

Article

# Adaptive Sampling for Learning Gaussian Processes Using Mobile Sensor Networks

Yunfei Xu <sup>1</sup> and Jongeun Choi <sup>1,2,\*</sup>

<sup>1</sup> Department of Mechanical Engineering, Michigan State University, East Lansing, MI 48823, USA;  
E-Mail: xuyunfei@egr.msu.edu

<sup>2</sup> Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48823, USA

\* Author to whom correspondence should be addressed; E-Mail: jchoi@egr.msu.edu;  
Tel.: +1-517-432-3164, Fax: +1-517-353-1750.

Received: 4 January 2011; in revised form: 25 February 2011 / Accepted: 27 February 2011 /  
Published: 9 March 2011

---

**Abstract:** This paper presents a novel class of self-organizing sensing agents that adaptively learn an anisotropic, spatio-temporal Gaussian process using noisy measurements and move in order to improve the quality of the estimated covariance function. This approach is based on a class of anisotropic covariance functions of Gaussian processes introduced to model a broad range of spatio-temporal physical phenomena. The covariance function is assumed to be unknown a priori. Hence, it is estimated by the maximum a posteriori probability (MAP) estimator. The prediction of the field of interest is then obtained based on the MAP estimate of the covariance function. An optimal sampling strategy is proposed to minimize the information-theoretic cost function of the Fisher Information Matrix. Simulation results demonstrate the effectiveness and the adaptability of the proposed scheme.

**Keywords:** mobile sensor networks; Gaussian processes; adaptive sampling

---

## 1. Introduction

In recent years, due to global climate changes, more environmental scientists are interested in the change of ecosystems over vast regions in lands, oceans, and lakes. For instance, for certain environmental conditions, rapidly reproducing harmful algal blooms in the Great Lakes can produce

cyanotoxins. Besides such natural disasters, there exist growing ubiquitous possibilities of the release of toxic chemicals and contaminants in the air, lakes, and public water systems. This resulted in the rising demands to utilize autonomous robotic systems that can perform a series of tasks such as estimation, prediction, monitoring, tracking and removal of a scalar field of interest undergoing often complex transport phenomena (Common examples are diffusion, convection, and advection).

Significant enhancements have been made in the areas of mobile sensor networks and mobile sensing vehicles such as unmanned ground vehicles, autonomous underwater vehicles, and unmanned aerial vehicles. Emerging technologies have been reported on the coordination of mobile sensing agents [1–6]. Mobile sensing agents form an ad-hoc wireless communication network in which each agent usually operates under a short communication range, with limited memory and computational power. Mobile sensing agents are often spatially distributed in an uncertain surveillance environment.

The mobility of mobile agents can be designed in order to perform the optimal sampling of the field of interest. Recently in [5], Leonard *et al.* developed mobile sensor networks that optimize ocean sampling performance defined in terms of uncertainty in a model estimate of a sampled field. In [6], distributed learning and cooperative control were developed for multi-agent systems to discover peaks of the unknown field based on the recursive estimation of an unknown field. In general, we design the mobility of sensing agents to find the most informative locations to make observations for a spatio-temporal phenomenon. To find these locations that predict the phenomena best, one needs a model of the spatio-temporal phenomenon itself. In our approach, we focus on Gaussian processes to model fields undergoing transport phenomena. A Gaussian process (or kriging in geostatistics) has been widely used as a nonlinear regression technique to estimate and predict geostatistical data [7–11]. A Gaussian process is a natural generalization of the Gaussian probability distribution. It generalizes the Gaussian distribution with a finite number of random variables to a Gaussian process with an infinite number of random variables in the surveillance region. Gaussian process modeling enables us to predict physical values, such as temperature and plume concentration, at any of spatial points with a predicted uncertainty level efficiently. For instance, near-optimal static sensor placements with a mutual information criterion in Gaussian processes were proposed by [12,13]. Distributed kriged Kalman filter for spatial estimation based on mobile sensor networks are developed by [14]. A distributed adaptive sampling approach was proposed by [15] for sensor networks to find locations that maximize the information contents by assuming that the covariance function is known up to a scaling parameter. Multi-agent systems that are versatile for various tasks by exploiting predictive posterior statistics of Gaussian processes were developed by [16,17].

