically based process models and corresponding ICM methods.

APPENDIX A. EMPIRICAL LIKELIHOOD AND CRB FOR ESTIMATING μ_k

We derive the concentrated empirical log-likelihood expression in (4.8). This derivation is similar to that in [32, Chs. 2.9 and 3.14] and [34, Sect. 2.4] and is given here for completeness. We utilize the method of Lagrange multipliers to solve the constrained optimization problem in (4.6): Define

$$G_k = \Big(\sum_{t=1}^{N} \ln p_{k,t} \Big) + \gamma_k \cdot \Big(\sum_{t=1}^{N} p_{k,t} - 1 \Big) - N\lambda_k \cdot \sum_{t=1}^{N} p_{k,t} \left[y_k(t) - \mu_k \right] \tag{A.1}$$

where γ_k and λ_k are Lagrange multipliers. Forming a weighted sum of the partial derivatives of G_k with respect to $p_{k,t}$ and setting the result to zero yields

$$0 = \sum_{t=1}^{N} p_{k,t} \frac{\partial G_k}{\partial p_{k,t}} = N + \gamma_k \tag{A.2}$$

where the second equality follows by using the constraints $\sum_{t=1}^{N} p_{k,t} = 1$ and $\sum_{t=1}^{N} p_{k,t} \left[y_k(t) - \mu_k \right]$. Therefore $\gamma_k = -N$ implying that

$$p_{k,t} = \frac{1}{N} \cdot \frac{1}{1 + \lambda_k \left[y_k(t) - \mu_k \right]} \tag{A.3}$$

where $\lambda_k = \lambda_k(\mu_k)$ is chosen as a solution to

$$\sum_{t=1}^{N} p_{k,t} \left[y_k(t) - \mu_k \right] = \frac{1}{N} \cdot \sum_{t=1}^{N} \frac{y_k(t) - \mu_k}{1 + \lambda_k \left[y_k(t) - \mu_k \right]} \tag{A.4a}$$

$$= 0. \tag{A.4b}$$

Substituting (A.3) into the multinomial log likelihood yields

$$\sum_{t=1}^{N} \ln p_{k,t} = -N \ln N + \Xi_k(\lambda_k; \mu_k) \tag{A.5}$$

where $\Xi_k(\lambda_k; \mu_k)$ was defined in (4.8b). To satisfy (A.4a), we need to minimize the above expression with respect to λ_k, yielding the convex dual formulation in (4.8). Assuming (4.5), all estimates of the multinomial probabilities need to satisfy

$$0 < p_{k,t} = \frac{1}{N} \cdot \frac{1}{1 + \lambda_k \left[y_k(t) - \mu_k \right]} < 1 \tag{A.6}$$

and (4.9) is obtained by using the second inequality in (A.6) for all $t \in \{1, 2, \ldots, N\}$. Finally, the first two derivatives of $\Xi_k(\lambda_k; \mu_k)$ with respect to λ_k are

$$\frac{\partial \Xi_k(\lambda_k; \mu_k)}{\partial \lambda_k} = -\sum_{t=1}^{N} \frac{y_k(t) - \mu_k}{1 + \lambda_k \left[y_k(t) - \mu_k \right]} \tag{A.7a}$$

$$\frac{\partial^2 \Xi_k(\lambda_k; \mu_k)}{\partial \lambda_k^2} = \sum_{t=1}^{N} \frac{\left[y_k(t) - \mu_k \right]^2}{\left\{ 1 + \lambda_k \left[y_k(t) - \mu_k \right] \right\}^2} \tag{A.7b}$$

and the Newton-Raphson iteration (4.10) follows.

Least Favorable Families and CRB for μ_k Under the Empirical Likelihood Model: We derive the CRB for μ_k under the empirical likelihood measurement-error model and sketch a proof that the empirical-likelihood approach employs a *least favorable nonparametric distribution family* for estimating μ_k.

