Optimal Energy Efficient Level Set Estimation of Spatially-Temporally Correlated Random Fields

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 classifying the region(s) that the values of an unknown function exceed a certain threshold. It has a wide range of applications such as spectrum sensing or environment monitoring. In this paper, we study the the optimal LSE of a linear random field that changes with respect to time. A linear sensor network is used to take discrete samples of the spatially-temporally correlated random field in both the space and time domain, and the sensors operate under a total power constraint. The samples are congregated at a fusion center (FC), which performs LSE of the random field by using the noisy observation of the samples. Under the Gaussian process (GP) framework, we first develop an optimal LSE algorithm that can minimize the LSE error probability. The results are then used to derive the exact LSE error probability with the assistance of frequency domain analysis. The analytical LSE error probability is expressed as an explicit function of a number of system parameters, such as the distance between two adjacent nodes, the sampling period in the time domain, the signal-to-noise ratio (SNR), and the spatial-temporal correlation of the random field. With the analytical results, we can identify the optimum node distance and sampling period that can minimize the LSE error probability.**

I. Introduction

Wireless sensor networks (WSNs) have been widely used in many scientific and engineering applications, including search and rescue, disaster relief, spectrum sensing, and environment monitoring, etc. Many WSNs are designed to monitor a random event, which can be modeled as a random function in the space and time domains. It is usually difficult and costly to estimate the precise values of the random function. For many applications, it is sufficient to find out the regions over which the function values exceeds a certain threshold, and this is denoted as level set estimation (LSE). The applications of LSE include terra in mapping, spectrum sensing [1], and monitoring the contours of sunlight, water pollution, and rainfall [2].

A large number of works are devoted to the development of LSE algorithms [3]– [4]. In [5], the LSE is performed by identifying the difference between two probability densities, and the method is closely related to standard binary classifications. The binary classification based method do not consider the difference between the threshold and the actual function value, which contains salient information helpful to LSE. The methods in [6]–[8] are developed by studying the statistical properties of the random field. These methods do not involve an intermediate reconstruction step and it is usually hard to obtain analytical conclusions on consistency and convergence. A popular LSE method is to estimate the values of the underlying function and then thresholding at the critical value [3], [9], [10]. Such an approach is easy to implement and the consistency and convergence of the algorithms can be analyzed based on certain smooth prior assumptions. In [11]–[13], the domains defining the function of interest are discretized into a set of small regions to perform the LSE, and the price is the larger estimation error introduced by the discretization. The LSE of a single spatial point over infinite continuous time domain are developed in [4]. However, it does not consider the spatial variation of the random field.

In this paper, we study the LSE of a spatially-temporally correlated random field with a linear WSN. The sensor nodes are evenly distributed on a line in the spatial domain, and they periodically sample a time-varying physical quantity, such as temperature or pollution level, under a constraint on the total power per unit area. The collected samples are transmitted to a fusion center (FC), which performs the LSE by using noisy observations of the samples. The performance of the LSE depends critically on the number of sensors in a unit area, i.e. sensor density, and the sampling rate. A higher sensor density and/or sampling rate means denser sampling of the random event, which benefit the LSE estimation. On the other hand, under the constraint of a total power per unit area, a higher sensor density and/or sampling rate means less energy per sample or a lower signal-to-noise ratio (SNR), which negatively affects the LSE performance. Therefore it is important to identify the optimum sensor density in the space domain and sampling rate in the time domain.

Under GP framework , we first propose an optimum LSE algorithm that can minimize the LSE error probability. The results are then used to derive an asymptotic LSE error probability when the size of the field and the time go to infinity. The analytical LSE error probability are explicitly expressed as a function of various system parameters, such as the node distance (inverse of node density), sampling period, SNR, and spatial and temporal covariance kernels. The asymptotically optimum node density and sampling rate can then be obtained by minimizing the asymptotic LSE error probability. Simulation results show that the asymptotic results can accurately predict the performance of practical random field of finite size and finite time duration.

The work was supported in part by the U.S. National Science Foundation (NSF) under Grants ECCS-1202075, ECCS-1405403, and 1454471.

