Optimum Level Set Estimation of a Time-varying Random Field Under a Power Constraint

Zuoen Wang*, Jingxian Wu*, Jing Yang*, and Hai Lin†
* Department of Electrical Engineering, University of Arkansas, Fayetteville, AR 72701, USA
†Graduate School of Engineering, Osaka Prefecture University, Osaka, 599-8531, Japan

***Abstract*—Level set estimation (LSE) is the process of using noisy observations of an unknown function to estimate the region(s) where the function values lie above a given threshold. It has a wide range of applications in many scientific and engineering areas, such as spectrum sensing or environment monitoring. In this paper, we study the optimum LSE of a time-varying random field under a total power constraint. A sensor performs uniform sampling of the random field and sends the samples to a fusion center, which estimates the level set by using distorted observations of the samples. Under a total power constraint, a higher sampling rate means less energy per sample, which may negatively impact the estimation performance, but also a stronger correlation between adjacent samples, which can improve the estimation accuracy. Thus it is critical to identify the optimum sampling rate that can minimize the LSE error provability. With the help of a Gaussian process (GP) prior model, we first develop an optimum LSE algorithm based on GP regression. The exact analytical LSE error probability of the LSE algorithm is then derived by considering a number of factors, such as the power consumptions of both sensing and transmission, the power constraint of the sensor, the sampling rate, and the probability distributions of the random field. To simplify analysis, we also obtain a closed-form upper bound of the LSE error probability. The optimum sampling rate is identified by using the analytical error probabilities.**

I. Introduction

Level set estimation (LSE) is the process of using noisy observations of an unknown function defined on a Hilbert space to estimate the region(s) where the function amplitude lies above a given threshold. It has a wide range of applications in many scientific and engineering areas. For example, the objective of spectrum sensing in cognitive radio networks is to identify the boundary of "spectrum holes" in the space, time, and frequency domains [1]. Other applications include the monitoring of pollution [2] or the contours or sunlight, temperature, or rainfalls for biosystem ecology tracking [3], etc.

LSE can be performed by applying standard binary classifications to the implicit function using probability models [4]–[6]. Such a binary classification approach ignores the difference between the actual function value and the threshold, and such information contains salient information that can improve the LSE accuracy. Another popular approach is to estimate the values of the underlying functions through regression, and then obtain the level set by thresholding the estimated function values at the critical value [7], [8].

Most LSE methods are applicable in a static setting, that is, the measurements in the field are given or passively provided [8]–[10]. Recently, it has been proposed to dynamically adjust the sensing strategy based on past sensing results [11]–[13] by using poop-based active learning [14]. The dynamic LSE employs sequential decision makings, and it can accurately track the level set in a time-varying random field. None of the above mentioned works consider the constraints imposed by the limited energy supply, which is one of the main performance limiting factors in wireless sensing systems.

In this paper, we study the optimum LSE of a time-varying random event in a power-constrained wireless sensing system. A wireless sensor samples the time-varying random event and transmits the discrete-time samples to a fusion center (FC), which performs LSE by using distorted observations of the discrete-time samples. The sampling rate of the wireless sensor plays a critical role in the LSE performance. A higher sampling rate means a stronger correlation between adjacent samples, and this can positively impact the LSE performance. On the other hand, under a total power constraint, a higher sampling rate results in less energy per sample, or a lower signal-to-noise ratio (SNR) per sample at the FC, and this will negatively affect the estimation performance. Thus it is critical to identify the optimum sampling rate that can minimize the LSE error. The optimum sensor density or sampling rate that can minimize the estimation mean squared error (MSE) of a random field has been studied in [15], [16]. The LSE problem studied in this paper is different from [15], [16], in that our objective is not to reconstruct the entire function, but to estimate the level set of the underlying function.

We introduce a Gaussian process (GP) prior model to capture the temporal correlation inherent in the random field [17]. Under the GP framework, we first show that the optimal level set estimation can be achieved by performing a GP regression with all discrete-time data samples and then thresholding the regression results. With the GP regression based LSE algorithm, the exact analytical LSE error probability is expressed as a function of a number of parameters, such as the sampling rate, the power constraint at the sensor, the power consumption of both sensing and transmission, and the temporal correlation of the random field, etc. To simplify analysis, we also obtain a closed-form upper bound of the LSE error probability. The optimum sampling rate is obtained by using the analytical LSE error probability results.

978-1-4799-5952-5/15/$31.00 ©2015 IEEE

II. System Model

We consider the sensing and monitoring of the level set of a time-varying random event, $x(t)$, where t is the time variable. The random event can be used to model temperature, air pressure, or density of toxic gases, etc.

Assumptions 1 *We make the following assumptions about the random event $x(t)$:*

1) *The prior distribution of $\{x(t)\}$ is a zero-mean Gaussian process (GP) with covariance function $k(t,t') = \mathbb{E}[x(t)x(t')]$, i.e, $x \sim GP(0, k(t,t'))$.*
2) *The Gaussian process is wide sense stationary in time, and*

$$k(t,t') = \rho^{|t-t'|}, \tag{1}$$

where $\rho \in [0,1]$ is the temporal correlation coefficient.

At time t , we are interested in identifying the γ-level set of $\{x(t)\}$, which is defined as

$$\mathcal{S}(t) := \{t' \in [0,t] : x(t') > \gamma\} \tag{2}$$

We assume $\gamma > 0$ without loss of generality.