The motivation of our work is as follows. Even though there have been efforts to utilize Gaussian processes to model and predict the spatio-temporal field of interest, most of recent papers assume that Gaussian processes are isotropic, implying that the covariance function only depends on the distance between locations. Many studies also assume that the corresponding covariance functions are known a priori for simplicity. However, this is not the case in general as pointed out in literature [12,13,18], in which they treat the non-stationary process by fusing a collection of isotropic spatial Gaussian processes associated with a set of local regions. Hence our motivation is to develop theoretically-sound algorithms for mobile sensor networks to learn the anisotropic covariance function of a spatio-temporal Gaussian

process. Mobile sensing agents can then predict the Gaussian process based on the estimated covariance function in a nonparametric manner.

The contribution of this paper is to develop covariance function learning algorithms for the sensing agents to perform nonparametric prediction based on a properly adapted Gaussian process for a given spatio-temporal phenomenon. By introducing a generalized covariance function, we expand the class of Gaussian processes to include the anisotropic spatio-temporal phenomena. The maximum a posteriori probability (MAP) estimator is used to find hyperparameters for the associated covariance function. The proposed optimal navigation strategy for autonomous vehicles minimizes the information-theoretic cost function such as D-or A-optimality criterion using the Fisher Information Matrix (or Cramér-Rao lower bound (CRLB)[19], improving the quality of the estimated covariance function. A Gaussian process with a time-varying covariance function has been proposed to demonstrate the adaptability of the proposed scheme.

This paper is organized as follows. In Section 2, we briefly review the mobile sensing network model and the notation related to a graph. A nonparametric approach to predict a field of interest based on measurements is presented in Section 3. Section 4 introduces a covariance function learning algorithm for an anisotropic, spatio-temporal Gaussian process. An optimal navigation strategy is described in Section 5. In Section 6, simulation results illustrate the usefulness of our approach and its adaptability for unknown and/or time-varying covariance functions.

The standard notation will be used in the paper. Let  $\mathbb{R}, \mathbb{R}_{\geq 0}, \mathbb{Z}$  denote, respectively, the set of real, non-negative real, and integer numbers. The positive semi-definiteness of a matrix  $A$  is denoted by  $A \succeq 0$ . Let  $|B|$  denotes the determinant of a matrix  $B$ .  $\mathbb{E}$  denote the expectation operator.

## 2. Mobile Sensor Networks

First, we explain the mobile sensing network and the measurement model used in this paper. Let  $N_s$  be the number of sensing agents distributed over the surveillance region  $\mathcal{Q} \in \mathbb{R}^2$ . Assume that  $\mathcal{Q}$  is a compact set. The identity of each agent is indexed by  $\mathcal{I} := \{1, 2, \dots, N_s\}$ . Let  $q_i(t) \in \mathcal{Q}$  be the location of the  $i$ -th sensing agent at time  $t \in \mathbb{R}_{\geq 0}$ . We assume that the measurement  $y(q_i(t), t)$  of agent  $i$  is the sum of the scalar value of the Gaussian process  $z(q_i(t), t)$  and sensor noise  $w_i(t)$ , at its position  $q_i(t)$  and some measurement time  $t$ ,

$$y(q_i(t), t) := z(q_i(t), t) + w_i(t).$$

The communication network of mobile agents can be represented by a graph with edges. Let  $G(t) = (\mathcal{I}, \mathcal{E}(t))$  be an undirected communication graph such that an edge  $(i, j) \in \mathcal{E}(t)$  if and only if agent  $i$  can communicate with agent  $j \neq i$ . We define the neighborhood of agent  $i$  at time  $t$  by  $N_i(t) := \{j \mid (i, j) \in \mathcal{E}(t), i \in \mathcal{I}\}$ . We also define the closed neighborhood of agent  $i$  at time  $t$  by the union of its index and its neighbors, *i.e.*,  $\bar{N}_i(t) := \{i\} \cup N_i(t)$ .

## 3. The Nonparametric Approach

With the spatially distributed sampling capability, agents need to estimate and predict the field of interest by fusing the collective samples from different space and time indices. We show a nonparametric approach to predict a field of interest based on measurements. We assume that a field undergoing a

physical transport phenomenon can be modeled by a spatio-temporal Gaussian process, which can be used for nonparametric prediction.