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

II. System Model

Consider a linear sensor network with M sensor nodes evenly distributed on a line. Denote the coordinate of the m-th node as $s_m = ml$, for $m = 1, 2, \ldots, M$, where l is the distance between two adjacent sensors. Define the two-dimensional space-time coordinate vector as $\mathbf{c} = [s, t]^T \in \mathcal{X} \times \mathcal{R}_+$, where s is the space coordinate, t is the time variable, $\mathcal{X}$ is the linear field and $\mathcal{R}$ is the set of real numbers. Sensors are used to measure a spatial-temporally dependent physical quantity, $x(\mathbf{c})$, such as air pressure, temperature, aggregated power level of wireless signals, or density of toxic gases, etc.

It is assumed that the *prior* distribution of $\{x(\mathbf{c})\}$ is a zero-mean Gaussian process that is wide sense stationary (WSS) in both space and time. The covariance function of $\{x(\mathbf{c})\}$ is

$$k(\mathbf{c}, \mathbf{c}') = \mathbb{E}[x(\mathbf{c})x(\mathbf{c}')] = k_s(|s - s'|) \cdot k_t(|t - t'|), \quad (1)$$

where $k_s(\cdot), k_t(\cdot)$ are the spatial and temporal covariance functions, respectively, and both are absolutely integrable.

Due to energy limit, the sensors take discrete-time samples of the random field. The collected discrete-time samples are transmitted to a FC. Assume uniform sampling is used. Denote the sampling instants as $t_n = nd$, for $n = 1, 2, \cdots$, where d is the sampling period. It is assumed that an energy $E_0 = E_c + E$ is allocated for each sample, where the constant E_c is due to hardware power consumption of the sensing operation, and E is the transmission energy of a sample. Denote $\mathbf{c}_{in} := [s_i, t_n]^T$ as the space-time coordinate of the n-th sample from the i-th sensor. The samples observed at the FC can be represented as

$$y(\mathbf{c}_{in}) = \sqrt{E}x(\mathbf{c}_{in}) + \xi(\mathbf{c}_{in}) \quad (2)$$

where $\xi(\mathbf{c}_{in})$ includes the effects of observation noise and channel distortions. It is assumed that $\xi(\mathbf{c}_{in})$ is a white Gaussian process with zero-mean and variance σ^2. The sensor nodes operate under the constraint of a fixed power P_0 per unit area. Given a sensing system with node density $\delta = \frac{1}{l}$, the energy allocated to one sample is thus $E_0 = \frac{P_0 d}{\delta} = P_0 dl$.

The FC uses the discrete-time samples to estimate the γ-level set of $\{x(\mathbf{c})\}$, which is defined as

$$\mathcal{S}(t) := \{[s, t]^T : s \in \mathcal{X}, x(s, t) > \gamma\}. \quad (3)$$

Without loss of generality, we assume $\gamma > 0$.

Denote the estimated level set as $\hat{\mathcal{S}}(t)$. Then the level set estimation error up to time t is defined as

$$e(t) := \int_0^t \int_{\mathcal{X}} \mathbb{I}\left\{[s, \mu]^T \in \Delta(\mathcal{S}(\mu), \hat{\mathcal{S}}(\mu))\right\} ds\, d\mu \quad (4)$$

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 between two sets, $\mathcal{S}^c$ is the complement of $\mathcal{S}$, and $\mathbb{I}\{\mathcal{E}\} = 1$ if the event $\mathcal{E}$ is true and 0 otherwise.

Then, the LSE problem can be formulated as

$$\begin{aligned} \min. \quad & \lim_{t, M \to \infty} \frac{\delta}{Mt} \mathbb{E}[e(t)] \\ \text{s.t.} \quad & E = P_0 d/\delta - E_c \geq 0 \end{aligned} \quad (5)$$

In the cost function, the LSE error is normalized by the time duration t and the length of the area of interest $V = M/\delta$. The optimization is performed with respect to the node density δ and the sampling rate $r = 1/d$.

III. Optimal Level Set Estimation in GP

We first study the optimal LSE for given δ and d and derive the corresponding LSE error probability. The results will be used to identify the solutions to (5) in the next section.