The level set will be estimated by using distorted observations of the random event. Due to energy limit, the sensing system can only take discrete-time samples. The collected discrete-time samples are transmitted to a FC. Denote the sampling instants as $t_i = id$, for $i = 1, 2, \cdots$, where d is the sampling period. It is assumed that the sampling operation consumes an energy of $E_0 = E_c + E$ joules, where the constant E_c is due to hardware power consumption of the sensing operation, and E is the transmission energy of a sample. The samples observed at the FC can be represented as

$$y_i = \sqrt{E} x_i + \xi_i \tag{3}$$

where $x_i = x(t_i)$, and ξ_i includes the effects of observation noise and channel distortions. It is assumed that ξ_i is zero mean Gaussian distributed with variance σ^2.

The sensor operates under the constraint of a fixed power P_0. The energy allocated to the one sample is thus $E_0 = P_0 d$. Consequently, the transmission energy per sample is $E = P_0 d - E_c$.

At time t, the FC will obtain an estimated level set, $\hat{\mathcal{S}}(t)$, by using the set of discrete-time samples, $\{y_i | t_i \le t\}$. We define the level set estimation error at time t as the symmetric difference between the level set of interest, $\mathcal{S}(t)$, and the estimated level set, $\hat{\mathcal{S}}(t)$:

$$e(\hat{\mathcal{S}}(t)) := \int_0^t \mathbb{I}\left\{\mu \in \Delta(\mathcal{S}(t), \hat{\mathcal{S}}(t))\right\} d\mu \tag{4}$$

where $\Delta(\mathcal{S}(t), \hat{\mathcal{S}}(t)) = (\mathcal{S}(t) \cap \hat{\mathcal{S}}^c(t)) \cup (\mathcal{S}^c(t) \cap \hat{\mathcal{S}}(t))$ denotes the symmetric difference, $\mathcal{S}^c$ is the complement of $\mathcal{S}$, and $\mathbb{I}\{E\} = 1$ if event E is true and 0 otherwise.

The level set estimation (LSE) problem can then be formulated as

$$\begin{aligned} \min . \quad & \lim_{t\to\infty} \frac{1}{t} \mathbb{E}[e(\hat{\mathcal{S}}(t))] \\ \text{s.t.} \quad & E = P_0 d - E_c,\ d \ge \frac{P_0}{E_c} \end{aligned} \tag{5}$$

The optimization problem involves two steps: first, how to choose the sampling period d to obtain the samples $\{y_i\}_i$; second, once $\{y_i\}_i$ is given, how to estimate the level set by using the knowledge of $\{y_i\}_i$.

Before identifying the optimum sampling period, we first study in Section III the optimum estimation of level set once $\{y_i\}_i$ are known at the receiver. The results will provide the analytical form of the conditional LSE error probability given $\{y_i\}_i$, which can be used to facilitate the optimum sampling design in Section IV.

III. Optimal Level Set Estimation in GP

In this section, we study the optimal level set estimation if the the discrete-time samples, $\mathbf{y}_n = [y_1, y_2, \ldots, y_n]^T$, are known at the FC. This is the operation performed at the FC after the sampling period has already been chosen, and how to choose the sampling period will be discussed in the next section.

Define $\mathbf{r}_{x\mathbf{y}_n}(t) := \mathbb{E}[x(t)\mathbf{y}_n^T] \in \mathcal{R}^n$ and $\mathbf{R}_{\mathbf{y}_n\mathbf{y}_n} := \mathbb{E}[\mathbf{y}_n\mathbf{y}_n^T] \in \mathcal{R}^{n\times n}$, where $\mathcal{R}$ is the set of real numbers. From (1) and (3), the i-th element of the vector $\mathbf{r}_{x\mathbf{y}_n}(t)$ is $\sqrt{E}k(t,t_i)$, and the (i,j)-th element of the matrix $\mathbf{R}_{\mathbf{y}_n\mathbf{y}_n}$ is $\sqrt{E_iE_j}k(t_i,t_j)+\sigma^2\delta_{ij}$, with $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

Due to the GP modeling, given $\mathbf{y}_n$, the distribution of $x(t)$ is still Gaussian, with mean $\hat{m}_n(t)$ and variance $\hat{k}_n(t)$ given by

$$\hat{m}_n(t) = \mathbf{r}_{x\mathbf{y}_n}(t)\mathbf{R}^{-1}_{\mathbf{y}_n\mathbf{y}_n}\mathbf{y}_n \tag{6}$$

$$\hat{k}_n(t) = k(t,t) - \mathbf{r}_{x\mathbf{y}_n}(t)\mathbf{R}^{-1}_{\mathbf{y}_n\mathbf{y}_n}\mathbf{r}_{x\mathbf{y}_n}(t)^T. \tag{7}$$

The GP regression based LSE algorithm is given in Algorithm 1.

Algorithm 1 GP regression based level set estimation

1: Input: $\mathbf{y}_n$
2: Run GP regression for $\forall t \in [0, t_n]$:

$$\hat{m}_n(t) := \mathbf{r}_{x\mathbf{y}_n}(t)\mathbf{R}^{-1}_{\mathbf{y}_n\mathbf{y}_n}\mathbf{y}_n$$

3: Threshold $\hat{m}_n(t)$:

$$\hat{\mathcal{S}}(t_n) = \{t \in [0,t_n] : \hat{m}(t) > \gamma\}$$

4: Output $\hat{\mathcal{S}}(t_n)$.

Theorem 1 *Algorithm 1 is optimal with given $\{\mathbf{y}_n\}$, i.e., it minimizes the conditional LSE error, $\mathbb{E}[e(\hat{\mathcal{S}}(t_n))|\mathbf{y}_n]$.*

Proof: Under the GP modeling on $x(t)$ and $y(t)$, $\mathcal{S}(t_n)$ is also a random process. Given the observation history $\mathbf{y}_n$, we can always obtain a posterior distribution of $\mathcal{S}(t_n)$.