Consider a spatio-temporal Gaussian process:

$$z(s, t) \sim \mathcal{GP}(\mu(s, t), \mathcal{K}(s, t, s', t')) \quad (1)$$

where  $s, s' \in \mathcal{Q}$ ,  $t, t' \in \mathbb{R}_{\geq 0}$  and  $\mu(s, t)$  denotes the mean value at location  $s$  and time  $t$ . We then propose the following generalized covariance function  $\mathcal{K}(s, t, s', t'; \Psi)$  with a hyperparameter vector  $\Psi := [\sigma_f \ \sigma_x \ \sigma_y \ \sigma_t]^T$ :

$$\mathcal{K}(s, t, s', t'; \Psi) = \sigma_f^2 \exp\left(-\sum_{l \in \{x, y\}} \frac{(s_l - s'_l)^2}{2\sigma_l^2}\right) \exp\left(-\frac{(t - t')^2}{2\sigma_t^2}\right) \quad (2)$$

where  $s_l$  is the  $l$ -th entry of  $s$ .  $\{\sigma_x, \sigma_y\}$  and  $\sigma_t$  are kernel bandwidths for space and time, respectively. Equation (2) shows that points close in the measurement space and time indices are strongly correlated and produce similar values. In reality, the larger temporal distance two measurements are taken with, the less correlated they become, which strongly supports our generalized covariance function in Equation (2). This may also justify the truncation (or windowing) of the observed time series data to limit the size of the covariance matrix for reducing the computational cost.

In the case that the global coordinates are different from the local model coordinates, a similarity transformation can be used to address this issue. For instance, a rotational relationship between the model basis  $\{\vec{e}_x, \vec{e}_y\}$  and the global basis  $\{\vec{E}_x, \vec{E}_y\}$  is:

$$\begin{bmatrix} \vec{e}_x \\ \vec{e}_y \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \vec{E}_x \\ \vec{E}_y \end{bmatrix}$$

where  $\theta$  represents the angle of rotation. We can then use the following relationship to change the coordinates:

$$\begin{cases} x = X \cos \theta + Y \sin \theta \\ y = -X \sin \theta + Y \cos \theta \end{cases}$$

where  $x$  and  $y$  indicate coordinates in the local basis and  $X$  and  $Y$  indicate their counterparts in the global basis. Equation (2) can then be rewritten in terms of global coordinates as

$$\begin{aligned} \mathcal{K}(s, t, s', t'; \Psi) &= \sigma_f^2 \exp\left(-\frac{[(s_X - s'_X) \cos \theta + (s_Y - s'_Y) \sin \theta]^2}{2\sigma_x^2}\right) \\ &\cdot \exp\left(-\frac{[-(s_X - s'_X) \sin \theta + (s_Y - s'_Y) \cos \theta]^2}{2\sigma_y^2}\right) \exp\left(-\frac{(t - t')^2}{2\sigma_t^2}\right) \end{aligned}$$

where  $s_X$  and  $s_Y$  are the coordinates in the global basis. In this case, the parameter vector  $\Psi$  is redefined as  $\Psi := [\sigma_f \ \sigma_x \ \sigma_y \ \sigma_t \ \theta]^T$ .

Up to time  $t_k$ , agent  $i$  has noisy collective data  $\{y(q_j(t_m), t_m) \mid m \in \mathbb{Z}, j \in \bar{N}_i(t_m), 1 \leq m \leq k\}$ , where  $\bar{N}_i(t_m)$  denotes the closed neighborhood of agent  $i$  at time  $t_m$ . The measurements  $y(q_j(t_m), t_m) = z(q_j(t_m), t_m) + w_j(t_m)$  are taken at different positions  $q_j(t_m) \in \mathcal{Q}$  and different times  $t_m \in \mathbb{R}_{\geq 0}$ . The measurements are corrupted by the sensor and communication noises

represented by Gaussian white noise  $w_j \sim \mathcal{N}(0, \sigma_w^2)$ . For the case in which the noise level  $\sigma_w$  is not known and needs to be estimated, the hyperparameter vector can be expanded to include  $\sigma_w$ , *i.e.*,  $\Psi := [\sigma_f \ \sigma_x \ \sigma_y \ \sigma_t \ \sigma_w]^T$ . The column-vectorized measurements collected by agent  $i$  is denoted by