We first differentiate the empirical log likelihood in (4.8a) with respect to μ_k:

$$\frac{dl_k(\mu_k)}{d\mu_k} = N\lambda_k(\mu_k) \tag{A.8}$$

which follows by using (A.3)–(A.4) and the constraint $\sum_{t=1}^{N} p_{k,t} = 1$. Then

$$\frac{d^2 l_k(\mu_k)}{d\mu_k^2} = N \cdot \frac{d\lambda_k(\mu_k)}{d\mu_k} \tag{A.9}$$

where $d\lambda_k(\mu_k)/d\mu_k$ can be computed by differentiating (A.4a) [with λ_k evaluated at $\lambda_k(\mu_k)$]:

$$\frac{d}{d\mu_k}\Big\{\sum_{t=1}^{N} \frac{y_k(t)-\mu_k}{1+\lambda_k(\mu_k)\,[y_k(t)-\mu_k]}\Big\} = 0 \tag{A.10a}$$

leading to

$$\begin{aligned}\frac{d\lambda_k(\mu_k)}{d\mu_k} \;=\; & \Big\{\sum_{t=1}^{N} \frac{[y_k(t)-\mu_k]^2}{\{1+\lambda_k(\mu_k)\,[y_k(t)-\mu_k]\}^2}\Big\}^{-1} \\ & \cdot\Big\{\lambda_k(\mu_k)\cdot\sum_{t=1}^{N} \frac{y_k(t)-\mu_k}{\{1+\lambda_k(\mu_k)\,[y_k(t)-\mu_k]\}^2} - N\Big\}\end{aligned} \tag{A.10b}$$

and, consequently,

$$\begin{aligned}-\frac{d^2 l_k(\mu_k)}{d\mu_k^2} \;=\; & N\Big\{\sum_{t=1}^{N} \frac{[y_k(t)-\mu_k]^2}{\{1+\lambda_k(\mu_k)\,[y_k(t)-\mu_k]\}^2}\Big\}^{-1} \\ & \cdot\Big\{N - \lambda_k(\mu_k)\cdot\sum_{t=1}^{N} \frac{y_k(t)-\mu_k}{\{1+\lambda_k(\mu_k)\,[y_k(t)-\mu_k]\}^2}\Big\}.\end{aligned} \tag{A.11}$$

Then, assuming the discrete uniform distribution of the observations $y_k(1), y_k(2), \ldots, y_k(N)$, the CRB for estimating μ_k is given by (4.11), which follows from the fact that the discrete uniform distribution of the observations implies $\mu_k = \overline{y}_k$ and $\lambda_k(\overline{y}_k) = 0$. Note that (4.11) closely resembles the well-known CRB expression for μ_k under the *parametric* Gaussian measurement-error model in Section IV-A.1 (see e.g. [45, eq. (3.9)]):

$$\mathrm{CRB_G} = \frac{\sigma_k^2}{N}. \tag{A.12}$$

In particular, (4.11) is a *good estimate* of (A.12). Hence, the empirical likelihood approach employs a *least favorable nonparametric distribution family* for estimating μ_k. This conclusion follows from the notion that a least favorable nonparametric family is one in which the estimation problem (i.e. estimating μ_k in our case) is "as hard as in a parametric problem" (corresponding to the Gaussian measurement-error model in the above example), see also the discussion in [32, Ch. 9.6], [33], [42, Sect. 2.3], and [46, Ch. 22.7].

APPENDIX B. EMPIRICAL ENTROPY AND CRB FOR ESTIMATING μ_k

We utilize Lagrange multipliers to solve the constrained optimization problem in (4.13) [subject to the constraints (4.4b)]: Define

$$\begin{aligned} G_k &= \sum_{t=1}^{N} N p_{k,t} \ln(N p_{k,t}) + \gamma_k \cdot \Big(\sum_{t=1}^{N} p_{k,t} - 1 \Big) - N\lambda_k \cdot \sum_{t=1}^{N} p_{k,t} \left[y_k(t) - \mu_k \right] \\ &= N \ln N + N \sum_{t=1}^{N} p_{k,t} \ln(p_{k,t}) + \gamma_k \cdot \Big(\sum_{t=1}^{N} p_{k,t} - 1 \Big) - N\lambda_k \cdot \sum_{t=1}^{N} p_{k,t} \left[y_k(t) - \mu_k \right] \end{aligned} \tag{B.1}$$