Given $\mathbf{y}_n$, the expected LSE error can be calculated as

$$\begin{aligned} \mathbb{E}[e(\hat{\mathcal{S}}(t_n))|\mathbf{y}_n] &:= \int_0^{t_n} \mathbb{P}\left[t \in \Delta(\mathcal{S}(t_n), \hat{\mathcal{S}}(t_n))\middle|\, \mathbf{y}_n\right] dt \\ &= \int_0^{t_n} \Big(\mathbb{P}\left[x(t) \le \gamma |\, \mathbf{y}_n\right] \cdot \mathbb{I}\{t \in \hat{\mathcal{S}}(t_n)\} \\ &\quad + \mathbb{P}\left[x(t) > \gamma |\, \mathbf{y}_n\right] \cdot \mathbb{I}\{t \in \hat{\mathcal{S}}^c(t_n)\}\Big) dt \end{aligned} \tag{8}$$

Therefore, the optimal estimator that minimizes (8) is to let

$$t \in \begin{cases} \hat{\mathcal{S}}(t_n) & \text{if } \mathbb{P}\left[x(t) > \gamma | \mathbf{y}_n\right] > \mathbb{P}\left[x(t) \le \gamma | \mathbf{y}_n\right] \\ \hat{\mathcal{S}}^c(t_n) & \text{if } \mathbb{P}\left[x(t) > \gamma | \mathbf{y}_n\right] \le \mathbb{P}\left[x(t) \le \gamma | \mathbf{y}_n\right] \end{cases} \quad (9)$$

for every $t \in [0, t_n]$.

Since $x(t)$ given $\mathbf{y}_n$ is still Gaussian distributed with mean and variance given in (6) and (7), we have

$$\mathbb{P}\left[x(t) > \gamma | \mathbf{y}_n\right] = Q\left(\frac{\gamma - \hat{m}_n(t)}{\sqrt{\hat{k}_n(t)}}\right) \quad (10)$$

where $Q(\cdot)$ is the Gaussian-Q function. The optimal estimator defined in (9) is then reduced to compare $\hat{m}_n(t)$ with γ. If $\hat{m}_n(t) > \gamma$, the probability in (10) is greater than 1/2, thus, we should let $t \in \hat{\mathcal{S}}(t_n)$; otherwise, we let $t \in \hat{\mathcal{S}}^c(t_n)$. ■

A byproduct of the proof of Theorem 1 is the conditional LSE error given $\mathbf{y}_n$, and the result is given as follows.

Corollary 1 *The minimum conditional LSE error with given* $\mathbf{y}_n$ *is*

$$\mathbb{E}[e(\hat{\mathcal{S}}(t_n))|\mathbf{y}_n] = \int_0^{t_n} Q\left(\frac{|\gamma - \hat{m}_n(t)|}{\sqrt{\hat{k}_n(t)}}\right) dt \quad (11)$$

The results in Theorem 1 and Corollary 1 will be used to facilitate the design of uniform sampling in the next section.

IV. Optimal Uniform Sampling

In this section, we will first find the unconditional LSE error probability using the results from Algorithm 1. The analytical results will then be used to identify the optimum sampling rate that can minimize the LSE error probability.

A. LSE Error Probability

The cost function in (5) is the time-averaged unconditional error probability. The unconditional LSE error probability can be alternatively expressed as

$$\mathbb{E}[e(\hat{\mathcal{S}}(t))] = \mathbb{E}_{\mathbf{y}_n}\left\{\mathbb{E}[e(\hat{\mathcal{S}}(nd))|\mathbf{y}_n]\right\} \quad (12)$$

We take another layer of expectation with respect to $\mathbf{y}_n$ on the right hand side (RHS) of (12), because the values of $\mathbf{y}_n$ are not yet known to the sensor during the optimum designs.

From (12) and Corollary 1, the cost function depends on the posteriori mean $\hat{m}_n(t)$ and variance $\hat{k}_n(t)$. From (6) and (7), we have

$$\hat{m}_n(t) = \sqrt{E}\mathbf{r}_{x\mathbf{x}_n}(t)\left(E\,\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n} + \sigma^2\mathbf{I}_n\right)^{-1}\mathbf{y}_n \quad (13)$$

$$\hat{k}_n(t) = k(t,t) - E\,\mathbf{r}_{x\mathbf{x}_n}(t)\left(E\,\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n} + \sigma^2\mathbf{I}_n\right)^{-1}\mathbf{r}^T_{x\mathbf{x}_n}(t). \quad (14)$$

where $\mathbf{r}_{x\mathbf{x}_n}(t) = \mathbb{E}\left[x(t)\mathbf{x}_n\right] \in \mathcal{R}^n$ and $\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n} = \mathbb{E}[\mathbf{x}_n\mathbf{x}_n^T] \in \mathcal{R}^{n\times n}$, with $\mathbf{x}_n = [x(d), x(2d), \cdots, x(nd)]^T \in \mathcal{R}^n$.

The posteriori mean is a function of $\mathbf{y}_n$, whereas the posteriori variance is a constant independent of $\mathbf{y}_n$. Since $\mathbf{y}_n$ is zero-mean Gaussian distributed, it can be easily shown that $\hat{m}_n(t)$ is zero-mean Gaussian distributed with variance being

$$\sigma^2_{\hat{m}_n}(t) = E\,\mathbf{r}_{x\mathbf{x}_n}(t)\left(E\,\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n} + \sigma^2\mathbf{I}_n\right)^{-1}\mathbf{r}^T_{x\mathbf{x}_n}(t) \quad (15)$$

From (14) and (15), we have $\hat{k}_n(t) = k(t,t) - \sigma^2_{\hat{m}}(t)$.