Denote $\mathbf{x}_n$ and $\mathbf{y}_n \in \mathcal{R}^{M \times 1}$ be the vectors containing the true and observed data samples that the FC collects from all M sensors at time t_n, respectively. Define $\mathbf{x}_{1:n} := \{\mathbf{x}_i\}_{i=1}^n$ and $\mathbf{y}_{1:n} := \{\mathbf{y}_i\}_{i=1}^n$ be the sets of true and observed discrete-time data from time t_1 to t_n, respectively.

Since $\{x(\mathbf{c})\}$ is Gaussian process, given $\mathbf{y}_{1:n}$, the distribution of $x(\mathbf{c})$ is still Gaussian, with conditional mean, $\hat{m}_n(\mathbf{c})$, and conditional variance, $\hat{k}_n(\mathbf{c})$, given by

$$\hat{m}_n(\mathbf{c}) = \mathbf{r}_{x\mathbf{y}_{1:n}}(\mathbf{c}) \mathbf{R}^{-1}_{\mathbf{y}_{1:n}\mathbf{y}_{1:n}} \mathbf{y}_{1:n}, \quad (6)$$

$$\hat{k}_n(\mathbf{c}) = k(\mathbf{c}, \mathbf{c}) - \mathbf{r}_{x\mathbf{y}_{1:n}}(\mathbf{c}) \mathbf{R}^{-1}_{\mathbf{y}_{1:n}\mathbf{y}_{1:n}} \mathbf{r}_{x\mathbf{y}_{1:n}}(\mathbf{c})^T, \quad (7)$$

where $\mathbf{r}_{x\mathbf{y}_{1:n}}(\mathbf{c}) := \mathbb{E}[x(\mathbf{c})\mathbf{y}_{1:n}^T] \in \mathcal{R}^{nM \times 1}$ and $\mathbf{R}_{\mathbf{y}_{1:n}\mathbf{y}_{1:n}^T} := \mathbb{E}[\mathbf{y}_{1:n}\mathbf{y}_{1:n}^T] \in \mathcal{R}^{nM \times nM}$.

Based on (6), the GP regression based LSE algorithm is given in Algorithm 1.

Algorithm 1 GP regression based level set estimation

1: Input: $\mathbf{t}_n$ and $\mathbf{y}_{1:n}$
2: Run GP regression for $\forall t \in [0, t_n], s \in \mathcal{X}$ to get $\hat{m}_n(\mathbf{c})$ with (6).
3: Threshold $\hat{m}_n(\mathbf{c})$:

$$\hat{\mathcal{S}}(t) = \{[s, t]^T : s \in \mathcal{X}, \hat{m}_n(\mathbf{c}) > \gamma\}, \quad 0 \leq t \leq t_n$$

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

Theorem 1 *Algorithm 1 is optimal with given $\mathbf{y}_{1:n}$, i.e., it minimizes the conditional LSE error probability, $\mathbb{E}[e(t_n)|\mathbf{y}_{1:n}]$. The corresponding minimum LSE error probability is*

$$\mathbb{E}[e(t_n)|\mathbf{y}_{1:n}] = \int_0^{t_n} \int_{\mathcal{X}} Q\left(\frac{|\gamma - \hat{m}_n(\mathbf{c})|}{\sqrt{\hat{k}_n(\mathbf{c})}}\right) ds\, dt \quad (8)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du$ is the Gaussian-Q function.

Proof:

Given $\mathbf{y}_{1:n}$, the expected LSE error can be calculated as

$$\begin{aligned} &\mathbb{E}[e(t_n)|\mathbf{y}_{1:n}] \\ &= \int_0^{t_n} \int_{\mathcal{X}} \mathbb{P}\left[[s, t]^T \in \Delta(\mathcal{S}(t), \hat{\mathcal{S}}(t))\middle| \mathbf{y}_{1:n}\right] ds\, dt \\ &= \int_0^{t_n} \int_{\mathcal{X}} \Big(\mathbb{P}\left[x(\mathbf{c}) \leq \gamma | \mathbf{y}_{1:n}\right] \cdot \mathbb{I}\{[s, t]^T \in \hat{\mathcal{S}}(t)\} \\ &\quad + \mathbb{P}\left[x(\mathbf{c}) > \gamma | \mathbf{y}_{1:n}\right] \cdot \mathbb{I}\{[s, t]^T \in \hat{\mathcal{S}}^c(t)\}\Big) ds\, dt \end{aligned} \quad (9)$$