where γ_k and λ_k are Lagrange multipliers. Setting the partial derivatives of G_k with respect to $p_{k,t}$ to zero yields

$$N + \gamma_k + N \ln(p_{k,t}) - N\lambda_k \left[y_k(t) - \mu_k \right] = 0 \tag{B.2}$$

for $t = 1, 2, \ldots, N$. Finding γ_k that satisfies the constraint $\sum_{t=1}^{N} p_{k,t} = 1$ leads to the following expressions for the multinomial probabilities:

$$p_{k,t} = \frac{\exp\{\lambda_k \left[y_k(t) - \mu_k \right]\}}{\sum_{\tau=1}^{N} \exp\{\lambda_k \left[y_k(\tau) - \mu_k \right]\}} = \frac{\exp[\lambda_k \, y_k(t)]}{\sum_{\tau=1}^{N} \exp[\lambda_k \, y_k(\tau)]}. \tag{B.3}$$

Finally, the constraint $\sum_{t=1}^{N} p_{k,t} \left[y_k(t) - \mu_k \right] = 0$ is satisfied by finding $\lambda_k = \lambda_k(\mu_k)$ that solves

$$\sum_{t=1}^{N} \exp\{\lambda_k \left[y_k(t) - \mu_k \right]\} \cdot \left[y_k(t) - \mu_k \right] = 0. \tag{B.4}$$

Note that (B.4) is an increasing function of λ_k and that satisfying (B.4) is equivalent to minimizing $\zeta_k(\lambda_k; \mu_k)$ in (4.15) with respect to λ_k. Finally, the first two derivatives of $\zeta_k(\lambda_k; \mu_k)$ with respect to λ_k are

$$\frac{\partial \zeta_k(\lambda_k; \mu_k)}{\partial \lambda_k} = \sum_{t=1}^{N} \exp\{\lambda_k \left[y_k(t) - \mu_k \right]\} \cdot \left[y_k(t) - \mu_k \right] \tag{B.5a}$$

$$\frac{\partial^2 \zeta_k(\lambda_k; \mu_k)}{\partial \lambda_k^2} = \sum_{t=1}^{N} \exp\{\lambda_k \left[y_k(t) - \mu_k \right]\} \cdot \left[y_k(t) - \mu_k \right]^2 \tag{B.5b}$$

and the Newton-Raphson iteration (4.16) follows.

Least Favorable Families and CRB for μ_k Under the Empirical Entropy Model: We derive the CRB for μ_k under the empirical entropy measurement-error model and sketch a proof that the empirical-entropy approach employs a *least favorable nonparametric distribution family* for estimating μ_k.

We first differentiate the nonparametric log likelihood (4.17) for the empirical entropy model with respect to μ_k:

$$\frac{dl_k(\mu_k)}{d\mu_k} = N \cdot \frac{d\lambda_k(\mu_k)}{d\mu_k} \cdot (\overline{y}_k - \mu_k). \tag{B.6}$$

To derive (B.6), we have used the identity:

$$\sum_{t=1}^{N} \exp[\lambda_k(\mu_k) \, y_k(t)] \cdot \left[y_k(t) - \mu_k \right] = 0 \tag{B.7}$$

which follows from (B.4). We now compute $d\lambda_k(\mu_k)/d\mu_k$ by differentiating (B.7):

$$\frac{d}{d\mu_k}\Big\{\sum_{t=1}^{N}\exp[\lambda_k(\mu_k)\,y_k(t)]\cdot[y_k(t)-\mu_k]\Big\}=0 \tag{B.8a}$$