The variance in (15) depends on a number of factors, such as the correlation coefficient ρ, the sampling period d, the energy per sample E, and the time instant t. As $n \to \infty$, we have the following asymptotic results of $\sigma^2_{\hat{m}_n}(t)$ and $\hat{k}_n(t)$.

Theorem 2 *Define the asymptotic posteriori variance* $\sigma^2_e(\mu) := \lim_{n\to\infty}\hat{k}_n(t)$*, where* $\mu = \frac{t}{d} - \lfloor\frac{t}{d}\rfloor \in [0,1]$ *is the relative position of* t *between two adjacent samples. We have*

$$\sigma^2_e(\mu) = \left[\frac{1}{\gamma_0(d-d_c)} + \frac{1+\rho^{2d}-\rho^{2\mu d}-\rho^{2(1-\mu)d}}{1-\rho^{2d}}\right] \left(\frac{1}{\gamma_0(d-d_c)} + \frac{1-\rho^d}{1+\rho^d}\right)^{-\frac{1}{2}} \left(\frac{1}{\gamma_0(d-d_c)} + \frac{1+\rho^d}{1-\rho^d}\right)^{-\frac{1}{2}} \quad (16)$$

where $\gamma_0 := \frac{P_0}{\sigma^2}$ *being the SNR, and* $d_c := \frac{E_c}{P_0}$ *is hardware energy normalized by the average power constraint. In addition,*

$$\lim_{n\to\infty}\sigma^2_{\hat{m}_n}(t) = 1 - \sigma^2_e(\mu). \quad (17)$$

Proof: Define a new vector, $\mathbf{x}'_n = [x(d+\mu d), x(2d+\mu d), \cdots, x(nd+\mu d)]^T$, which is obtained by shifting $\mathbf{x}_n$ to the right by μd seconds. The posterior covariance matrix of $\mathbf{x}'_n$ given $\mathbf{y}_n$ is $\hat{\mathbf{R}}_{\mathbf{x}'_n\mathbf{x}'_n} := \mathbb{E}[\mathbf{x}'_n\mathbf{x}'^T_n|\mathbf{y}_n]$, which can be expressed as

$$\hat{\mathbf{R}}_{\mathbf{x}'_n\mathbf{x}'_n} = \mathbf{R}_{\mathbf{x}'_n\mathbf{x}'_n} - E\,\mathbf{R}_{\mathbf{x}'_n\mathbf{x}_n}\left(E\,\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n} + \sigma^2\mathbf{I}_n\right)^{-1}\mathbf{R}^T_{\mathbf{x}'_n\mathbf{x}_n} \quad (18)$$

where $\mathbf{R}_{\mathbf{x}'_n\mathbf{x}'_n} = \mathbb{E}[\mathbf{x}'_n\mathbf{x}'^T_n] = \mathbf{R}_{\mathbf{x}_n\mathbf{x}_n}$ and $\mathbf{R}_{\mathbf{x}'_n\mathbf{x}_n} = \mathbb{E}[\mathbf{x}'_n\mathbf{x}^T_n]$. Since the value of $\hat{k}_n(t)$ in (14) is on the diagonal of $\hat{\mathbf{R}}_{\mathbf{x}'_n\mathbf{x}'_n}$, we have

$$\sigma^2_e(\mu) = \lim_{n\to\infty}\hat{k}_n(t) = \lim_{n\to\infty}\frac{1}{n}\text{trace}(\hat{\mathbf{R}}_{\mathbf{x}'_n\mathbf{x}'_n}) \quad (19)$$

From (1), $\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n}$ is a symmetric Toeplitz matrix with the (i,j)-th element being $\rho^{|i-j|d}$. Similarly, the matrix $\mathbf{R}_{\mathbf{x}'_n\mathbf{x}_n}$ is an asymmetric Toeplitiz matrix with the first row being $[\rho^{|\mu d|}, \rho^{|\mu-1|d}, \cdots, \rho^{|\mu-(n-1)|d}]$, and the first column $[\rho^{|\mu d|}, \rho^{|\mu+1|d}, \cdots, \rho^{\mu+(n-1)|d}]^T$.

The Toeplitz matrix, $\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n}$ is uniquely determined by the sequence $\{\rho^{|n|d}\}_n$, whose discrete-time Fourier transform (DTFT), $\Psi_{\mathbf{x}_n\mathbf{x}_n}(\omega) = \sum_{n=-\infty}^{+\infty}\rho^{|n|d}e^{-jn\omega}$, is

$$\Psi_{\mathbf{x}_n\mathbf{x}_n}(\omega) = \frac{1-\rho^{2d}}{1+\rho^{2d}-2\rho^d\cos(\omega)} \quad (20)$$

The Toeplitz matrix, $\mathbf{R}_{\mathbf{x}'_n\mathbf{x}_n}$, is uniquely determined by the sequence, $\{\rho^{|\mu+n|d}\}_n$. The DTFT, $\Psi_{\mathbf{x}'_n\mathbf{x}_n}(\omega) = \sum_{n=-\infty}^{+\infty}\rho^{|\mu+n|d}e^{-jn\omega}$, can be calculated as

$$\Psi_{\mathbf{x}'_n\mathbf{x}_n}(\omega) = \frac{\rho^{-\mu d}\left[\rho^d(1-\rho^{2\mu d})e^{j\omega} + \rho^{2\mu d} - \rho^{2d}\right]}{1+\rho^{2d}-2\rho^d\cos(\omega)}. \quad (21)$$

Based on [18, Lemma 2], $\mathbf{R}_{\mathbf{x}'_n\mathbf{x}_n}$ is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{\mathbf{x}'_n\mathbf{x}_n} = \mathbf{U}_n^H\mathbf{D}_{\mathbf{x}'_n\mathbf{x}_n}\mathbf{U}_n$, where $\mathbf{U}_n$ is the unitary discrete Fourier transform (DFT) matrix with the (i,l)-th element being $(\mathbf{U}_n)_{i,l} = \frac{1}{\sqrt{n}}\exp\left[-j2\pi\frac{(i-1)(l-1)}{n}\right]$, and $\mathbf{D}_{\mathbf{x}'_n\mathbf{x}_n}$ is a diagonal matrix with its k-th element being $(\mathbf{D}_{\mathbf{x}'_n\mathbf{x}_n})_k = \Psi_{\mathbf{x}'_n\mathbf{x}_n}(2\pi\frac{k-1}{n})$.