Therefore, the optimal estimator that minimizes (9) is to let

$$\mathbf{c} \in \begin{cases} \hat{\mathcal{S}}(t), & \text{if } \mathbb{P}\left[x(\mathbf{c}) > \gamma | \mathbf{y}_{1:n}\right] > \mathbb{P}\left[x(\mathbf{c}) \leq \gamma | \mathbf{y}_{1:n}\right] \\ \hat{\mathcal{S}}^c(t), & \text{if } \mathbb{P}\left[x(\mathbf{c}) > \gamma | \mathbf{y}_{1:n}\right] \leq \mathbb{P}\left[x(\mathbf{c}) \leq \gamma | \mathbf{y}_{1:n}\right] \end{cases} \tag{10}$$

for every $\mathbf{c} = [s,t]^T$ with $t \in [0, t_n]$ and $s \in \mathcal{X}$.

Since $x(\mathbf{c})$ given $\mathbf{y}_{1:n}$ is still Gaussian distributed with mean and variance given in (6) and (7), we have

$$\mathbb{P}\left[x(\mathbf{c}) > \gamma | \mathbf{y}_{1:n}\right] = Q\left(\frac{\gamma - \hat{m}_n(\mathbf{c})}{\sqrt{\hat{k}_n(\mathbf{c})}}\right) \tag{11}$$

The optimal estimator defined in (10) is then reduced to compare $\hat{m}_n(\mathbf{c})$ with γ. If $\hat{m}_n(\mathbf{c}) > \gamma$, the probability in (11) is greater than 1/2, thus, we should let $\mathbf{c} \in \hat{\mathcal{S}}(t_n)$; otherwise, we let $\mathbf{c} \in \hat{\mathcal{S}}^c(t_n)$. A combination of (9) and (11) results in (8). ■

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 period and node density that can minimize the LSE error probability.

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

$$\mathbb{E}[e(t_n)] = \mathbb{E}_{\mathbf{y}_{1:n}}\left\{\mathbb{E}[e(t_n)|\mathbf{y}_{1:n}]\right\} \tag{12}$$

From (8) and (12), the cost function depends on the posterior mean $\hat{m}_n(\mathbf{c})$ and variance $\hat{k}_n(\mathbf{c})$. From (6) and (7), we have

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

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

where $\mathbf{r}_{x\mathbf{x}_{1:n}}(\mathbf{c}) = \mathbb{E}\left[x(\mathbf{c})\mathbf{x}^T_{1:n}\right] \in \mathcal{R}^{1\times nM}$ and $\mathbf{R}_{\mathbf{x}_{1:n}\mathbf{x}_{1:n}} = \mathbb{E}[\mathbf{x}_{1:n}\mathbf{x}^T_{1:n}] \in \mathcal{R}^{nM\times nM}$.

The posterior mean is a function of $\mathbf{y}_{1:n}$, whereas the posterior variance is a constant independent of $\mathbf{y}_{1:n}$. Since $\mathbf{y}_{1:n}$ is zero-mean Gaussian distributed, it can be easily shown that $\hat{m}_n(\mathbf{c})$ is zero-mean Gaussian distributed with variance

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

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

The variance in (15) depends on a number of factors, such as the spatial and temporal covariance, the sampling period d, the node density δ, the energy per sample E, the number of sensors M and the time instant t. As $n \to \infty$ and $M \to \infty$, we have the following asymptotic results of $\sigma^2_{\hat{m}_n}(\mathbf{c})$ and $\hat{k}_n(\mathbf{c})$.