leading to

$$\frac{d\lambda_k(\mu_k)}{d\mu_k}=\frac{\sum_{t=1}^{N}\exp[\lambda_k(\mu_k)\,y_k(t)]}{\sum_{t=1}^{N}\exp[\lambda_k(\mu_k)\,y_k(t)]\cdot[y_k(t)-\mu_k]^2} \tag{B.8b}$$

where we have used (B.7) to obtain (B.8b). Finally,

$$\frac{d^2l_k(\mu_k)}{d\mu_k^2}=N\cdot\frac{d^2\lambda_k(\mu_k)}{d\mu_k^2}\cdot(\overline{y}_k-\mu_k)-N\cdot\frac{d\lambda_k(\mu_k)}{d\mu_k}. \tag{B.9}$$

Then, assuming the discrete uniform distribution of the observations $y_k(1), y_k(2), \ldots, y_k(N)$, we have $\mu_k = \overline{y}_k$, $\lambda_k(\overline{y}_k)=0$, and

$$-\frac{d^2l_k(\overline{y}_k)}{d\mu_k^2}=\frac{N}{s_k^2} \tag{B.10}$$

which follows by using (B.8b). Therefore, (4.11) holds, implying that estimating μ_k is as hard as in a parametric Gaussian model and, consequently, the empirical entropy approach employs a *least favorable nonparametric distribution family* (see also Appendix A).

APPENDIX C. ICM DETECTOR FOR THE GAUSSIAN MEASUREMENT-ERROR MODEL

Under the Gaussian measurement-error model (4.2) in Section IV-A.1, the conditional predictive log likelihoods in (3.1) simplify to

$$\begin{aligned}
L_k\big(1\,|\,\mathcal{N}_\beta(k)\big) &= \max_{\mu_k>0,\,\sigma_k^2}\Big\{\sum_{t=1}^{N}\ln p_{\mathrm{noise}_k}\Big(y_k(t)-\mu_k\,;\,\sigma_k^2\Big)\Big\}+\ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(1|\mathcal{N}_\beta(k)\big)\\
&= \begin{cases} -N/2-(N/2)\cdot\ln(s_k^2)+\ln p_{\beta_k|\mathcal{N}(k)}\big(1|\mathcal{N}(k)\big), & \overline{y}_k>0\\ -N/2-(N/2)\cdot\ln(s_{0,k}^2)+\ln p_{\beta_k|\mathcal{N}(k)}\big(1|\mathcal{N}_\beta(k)\big), & \overline{y}_k\le 0\end{cases} \qquad \text{(C.1a)}\\
L_k\big(0\,|\,\mathcal{N}_\beta(k)\big) &= \max_{\sigma_k^2}\Big\{\sum_{t=1}^{N}\ln p_{\mathrm{noise}_k}\Big(y_k(t)\,;\,\sigma_k^2\Big)\Big\}+\ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(0|\mathcal{N}(k)\big)\\
&= -N/2-(N/2)\cdot\ln(s_{0,k}^2)+\ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(0|\mathcal{N}_\beta(k)\big) \qquad \text{(C.1b)}
\end{aligned}$$

and (4.20b) follows.

APPENDIX D. ICM DETECTOR FOR NONPARAMETRIC MEASUREMENT-ERROR MODELS

We specialize Step (ICM2) of the ICM algorithm to the nonparametric measurement-error models in Section IV-A.2. Here, the conditional predictive log likelihoods in (3.1) simplify to

$$L_k\big(1\,|\,\mathcal{N}_\beta(k)\big) = \max_{\mu_k>0}\{l_k(\mu_k)\}+\ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(1|\mathcal{N}_\beta(k)\big) \tag{D.1a}$$

$$L_k\big(0\,|\,\mathcal{N}_\beta(k)\big) = l_k(0)+\ln p_{\beta_k|\mathcal{N}_\beta(k)}\big(0|\mathcal{N}_\beta(k)\big). \tag{D.1b}$$

We now show that, for $\kappa = 0$ and $\kappa = -1$

$$\max_{\mu_k>0}[l_k(\mu_k)] = \begin{cases} -N\ln N, & \overline{y}_k > 0 \\ l_k(0), & \overline{y}_k \leq 0 \end{cases} \tag{D.2}$$

Proof of (D.2) for Empirical Likelihood: Consider the empirical likelihood model ($\kappa = 0$). Then, the result for $\overline{y}_k > 0$ follows from (4.7a).