Similarly, the Toeplitz matrix, $\mathbf{R}_{\mathbf{x}_n\mathbf{x}_n}$, is asymptotically equivalent to a circulant matrix, $\mathbf{C}_{\mathbf{x}_n\mathbf{x}_n} = \mathbf{U}_n^H\mathbf{D}_{\mathbf{x}_n\mathbf{x}_n}\mathbf{U}_n$, where $\mathbf{D}_{\mathbf{x}_n\mathbf{x}_n}$ is a diagonal matrix with its k-th element being $(\mathbf{D}_{\mathbf{x}_n\mathbf{x}_n})_k = \Psi_{\mathbf{x}_n\mathbf{x}_n}(2\pi\frac{k-1}{n})$.

Based on [19, Theorem 2.1], $\hat{\mathbf{R}}_{\mathbf{x}'_n\mathbf{x}'_n}$ in (18) is asymptotically equivalent to a circulant matrix $\hat{\mathbf{C}}_{\mathbf{x}'_n\mathbf{x}'_n} = \mathbf{C}_{\mathbf{x}_n\mathbf{x}_n} - \mathbf{C}_{\mathbf{x}'_n\mathbf{x}_n}(\mathbf{C}_{\mathbf{x}_n\mathbf{x}_n} + \frac{1}{\gamma_0(d-d_c)})^{-1}\mathbf{C}^H_{\mathbf{x}'_n\mathbf{x}_n} = \mathbf{U}_n^H\hat{\mathbf{D}}_{\mathbf{x}_n\mathbf{x}_n}\mathbf{U}_n$, where $\hat{\mathbf{D}}_{\mathbf{x}_n\mathbf{x}_n} = \mathbf{D}_{\mathbf{x}_n\mathbf{x}_n} - \mathbf{D}_{\mathbf{x}'_n\mathbf{x}_n}(\mathbf{D}_{\mathbf{x}_n\mathbf{x}_n} + \frac{1}{\gamma_0(d-d_c)})^{-1}\mathbf{D}^H_{\mathbf{x}'_n\mathbf{x}_n}$.

From Szego's Theorem [18], we have

$$\lim_{n\to\infty}\hat{k}_n(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\Psi_{\mathbf{x}_n\mathbf{x}_n}(\omega) - \frac{|\Psi_{\mathbf{x}'_n\mathbf{x}_n}(\omega)|^2}{\Psi_{\mathbf{x}_n\mathbf{x}_n}(\omega) + \frac{1}{\gamma_0(d-d_c)}}\right]d\omega. \tag{22}$$

Substituting (20) and (21) into above equation and solving the integral with [20, eqn. (2553.3)], we have the result in (16). ■

In Theorem 2, the asymptotic posteriori variance $\sigma_e^2(\mu)$ is expressed as an explicit function of a number of parameters, such as the SNR γ_0, the sampling period d, the normalized hardware energy d_c, the temporal correlation coefficient ρ, and the relative time location μ.

From Theorem 2, we have the following results regarding the asymptotic distribution of $\hat{m}_n(t)$.

From Theorem 2, as $n\to\infty$, $\hat{m}_n(kd+\mu)$ is a zero-mean Gaussian random variable with variance $1-\sigma_e^2(\mu)$. Therefore, the statistical properties of $\hat{m}_n(t)$ as $n\to\infty$ are periodic in t with period d. Define $\hat{m}(\mu) := \lim_{n\to\infty}\hat{m}_n(kd+\mu)$. We have the following corollary regarding the distribution of $\hat{m}(\mu)$.

Corollary 2 *As $n\to\infty$, $\hat{m}(\mu) = \lim_{n\to\infty}\hat{m}_n(kd+\mu)$ is zero-mean Gaussian distributed with variance $1-\sigma_e^2(\mu)$, that is, $\hat{m}(\mu)\sim\mathcal{N}(0, 1-\sigma_e^2(\mu))$.*

With the asymptotic results in Theorem 2 and Corollary 2, we can get the explicit expression of the cost function in (5), and the result is given in the following theorem.

Theorem 3 *The cost function in (5), $\mathcal{J} := \lim_{n\to\infty}\frac{1}{nd}\mathbb{E}[e(\hat{\mathcal{S}}(nd))]$, can be expressed as*

$$\mathcal{J} = \frac{1}{\pi}\int_0^1\int_0^{\pi/2}\left[1+\frac{1-\sigma_e^2(\mu)}{\sigma_e^2(\mu)\sin^2(\theta)}\right]^{-1/2} \times\exp\left(-\frac{\gamma^2/2}{1-\sigma_e^2(\mu)\cos^2(\theta)}\right)d\mu d\theta. \tag{23}$$

Proof: We first consider the conditional cost function, $\lim_{n\to\infty}\frac{1}{nd}\mathbb{E}[e(\hat{\mathcal{S}}(nd))|\mathbf{y}_n]$. From (11)-(12), the cost function can be expressed as

$$\lim_{n\to\infty}\frac{1}{nd}\mathbb{E}[e(\hat{\mathcal{S}}(nd))|\mathbf{y}_n] = \lim_{n\to\infty}\frac{1}{nd}\sum_{i=1}^{n}\int_{(i-1)d}^{id}Q\left(\frac{|\gamma-\hat{m}_n(t)|}{\sqrt{\hat{k}_n(t)}}\right)dt$$