Theorem 2 *Define the asymptotic posterior variance* $\sigma^2_e(u,v) := \lim_{n,M\to\infty}\hat{k}_n(\mathbf{c})$, *where* $u = \frac{s}{l} - \lfloor\frac{s}{l}\rfloor \in [0,1]$ *is the relative position of* s *between two adjacent samples in space and* $v = \frac{t}{d} - \lfloor\frac{t}{d}\rfloor \in [0,1]$ *is the relative position of* t *between two adjacent samples in time. We have*

$$\sigma^2_e(u,v) = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}\left[\Psi_{00}(f_s,f_t) - \frac{|\Psi_{uv}(f_s,f_t)|^2}{\Psi_{00}(f_s,f_t) + \frac{1}{\gamma_0(d/\delta - d_c)}}\right]df_s\,df_t \tag{16}$$

where $\gamma_0 := \frac{P_0}{\sigma^2}$ *is the signal-to-noise ratio (SNR),* $d_c := \frac{E_c}{P_0}$ *is hardware energy normalized by the average power constraint, and*

$$\Psi_{uv}(f_s,f_t) := \sum_{m=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} k_s((m+u)l)k_t((n+v)d)e^{-2\pi j(mf_s+nf_t)} \tag{17}$$

is the 2-D discrete-time Fourier transform (DTFT) of sequences $\left\{k_s((m+u)l)k_t((n+v)d)\right\}_{m,n}$.

In addition,

$$\lim_{n,M\to\infty}\sigma^2_{\hat{m}_n}(\mathbf{c}) = 1 - \sigma^2_e(u,v). \tag{18}$$

Proof: To simplify notation, denote $x_{i+u,n+v} = x([(i+u)l,(n+v)d]^T)$ and $y_{i,n} = y([il,nd]^T)$. The linear minimum mean squared error (LMMSE) estimation of $x_{i+u,n+v}$ based on $\{y_{i,n}\}_{i,n}$ is

$$\hat{x}_{i+u,n+v} = \sum_{m=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} h(m,k)y_{i-m,n-k} \tag{19}$$

where $\{h(m,k)\}$ is the impulse response of the LMMSE filter.

Based on the orthogonal principle, $\mathbb{E}[(x_{i+u,n+v} - \hat{x}_{i+u,n+v})y_{i',n'}] = 0$, we have

$$\sum_{m=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} h(m,k)r_{yy}(i-m,n-k) = r_{xy}(i+u,n+v) \tag{20}$$

where

$$r_{yy}(i,n) = \mathbb{E}[y_{i'+i,n'+n}y_{i',n'}] = E\,k_s(il)k_t(nd) + \sigma^2 \tag{21}$$

$$\begin{aligned} r_{xy}(i+u,n+v) &= \mathbb{E}[x_{i'+i+u,n'+n+v}y_{i',n'}] \\ &= \sqrt{E}k_s((i+u)l)k_t((n+v)d) \end{aligned} \tag{22}$$

Based on the convolution theorem [14], converting (20)-(22) into the frequency domain with 2-D DTFT, we have

$$H(f_s,f_t) = \frac{\sqrt{E}\Psi_{uv}(f_s,f_t)}{E\,\Psi_{00}(f_s,f_t) + \sigma^2} \tag{23}$$

Based on the orthogonal principle, the MSE can be calculated as

$$\sigma^2_e(u,v) = \mathbb{E}[(x_{i+u,n+v} - \hat{x}_{i+u,n+v})x_{i+u,n+v}] \tag{24}$$

Combining (19) and (24) yields

$$\sigma_e^2(u,v) = k_s(0)k_t(0) - \sqrt{E}\sum_{m=-\infty}^{+\infty}\sum_{k=-\infty}^{+\infty} h(m,k)k_s(-(m+u)l)k_t(-(k+v)d) = \int_{-\frac{1}{2}}^{\frac{1}{2}}\int_{-\frac{1}{2}}^{\frac{1}{2}}[\Psi_{00}(f_s,f_t) - H(f_s,f_t)\Psi_{uv}^*(f_s,f_t)]\,df_s df_t \tag{25}$$

Then (16) can be obtained by combining the above equation with (23). ■

From Theorem 2, we can see that as $n, M \to \infty$, $\hat{m}_n(il+u, kd+v)$ is a zero-mean Gaussian random variable with variance $1-\sigma_e^2(u,v)$. Therefore, as $n, M \to \infty$, the statistical properties of $\hat{m}_n(\mathbf{c})$ are periodic in space and time with periods l and d, respectively. Define $\hat{m}(u,v) := \lim_{n,M\to\infty}\hat{m}_n([(i+u)l,(k+v)d]^T)$. We have the following corollary regarding the distribution of $\hat{m}(u,v)$.