We now focus on the case where $\overline{y}_k \leq 0$. Then, for $\mu_k > 0$, the expression in (A.4a) is negative at $\lambda_k = 0$. Since (A.4a) is a decreasing function of λ_k, the optimal λ_k which solves (A.4) for any $\mu_k > 0$ must be negative. Then, (A.8) implies that, in this case, $l_k(\mu_k)$ is a decreasing function of μ_k and (D.2) follows.

Proof of (D.2) for Empirical Entropy: Consider now the empirical entropy model ($\kappa = -1$). Then, the result for $\overline{y}_k > 0$ follows by noting that

- $\lambda_k(\overline{y}_k) = 0$ solves (B.4) and
- the nonparametric log likelihood for the empirical entropy model is maximized at $\mu_k = \overline{y}_k$, which follows by setting $dl_k(\mu_k)/d\mu_k$ in (B.6) to zero and noting that $d\lambda_k(\mu_k)/d\mu_k$ is always positive [see (B.8b)].

In the case where $\overline{y}_k \leq 0$ and $\mu_k > 0$, the derivative $dl_k(\mu_k)/d\mu_k$ in (B.6) is negative. Therefore, $l_k(\mu_k)$ is a decreasing function of μ_k and (D.2) follows.

Finally, substituting (D.1) and (D.2) into (4.24a) yields (4.24b).

APPENDIX E. GLR TESTS FOR μ_k UNDER NONPARAMETRIC MEASUREMENT-ERROR MODELS

We derive the empirical likelihood and entropy GLR tests in Section IV-D. Under the null hypotheses $\mathrm{H}_{0,k}$: $\mu_k = 0$, the asymptotic distribution of the GLR test statistics

$$2\max_{\mu_k>0}\{l_k(\mu_k)]\} - 2\,l_k(0) = [-2N\ln N - 2\,l_k(0)] \cdot i_{[0,\infty)}(\overline{y}_k) \tag{E.1}$$

is given by, for $l \geq 0$,

$$\lim_{N\to\infty} P(l_k \leq l) = \tfrac{1}{2}\,P(\chi_1^2 \leq l) + \tfrac{1}{2} \tag{E.2}$$

which follows by adapting the results in [32, Th. 2.2], [34, Sect. 2.3] (for empirical likelihood) and [41], [42] (for empirical entropy). to the one-sided testing problem in Section IV. Here, χ_1^2 denotes a random variable having a central χ^2 distribution with one degree of freedom and can be obtained by squaring a standard normal random variable. The second term in (E.2) corresponds to the probability that $\overline{y}_k < 0$ under $\mathrm{H}_{0,k}$, which is 1/2; in this case, the GLR test statistics (E.1) becomes zero.

Note that (4.25a) follows by using the square root of (E.1) as the test statistics, which is possible because $-N\ln N - l_k(0)$ are non-negative. Then, (E.2) implies that a specified false-alarm probability P_{FA} will be

achieved by comparing

$$\sqrt{2\left[-N\ln N - l_k(0)\right]}\cdot i_{[0,\infty)}(\overline{y}_k) \tag{E.3}$$

with the threshold τ_{NP}, computed using (4.25b).

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers for their helpful comments.