Performing change of variable, $t=(i-1)d+\mu d$, in the above integral yields, and using the results from Theorem 2, we have

$$\lim_{n\to\infty}\frac{1}{nd}\mathbb{E}[e(\hat{\mathcal{S}}(nd))|\mathbf{y}_n] = \int_0^1 Q\left(\frac{|\gamma-\hat{m}(\mu)|}{\sqrt{\sigma_e^2(\mu)}}\right)d\mu \tag{24}$$

In the conditional cost function in (24), there is only one random variable, $\hat{m}(\mu)\in\mathcal{N}(0,1-\sigma_e^2(\mu))$, which is a function of $\mathbf{y}_n$. Therefore the unconditional cost function can be expressed as

$$\mathcal{J} = \lim_{n\to\infty}\frac{1}{nd}\mathbb{E}[e(\hat{\mathcal{S}}(nd))] = \int_0^1\mathbb{E}_{\hat{m}(\mu)}\left[Q\left(\frac{|\gamma-\hat{m}(\mu)|}{\sqrt{\sigma_e^2(\mu)}}\right)\right]d\mu \tag{25}$$

With Craig's alternative expression of the Q-function [21], (25) can be reformulated as

$$\mathcal{J} = \frac{1}{\pi}\int_0^1\int_0^{\frac{\pi}{2}}\mathbb{E}_{\hat{m}(\mu)}\left[e^{-\frac{(\hat{m}(\mu)-\gamma)^2}{2\sigma_e^2(\mu)\sin^2(\theta)}}\right]d\mu d\theta \tag{26}$$

Define $Z := \frac{(\hat{m}(\mu)-\gamma)^2}{1-\sigma_e^2(\mu)}$, which is a non-central χ^2-distributed random variable with one degree-of-freedom and the non-centrality parameter $\frac{\gamma^2}{1-\sigma_e^2(\mu)}$. The moment generating function (MGF) of Z, $M_Z(s) = \mathbb{E}_Z[e^{sZ}]$, is

$$M_Z(s) = \exp\left[\frac{s}{(1-2s)}\frac{\gamma^2}{1-\sigma_e^2(\mu)}\right]\frac{1}{\sqrt{1-2s}}. \tag{27}$$

Combining (26) with (27) yields (23). ■

The results in Theorem 3 give the exact analytical expression of the cost function, which is expressed as a function of the optimization parameter d, and other system parameters such as the SNR γ_0, the temporal correlation coefficient ρ, and the normalized hardware energy consumption d_c. Thus, given $\{\gamma_0,\rho,d_c\}$, we can identify $d\geq d_c$ that minimizes the cost function $\mathcal{J}$ in (23). The integrand in (23) has only elementary functions, and the integration limits are finite. Thus the integrals in (23) can be easily evaluated with numerical integrations with high precision.

B. Optimum Sampling Rate

Due to the complicated form of the double integrals in (23), it might be difficult to directly minimize the exact cost function with respect to d. We resort to an upper bound of d to simplify the optimization.

The following corollary provides an upper bound of the cost function expressed in a closed-form.

Corollary 3 *The asymptotical expected LSE error in (23) is upper bounded by*

$$g(d) = \frac{1}{2}\exp\left(-\frac{\gamma^2}{2}\right)\left[\frac{1}{\gamma_0(d-d_c)}+\frac{1+\rho^{2d}}{1-\rho^{2d}}+\frac{1}{d\log\rho}\right]^{1/2} \left(\frac{1}{\gamma_0(d-d_c)}+\frac{1-\rho^d}{1+\rho^d}\right)^{-\frac{1}{4}}\left(\frac{1}{\gamma_0(d-d_c)}+\frac{1+\rho^d}{1-\rho^d}\right)^{-\frac{1}{4}} \tag{28}$$

Proof: It can be easily shown that the integrand in (23) is an increasing function with respect to θ for $\theta\in[0,\pi/2]$. Thus we have

$$\mathbb{E}[e(\hat{\mathcal{S}})] \leq \frac{1}{2}\exp\left(-\frac{\gamma^2}{2}\right)\int_0^1\sqrt{\sigma_e^2(\mu)}d\mu. \tag{29}$$

Due to the concavity of $\sqrt{x}$, the asymptotic LSE error can be upper-bounded again as

$$\mathbb{E}[e(\hat{\mathcal{S}})] \leq \frac{1}{2}\exp\left(-\frac{\gamma^2}{2}\right)\sqrt{\int_0^1 \sigma_e^2(\mu)d\mu}. \qquad (30)$$

Substituting (16) into (30) and simplifying yields (28). ■

Given the complicated form of the exact LSE error probability in (23), we propose to instead minimize the error probability upper bound in (28). It will be shown in the numerical results that minimizing the exact error probability or its upper bound yields almost the same values of the optimum sampling rates. Since the upper bound in (28) is continuous and differentiable with respect to d, the optimum value of d that minimizes $g(d)$ must be one of the zero-slope points of $g(d)$. Therefore, the optimum value of d must be one of the solutions to $g'(d) = 0$, which can be expressed as $g'(d) = c(d) \cdot w(d)$, where