Corollary 1 *As $n, M \to \infty$, $\hat{m}(u,v)$ is zero-mean Gaussian distributed with variance $1-\sigma_e^2(u,v)$, that is, $\hat{m}(u,v) \sim \mathcal{N}(0, 1-\sigma_e^2(u,v))$.*

With the asymptotic results in Theorem 2 and Corollary 1, we can get an explicit expression of the cost function in (5), and the result is given in the following theorem.

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

$$\mathcal{J}(d,l) = \frac{1}{\pi}\int_0^1\int_0^1\int_0^{\pi/2}\left[1+\frac{1-\sigma_e^2(u,v)}{\sigma_e^2(u,v)\sin^2(\theta)}\right]^{-1/2} \times \exp\left(-\frac{\gamma^2/2}{1-\sigma_e^2(u,v)\cos^2(\theta)}\right)du\,dv\,d\theta. \tag{26}$$

Proof: We first consider the conditional cost function, $\lim_{n,M\to\infty}\frac{1}{nd\cdot Ml}\mathbb{E}[e(nd)|\mathbf{y}_{1:n}]$. From (8)-(12), we have

$$\mathbb{E}[e(nd)|\mathbf{y}_{1:n}] = \sum_{i=1}^{M}\sum_{k=1}^{n}\int_{(i-1)l}^{il}\int_{(k-1)d}^{kd} Q\left(\frac{|\gamma-\hat{m}_n(\mathbf{c})|}{\sqrt{\hat{k}_n(\mathbf{c})}}\right)ds\,dt$$

Performing change of variable, $s = (i-1)l + ul$, and $t = (k-1)d + vd$ in the above integral , and using the results from Theorem 2, we have

$$\lim_{n,M\to\infty}\frac{1}{nd\cdot Ml}\mathbb{E}[e(\hat{\mathcal{S}}(nd))|\mathbf{y}_{1:n}] = \int_0^1\int_0^1 Q\left(\frac{|\gamma-\hat{m}(u,v)|}{\sqrt{\sigma_e^2(u,v)}}\right)du\,dv \tag{27}$$

In the conditional cost function in (27), there is only one random variable, $\hat{m}(u,v) \sim \mathcal{N}(0, 1-\sigma_e^2(u,v))$, which is a function of $\mathbf{y}_{1:n}$. Therefore the unconditional cost function can be expressed as

$$\mathcal{J}(d,l) = \int_0^1\int_0^1 \mathbb{E}_{\hat{m}(u,v)}\left[Q\left(\frac{|\gamma-\hat{m}(u,v)|}{\sqrt{\sigma_e^2(u,v)}}\right)\right]du\,dv \tag{28}$$

With Craig's alternative expression of the Q-function [15], (28) can be reformulated as

$$\mathcal{J}(d,l) = \frac{1}{\pi}\int_0^1\int_0^1\int_0^{\frac{\pi}{2}} \mathbb{E}_{\hat{m}(u,v)}\left[e^{-\frac{(\hat{m}(u,v)-\gamma)^2}{2\sigma_e^2(u,v)\sin^2(\theta)}}\right]du\,dv d\theta \tag{29}$$

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

$$M_Z(s) = \exp\left[\frac{s}{(1-2s)}\frac{\gamma^2}{1-\sigma_e^2(u,v)}\right]\frac{1}{\sqrt{1-2s}}. \tag{30}$$

Combining (29) with (30) yields (26). ■

Theorem 3 gives the exact analytical expression of the cost function, which is expressed as a function of d, δ and other system parameters such as the SNR γ_0, the temporal and spatial correlation covariance function $k_t(x)$ and $k_s(x)$, and the normalized hardware energy consumption d_c. Thus, given $\{\gamma_0, \rho, d_c, k_t, k_s\}$, we can identify d and δ that minimizes the cost function $\mathcal{J}$ in (26). The integration limits in (26) are finite, thus the integrals in (26) can be evaluated with numerical integration with high precision.