$$Y_{\leq k} := \text{col} (y(q_j(t_m), t_m) \mid m \in \mathbb{Z}, j \in \bar{N}_i(t_m), 1 \leq m \leq k)$$

with a joint distribution

$$p(Y_{\leq k} \mid \Psi) := \frac{1}{(2\pi)^{n/2} |\Sigma_{Y_{\leq k}}|^{1/2}} \exp \left( -\frac{1}{2} (Y_{\leq k} - \mu_{Y_{\leq k}})^T \Sigma_{Y_{\leq k}}^{-1} (Y_{\leq k} - \mu_{Y_{\leq k}}) \right)$$

where  $n$  is the total number of observations up to time  $t_k$ ,  $\mu_{Y_{\leq k}} := \mathbb{E}(Y_{\leq k})$  is the mean vector of  $Y_{\leq k}$ ,  $\Sigma_{Y_{\leq k}} := \mathbb{E}((Y_{\leq k} - \mu_{Y_{\leq k}})(Y_{\leq k} - \mu_{Y_{\leq k}})^T)$  is the covariance matrix of  $Y_{\leq k}$  obtained by  $[\Sigma_{Y_{\leq k}}]_{ij} = \mathcal{K}(s_i, t_i, s_j, t_j) + \sigma_w^2 \delta_{ij}$  in which  $\delta_{ij}$  denotes the Kronecker delta function.

If the covariance function is known a priori, the prediction of the random field  $z(s, t)$  at location  $s$  and time  $t$  is then obtained by

$$z(s, t \mid t_k) := z(s, t) \mid Y_{\leq k} \sim \mathcal{N}(\hat{z}(s, t \mid t_k), \sigma^2(s, t \mid t_k)) \quad (3)$$

where  $\hat{z}(s, t \mid t_k) := \mathbb{E}(z(s, t \mid t_k))$  is

$$\hat{z}(s, t \mid t_k) := \mu(s, t) + \Sigma_{zY_{\leq k}} \Sigma_{Y_{\leq k}}^{-1} (Y_{\leq k} - \mu_{Y_{\leq k}})$$

and the prediction error variance is

$$\sigma^2(s, t \mid t_k) := \Sigma_z - \Sigma_{zY_{\leq k}} \Sigma_{Y_{\leq k}}^{-1} \Sigma_{Y_{\leq k}z}$$

where  $\Sigma_z$  is the covariance of  $z$ , obtained by  $\mathcal{K}(s, t, s, t; \Psi)$ ,  $\Sigma_{zY_{\leq k}} = \Sigma_{Y_{\leq k}z}^T$  is the covariance matrix between  $z$  and  $Y_{\leq k}$ , obtained by  $[\Sigma_{zY_{\leq k}}]_j = \mathcal{K}(s, t, s_j, t_j; \Psi)$ . Each agent can then predict the field of interest at any location and time with the associated uncertainty in a nonparametric way. In the next section, we present a learning approach for unknown covariance functions.

#### 4. The MAP Estimate of the Hyperparameter Vector

Without loss of generality, we use a zero mean Gaussian process  $z(s, t) \sim \mathcal{GP}(0, \mathcal{K}(s, t, s', t'))$  for modeling the field undergoing a physical transport phenomenon. This is not a strong limitation since the mean of the posterior process is not confined to zero [11].

If the covariance function of a Gaussian process is not known a priori, mobile agents need to estimate parameters of the covariance function ( $\Psi$ ) based on the observed samples. Using Bayes' rule, the posterior  $p(\Psi \mid Y_{\leq k})$  is proportional to the likelihood  $p(Y_{\leq k} \mid \Psi)$  times the prior  $p(\Psi)$ , *i.e.*,

$$p(\Psi \mid Y_{\leq k}) \propto p(Y_{\leq k} \mid \Psi) p(\Psi)$$

At time  $t_k$ , the maximum a posteriori (MAP) estimate  $\hat{\Psi}_k$  of the hyperparameter vector can be obtained by

$$\begin{aligned} \hat{\Psi}_k &= \arg \max_{\Psi} p(\Psi \mid Y_{\leq k}) \\ &= \arg \max_{\Psi} p(Y_{\leq k} \mid \Psi) p(\Psi) \end{aligned} \quad (4)$$