$$\begin{aligned} w(d) = &-8\ln\rho^{-1}\rho^{2d}(1-\rho^{2d})^{-3} + \left[-4\ln\rho^{-1}\gamma_0(d-d_c)\rho^{2d}\right. \\ &\left. +4\ln\rho^{-1}\rho^{2d} + \frac{4}{d}\rho^{2d} + \frac{4}{d-d_c}\right](1-\rho^{2d})^{-2} \\ &+\left[-\frac{2}{\ln\rho^{-1}d(d-d_c)} - \frac{4}{d-d_c} + \frac{4}{\ln\rho^{-1}d^2}\right](1-\rho^{2d})^{-1} \\ &+\frac{1}{\ln\rho^{-1}d(d-d_c)}\left(1-\frac{1}{\gamma_0(d-d_c)}\right) \\ &+\frac{1}{d^2\ln\rho^{-1}}\left(\frac{1}{\gamma_0(d-d_c)} + \gamma_0(d-d_c) - 2\right) \qquad (31) \end{aligned}$$

and

$$\begin{aligned} c(d) = &\frac{1}{8}\exp(-\gamma^2)g^{-1}(d)\frac{1}{\gamma_0(d-d_c)}\left[\left(\frac{1}{\gamma_0(d-d_c)} - 1\right)^2\right. \\ &\left. + \frac{4}{\gamma_0(d-d_c)(1-\rho^{2d})}\right]^{-\frac{3}{2}} > 0 \end{aligned}$$

for all $d > d_c$.

The above problem can be solved numerically with the fsolve function in Matlab. Our numerical results indicate the LSE error probability upper bound, $g(d)$, defined in (28) is quasi-convex in d and there is always just a unique solution to the above equation for all configurations considered in this paper.

V. Numerical and Simulation Results

In this section, numerical and simulation results are presented to demonstrate the performance of the proposed level set sensing and estimation algorithm. Without loss of generality, the level set threshold γ is set to 0.1.

Fig. 1 shows the asymptotic LSE error probabilities and their corresponding upper bounds as a function of the sampling rate $r = \frac{1}{d} \leq \frac{1}{d_c}$ under various temporal correlation parameter ρ. The SNR is $\gamma_0 = 10$ dB, and the normalized hardware energy consumption d_c is fixed as 0.05, which corresponds to a maximum sampling rate of $r_{\max} = 20$ Hz. The simulation results are obtained with $n = 100$ samples yet the analytical results are derived by using $n \to \infty$. The simulation results with finite n match very well with the analytical results with infinite n, thus the asymptotic analytical results provide a very good approximation of the performance of practical systems with a finite n. For all system configurations, when the sampling rate approaches its boundaries at 0 or $r_{\max} = 20$ Hz, the LSE error probability approaches $Q(|\gamma|)$, the error probability of random decisions. At 0 Hz, no sample is collected by the sensor. At $r_{\max}$, all energy is consumed by the sensing operation thus no information is transmitted to the FC. The optimum sampling rates that minimize the error probability upper bound in (28) are obtained by zeroing (31) and marked in the figure. The optimum sampling rates that minimize the exact error probability in (23) are obtained through exhaustive search. It is observed that the sampling rates that minimize the upper bound or the exact expression are almost the same. For example, when $\rho = 0.9$, the two optimum sampling rates are 1.85 Hz and 1.81 Hz, respectively. Therefore minimizing the upper bound provides a reasonably accurate approximation of the true optimum sampling rate.

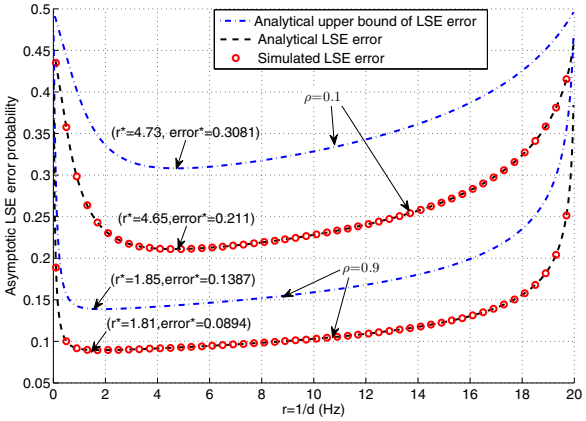

Fig. 1. LSE error probabilities as a function of sampling rate $r = 1/d$ under various temporal correlation coefficients ρ ($\gamma_0 = 10$ dB, $d_c = 0.05$ seconds).

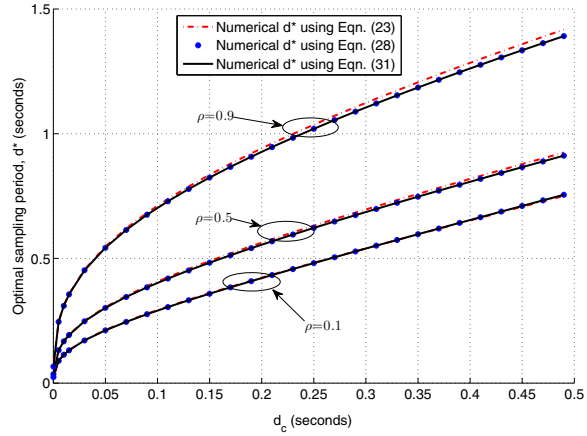

Fig. 2. Optimum sampling period d^* as a function of d_c under various temporal correlation coefficients ρ ($\gamma_0 = 10$ dB).