In the special case that the covariance functions are the power law kernels, i.e., $k_s(x) = \rho_s^{|x|}$ and $k_t(x) = \rho_t^{|x|}$ with ρ_s and ρ_t being the power law coefficients in the spatial and temporal domains, respectively, we can get the closed-form expressions of the 2-D DTFT in (17)

$$\Psi_{uv}(f_s,f_t) = \frac{\rho_t^{-vd}[\rho_t^d(1-\rho_t^{2vd})e^{2\pi j f_t} + \rho_t^{2vd} - \rho_t^{2d}]}{1+\rho_t^{2d}-2\rho_t^d\cos(2\pi f_t)} \times \frac{\rho_s^{-ul}[\rho_s^l(1-\rho_s^{2ul})e^{2\pi j f_s} + \rho_s^{2ul} - \rho_s^{2l}]}{1+\rho_s^{2l}-2\rho_s^l\cos(2\pi f_s)} \tag{31}$$

Numerical results show that the cost function with the power law covariance is convex in both sampling period d and node distance l, thus we can always identify the optimal d and l by using the Karush-Kuhn-Tucker (KKT) conditions.

V. Simulation Results

Simulation results are presented in this section to demonstrate the performance of the proposed level set estimation algorithm with power law kernels. The threshold of the level set γ is set to be 0.1.

Fig. 1 shows the asymptotic LSE error probability as a function of the sampling period d under different node distances l. The SNR is $\gamma_0 = 10$ dB. The temporal and spatial power law coefficients are set to be $\rho_t = 0.5$ and $\rho_s = 0.8$, respectively. The normalized hardware energy is $d_c = 0.05$. The simulation results are obtained with $n = 100$ and $M = 100$. The simulation results with finite n and M matches very well with the numerically analytical results with

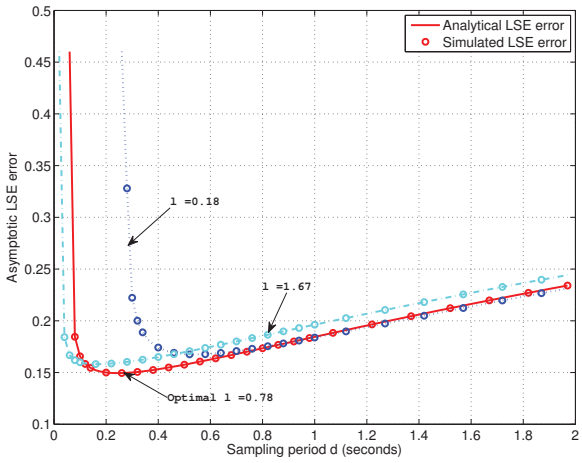

Fig. 1. Asymptotic LSE error probabilities as a function of sampling period d under various node distances l.

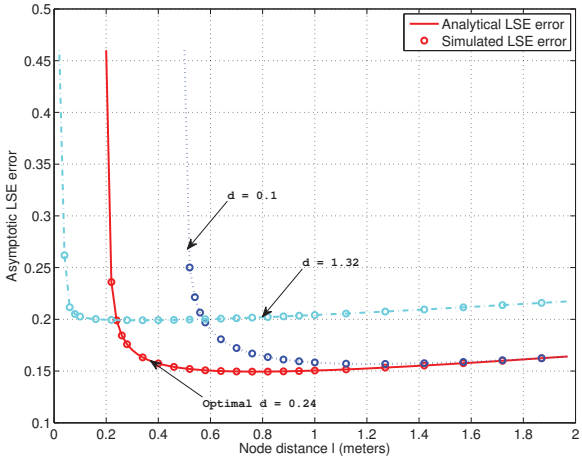

Fig. 2. Asymptotic LSE error probabilities as a function of node distances l under various sampling periods d.

infinite n and M. Thus the asymptotic analytical results give a very good approximation of the performance of practical systems with finite n and M. Given l, the asymptotic LSE error probability is a convex function of sampling period d for $d > d_c/l$. When $d = d_c/l$, all energy is consumed by the sensing operation and there is no energy to transmit information to FC, which leads to a maximum LSE error, $Q(|\gamma|) = 0.46$, the same error with a random decision estimator. The minimum LSE error probability is achieved with the optimal $l^* = 0.78$, which corresponds to an LSE error probability of 0.149 .