REFERENCES

[1] F. Zhao and L.J. Guibas, *Wireless Sensor Networks: An Information Processing Approach*, San Francisco, CA: Morgan Kaufmann, 2004.
[2] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks," in *Proc. 5th ACM Annu. Int. Conf. Mobile Computing and Networking,* Seattle, WA, Aug. 1999, pp. 263–270.
[3] C.-Y. Chong and S.P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proc. IEEE,* vol. 91, pp. 1247–1256, Aug. 2003.
[4] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
[5] *Embedded, Everywhere: A Research Agenda for Networked Systems of Embedded Computers,* National Academy Press, 2001, report of the National Research Council Committee on Networked Systems of Embedded Computers. Available online: http://books.nap.edu/html/embedded_everywhere.
[6] B.J. Feder, "Wireless sensor networks spread to new territory," *New York Times,* July 26, 2004.
[7] D. Culler, D. Estrin, and M. Srivastava, "Overview of sensor networks," *IEEE Computer,* vol. 37, pp. 41–49, Aug. 2004.
[8] J. Elson and D. Estrin, "Sensor networks: A bridge to the physical world," in *Wireless Sensor Networks,* C.S. Raghavendra, K.M. Sivalingam, and T. Znati, (Eds.), Boston, MA: Kluwer, 2004, ch. 1, pp. 3–20.
[9] N.A.C. Cressie, *Statistics for Spatial Data*, revised ed., New York: Wiley, 1993.
[10] J. Besag, "On the statistical analysis of dirty pictures," *J. R. Stat. Soc., Ser. B,* vol. 48, pp. 259–302, 1986.
[11] A.S. Willsky, "Multiresolution Markov models for signal and image processing," *Proc. IEEE,* vol. 90, pp. 1396–1458, Aug. 2002.
[12] A. Dogandžić and B. Zhang, "Distributed signal processing for sensor networks using hidden Markov random fields," in *Proc. 39th Annu. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 2005, pp. WA7.7–WA7.12.
[13] Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood,* New York: Oxford University Press, 2001.
[14] B. Krishnamachari and S. Iyengar, "Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks," *IEEE Trans. Computers*, vol. 53, pp. 241–250, March 2004.
[15] Z.-Q. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Select. Areas Commun.,* vol. 23, pp. 735–744, Apr. 2005.
[16] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2210–2219, June 2005.
[17] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A.J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Processing*, vol. 54, pp. 413–422, Feb. 2006.
[18] K. Plarre and P.R. Kumar, "Extended message passing algorithm for inference in loopy Gaussian graphical model," *Ad Hoc Networks*, vol. 2, pp. 153–169, Apr. 2004.
[19] J.W. Fisher, M.J. Wainwright, E.B. Sudderth, and A.S. Willsky, "Statistical and information-theoretic methods for self-organization and fusion of multimodal, networked sensors," *Int. J. High Performance Comput. Appl.,* vol. 15, pp. 337–353, Fall 2002.
[20] E.B. Sudderth, M.J. Wainwright, and A.S. Willsky, "Embedded trees: Estimation of Gaussian processes on graphs with cycles," *IEEE Trans. Signal Processing,* vol. 54, pp. 3136–3150, Nov. 2004.
[21] S.L. Lauritzen, *Graphical Models,* New York: Oxford University Press, 1996.
[22] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufmann, 1988.
[23] V. Delouille, R. Neelamani, V. Chandrasekaran, and R.G. Baraniuk, "The embedded triangles algorithm for distributed estimation in sensor networks," in *Proc. 2003 IEEE Workshop Stat. Signal Process.,* St. Louis, MO, Sept. 2003, pp. 371–374.
[24] S. Venkatesh and M. Alanyali, "M-ary hypothesis testing in sensor networks," in *Proc. 38th Annu. Conf. Inform. Sci. Syst.*, Princeton, NJ, March 2004, pp. 696–701.
[25] D.S. Scherber and H.C. Papadopoulos, "Distributed computation of averages over ad hoc networks," *IEEE J. Select. Areas Commun.,* vol. 23, pp. 776–787, Apr. 2005.
[26] D.V. Lindley and A.F.M. Smith, "Bayes estimates for the linear model," *J. R. Stat. Soc., Ser. B,* vol. 34, pp. 1–41, 1972.
[27] M.-H. Chen, Q.-M. Shao, and J.G. Ibrahim, *Monte Carlo Methods in Bayesian Computation,* New York:Springer-Verlag, 2000.
[28] D.B. Rowe, *Multivariate Bayesian Statistics*, New York: Chapman & Hall, 2002.