Fig. 2 shows the optimum sampling periods d^* as a function of d_c under various values of ρ. The SNR is set as $\gamma_0 = 10$ dB. The results that minimize the upper bound in (28) or the exact error probability in (23) are shown in the

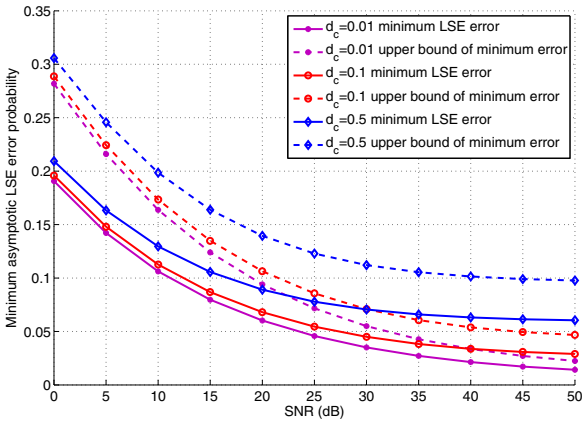

Fig. 3. Minimum LSE error probabilities as a function of SNR (γ_0) under various d_c ($\rho = 0.8$).

figure. Again, minimizing the upper bound or the exact error probability yields almost identical optimum sampling periods, for all system configurations. The optimum sampling period is an increasing function in ρ, in that a larger ρ renders a stronger correlation between two adjacent samples. It is also an increasing function in d_c, because more energy needs to be allocated for each sample with a higher hardware energy consumption.

The minimum LSE error probabilities and their upper bounds are shown in Fig. 3 as a function of the SNR, under various values of d_c. The results are obtained by first identifying the optimum sampling period d^* obtained by zeroing (31), and then plugging the values in (23) or (28). The value of ρ is 0.8. As expected, a higher d_c or a higher hardware energy consumption always results in a higher LSE error probability. The impact of d_c on the LSE is very small when the SNR is low, and it is more pronounced at high SNR. In addition, the gap between the minimum LSE error probability and its upper bound narrows as SNR increases.

VI. Conclusions

The optimum level set estimation of a time-varying random field under a power constraint has been studied in this paper. The optimum LSE algorithm has been developed by thresholding the Gaussian regressions of the noisy samples at the FC. With the optimum LSE algorithm, the exact LSE error probability and its upper bound has been expressed as explicit functions of a number of system parameters, such as the sampling rate, the SNR, the hardware energy consumption, and the temporal correlation coefficient of the random field. The optimum sampling rate that minimizes the upper bound of LSE error probability has been identified. Simulation results demonstrated that minimizing the exact error probability or its upper bound result in almost the same optimum sampling rates.

References

[1] I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Phys. Commun.*, vol. 4, no. 1, pp. 40–62, Mar. 2011.

[2] N. Maisonneuve, M. Stevens, and B. Ochab, "Participatory noise pollution monitoring using mobile phones," *Info. Pol.*, vol. 15, no. 1,2, pp. 51–71, Apr. 2010.

[3] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, ser. WSNA '02. New York, NY, USA: ACM, 2002, pp. 88–97.

[4] D. M. Mason, W. Polonik *et al.*, "Asymptotic normality of plug-in level set estimates," *The Annals of Applied Probability*, vol. 19, no. 3, pp. 1108–1142, 2009.

[5] P. Hall and K.-H. Kang, "Bandwidth choice for nonparametric classification," *Annals of statistics*, pp. 284–306, 2005.

[6] A. Goldenshluger and A. Zeevi, "The hough transform estimator," *Annals of statistics*, pp. 1908–1932, 2004.

[7] C. Scott and M. Davenport, "Regression level set estimation via cost-sensitive classification," *Signal Processing, IEEE Transactions on*, vol. 55, no. 6, pp. 2752–2757, 2007.

[8] K. Dantu and G. Sukhatme, "Detecting and tracking level sets of scalar fields using a robotic sensor network," in *Robotics and Automation, 2007 IEEE International Conference on*. IEEE, 2007, pp. 3665–3672.

[9] R. Willett and R. D. Nowak, "Minimax optimal level-set estimation," *Image Processing, IEEE Transactions on*, vol. 16, no. 12, pp. 2965–2979, 2007.

[10] S. Srinivasan, K. Ramamritham, and P. Kulkarni, "Ace in the hole: Adaptive contour estimation using collaborating mobile sensors," in *Information Processing in Sensor Networks, 2008. IPSN'08. International Conference on*. IEEE, 2008, pp. 147–158.

[11] A. Gotovos, N. Casati, G. Hitz, and A. Krause, "Active learning for level set estimation," in *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*. AAAI Press, 2013, pp. 1344–1350.

[12] J. Yang, Z. Wang, and J. Wu, "Level set estimation with dynamic sparse sensing," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*. IEEE, 2014, pp. 487–491.

[13] B. Bryan, J. Schneider, R. Nichol, C. Miller, C. Genovese, and L. Wasserman, "Active learning for identifying function threshold boundaries," in *NIPS*. Citeseer, 2005.

[14] B. Settles, "Active learning literature survey," *University of Wisconsin, Madison*, vol. 52, no. 55-66, p. 11, 2010.

[15] J. Wu and N. Sun, "Optimum sensor density in distortion tolerant wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 2056–2064, 2012.

[16] N. Sun and J. Wu, "Optimum sampling in spatial-temporally correlated wireless sensor networks," *EURASIP J. Wireless Commun. Networking*, 2013.

[17] C. E. Rasmussen, "Gaussian processes for machine learning," 2006.

[18] H. Gazzah, P. A. Regalia, and J.-P. Delmas, "Asymptotic eigenvalue distribution of block toeplitz matrices and application to blind simo channel identification," *Information Theory, IEEE Transactions on*, vol. 47, no. 3, pp. 1243–1251, 2001.

[19] R. M. Gray, "Toeplitz and circulant matrices: A review," *Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2005.

[20] D. Zwillinger, *Table of integrals, series, and products*. Elsevier, 2014.

[21] J. W. Craig, "A new, simple and exact result for calculating the probability of error for two-dimensional signal constellations," in *Military Communications Conference, 1991. MILCOM'91, Conference Record, Military Communications in a Changing World., IEEE*. IEEE, 1991, pp. 571–575.