Similarly, Fig. 2 presents the asymptotic LSE error probability as a function of node distance l with different sampling periods d. All other parameters are the same as Fig. 1. Given d, the asymptotic LSE error probability is also a convex function of node distance l for $l > d_c/d$. The special case of $l = d_c/d$ results in a maximum LSE error, $Q(|\gamma|) = 0.46$. The minimum

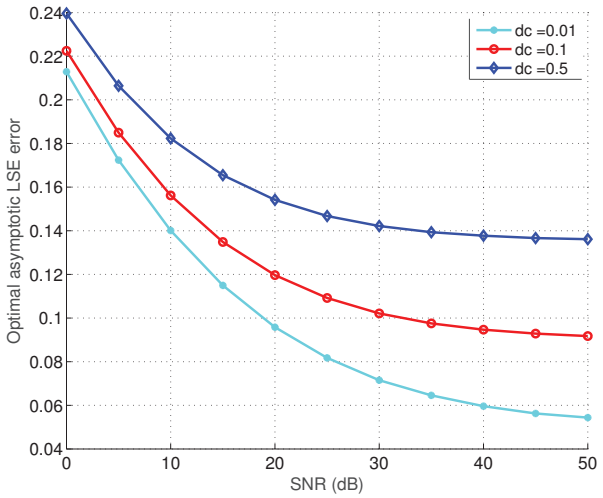

Fig. 3. Optimal asymptotic LSE error probabilities as a function of SNR under various d_c ($\rho = 0.5$).

LSE error probability 0.149 is achieved with $d^* = 0.24$ and $l^* = 0.78$.

Fig. 3 shows the optimal asymptotic LSE error probability as a function of the SNR, under different values of d_c. The power law coefficients are $\rho_t = 0.5$ and $\rho_s = 0.8$. The optimal asymptotic LSE error probability is obtained by identifying the optimal values of d and l for each configuration. As expected, the optimal asymptotic LSE error probability is a decreasing function of SNR. For a given SNR, a larger d_c yields a larger optimal asymptotic LSE error probability, due to the fact that more energy are consumed by the hardware.

VI. Conclusions

We have studied the optimal level set estimation of a temporally-spatially correlated random field with linear sensor network, under a total power constraint. The optimal LSE algorithm was developed by using regression of a 2-D Gaussian process. Then the exact LSE error probability was derived with the assistance of frequency domain analysis. The LSE error probability has been expressed as an explicit function of a number of system parameters, such as the node distance in the space domain, the sampling period in the time domain, the covariance kernel functions, and the SNR, etc. The optimum node distance and sampling period can then be identified to minimize the LSE error probability. Simulation results demonstrated that the proposed algorithm can achieve accurate and efficient LSE.

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] 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.

[3] 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.

[4] Z. Wang, J. Wu, J. Yang, and H. Lin, "Energy-efficient wireless sensing for level set estimations," *Access, IEEE*, vol. 3, pp. 1480–1490, 2015.

[5] M. Roederer and R. R. Hardy, "Frequency difference gating: a multivariate method for identifying subsets that differ between samples," *Cytometry*, vol. 45, no. 1, pp. 56–64, 2001.

[6] 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.

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

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

[9] 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.

[10] 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.

[11] 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.

[12] Y. Ma, R. Garnett, and J. Schneider, "Active area search via bayesian quadrature," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, 2014, pp. 595–603.

[13] C. Guestrin, A. Krause, and A. P. Singh, "Near-optimal sensor placements in gaussian processes," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 265–272.

[14] E. W. Weisstein, "Convolution theorem," *From MathWorld-A Wolfram Web Resource, 2006a. URL http://mathworld. wolfram. com/ConvolutionTheorem. html*, 2014.

[15] 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.