[29] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Stat. Prob,.* vol. 1, pp. 547–561, 1961.
[30] N. Cressie and T.R.C. Read, "Multinomial goodness-of-fit tests," *J. R. Stat. Soc., Ser. B,* vol. 46, pp. 440–464, 1984.
[31] T.R.C. Read and N.A.C. Cressie, *Goodness-of-fit Statistics for Discrete Multivariate Data,* New York:Springer-Verlag, 1988.
[32] A.B. Owen, *Empirical Likelihood*, New York: Chapman & Hall, 2001.
[33] C. Stein, "Efficient nonparametric testing and estimation," in *Proc. Third Berkeley Symp. Mathematical Statistics and Probability,* J. Neyman (Ed.), University of California Press, 1956, vol. 1, pp. 187–195.
[34] P. Hall and B. La Scala, "Methodology and algorithms of empirical likelihood," *Int. Statist. Rev.,* vol. 58, pp. 109–127, Aug. 1990.
[35] W.H. Press, S.A. Teukolsky, W.T. Vettering, and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing,* Second ed., Cambridge, UK: Cambridge Univ. Press, 1992.
[36] T.M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.
[37] B. Efron, "Nonparametric standard errors and confidence intervals," *Canad. J. Statist.,* vol. 9, no. 2, pp. 139–172, Aug. 1981.
[38] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans,* Philadelphia, PA: SIAM, 1982.
[39] A.C. Davison and D.V. Hinkley, *Bootrstrap Methods and their Application,* Cambridge, UK: Cambridge Univ. Press, 1997.
[40] P.D. Sampson and P. Guttorp, "Nonparametric estimation of nonstationary spatial covariance structure," *J. Amer. Stat. Assoc.,* vol. 87, pp. 108–119, March 1992.
[41] S.M.S. Lee and G.A. Young, "Nonparametric likelihood ratio confidence intervals," *Biometrika,* vol. 86, pp. 107–118, Mar. 1999.
[42] T.J. DiCiccio and J.P. Romano, "Nonparametric confidence limits by resampling methods and least favorable families," *Int. Statist. Rev.,* vol. 58, pp. 59–76, Apr. 1990.
[43] J. Besag, "Statistical analysis of non-lattice data," *The Statistician,* vol. 24, pp. 179–195, 1975.
[44] L. Xiao, S. Boyd, and S. Lall, "A scheme for asynchronous distributed sensor fusion based on average consensus," in *Proc. 4th International Symposium on Information Processing in Sensor Networks.,* Berkeley, CA, Apr. 2005, pp. 63–70.
[45] S.M. Kay, *Fundamentals of Statistical Signal Processing — Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.
[46] B. Efron and R.J. Tibshirani, *An Introduction to the Bootstrap*, New York: Chapman & Hall, 1993.

Aleksandar Dogandžić (S'96–M'01) received the Dipl. Ing. degree (*summa cum laude*) in Electrical Engineering from the University of Belgrade, Yugoslavia, in 1995, and the M.S. and Ph.D. degrees in electrical engineering and computer science from the University of Illinois at Chicago (UIC) in 1997 and 2001, respectively, under the guidance of Professor Arye Nehorai.

In August 2001, he joined the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA, as an Assistant Professor. His research interests are in statistical signal processing theory and applications.

Dr. Dogandžić received the 2003 Young Author Best Paper Award and 2004 Signal Processing Magazine Award by the IEEE Signal Processing Society. In 2006, he received the CAREER Award by the National Science Foundation and holds the 2006–2007 Litton Assistant Professorship in Electrical and Computer Engineering at the Iowa State University.

Benhong Zhang was born in Shanghai, China in 1980. He received the B.S. degree in Electronic Engineering from Tsinghua University, Beijing, China in 2002. He is a Ph.D. candidate in the Department of Electrical and Computer Engineering, Iowa State University.

His research interests are in statistical signal processing and applications.

LIST OF FIGURES