This is equivalent to maximize the logarithm of the posterior  $p(\Psi|Y_{\leq k})$ , *i.e.*,

$$\hat{\Psi}_k = \arg \max_{\Psi} (\ln p(Y_{\leq k}|\Psi) + \ln p(\Psi))$$

The log likelihood function is given by

$$\ln p(Y_{\leq k}|\Psi) = -\frac{1}{2} Y_{\leq k}^T \Sigma_{Y_{\leq k}}^{-1} Y_{\leq k} - \frac{1}{2} \ln |\Sigma_{Y_{\leq k}}| - \frac{n}{2} \ln 2\pi$$

where  $n$  is the size of  $Y_{\leq k}$ . Notice that if no prior information is given, the MAP estimate in Equation (4) is equal to the maximum likelihood (ML) estimate.

A gradient ascent algorithm is used to find a MAP estimate of  $\Psi$ :

$$\hat{\Psi}_k^{(i+1)} = \hat{\Psi}_k^{(i)} + \epsilon_k^{(i)} \nabla_{\hat{\Psi}_k^{(i)}} \ln p(\hat{\Psi}_k^{(i)}|Y_{\leq k}), \quad i \geq 0$$

where  $\epsilon_k^{(i)}$  is a small positive number which can be obtained by using a backtracking line search and  $\nabla_x f(x)$  is the partial derivative of  $f(x)$  with respect to  $x$ . The partial derivative of the log likelihood function with respect to a hyperparameter  $\psi_j$  is given by

$$\begin{aligned} \frac{\partial \ln p(Y_{\leq k}|\Psi)}{\partial \psi_j} &= \frac{1}{2} Y_{\leq k}^T \Sigma_{Y_{\leq k}}^{-1} \frac{\partial \Sigma_{Y_{\leq k}}}{\partial \psi_j} \Sigma_{Y_{\leq k}}^{-1} Y_{\leq k} - \frac{1}{2} \text{tr} \left( \Sigma_{Y_{\leq k}}^{-1} \frac{\partial \Sigma_{Y_{\leq k}}}{\partial \psi_j} \right) \\ &= \frac{1}{2} \text{tr} \left( (\beta \beta^T - \Sigma_{Y_{\leq k}}^{-1}) \frac{\partial \Sigma_{Y_{\leq k}}}{\partial \psi_j} \right) \end{aligned} \quad (5)$$

where  $\beta = \Sigma_{Y_{\leq k}}^{-1} Y_{\leq k}$ . Alternatively, a simplex search method [20] can be used to find a MAP estimate of  $\Psi$ . This is a direct search method that does not use numerical or analytic gradients.

After finding a MAP estimate of  $\Psi$ , agents can proceed the prediction of the field of interest using Equation (3).

## 5. An Adaptive Sampling Strategy

Agents should find new sampling positions to improve the quality of the estimated covariance function in the next iteration at time  $t_{k+1}$ . For instance, to precisely estimate the anisotropic phenomenon, *i.e.*, processes with different correlations along  $x$ -axis and  $y$ -axis directions, sensing agents need to explore and sample measurements along different directions.

To this end, we consider a centralized scheme. Suppose that a leader agent (or a central station) knows the communication graph at the next iteration time  $t_{k+1}$  and also has access to all measurements collected by agents. Let  $Y_{k+1}$  and  $Y_{\leq k}$  be the measurements at time  $t_{k+1}$  and the collective measurements up to time  $t_k$ , respectively, *i.e.*,

$$\begin{aligned} Y_{k+1} &:= \text{col}(y(q_i(t_{k+1}), t_{k+1}) \mid i \in \mathcal{I}), \\ Y_{\leq k} &:= \text{col}(y(q_i(t_m), t_m) \mid m \in \mathbb{Z}, i \in \mathcal{I}, 1 \leq m \leq k) \end{aligned}$$

To derive the optimal navigation strategy, we compute the log likelihood function of observations of  $Y_{\leq k+1}$ :

$$\begin{aligned} \mathcal{L}(Y_{\leq k+1}, \Psi) &:= \ln p(Y_{\leq k+1}|\Psi) \\ &= -\frac{1}{2} Y_{\leq k+1}^T \Sigma_{Y_{\leq k+1}}^{-1} Y_{\leq k+1} - \frac{1}{2} \ln |\Sigma_{Y_{\leq k+1}}| - \frac{n_{\leq k+1}}{2} \ln 2\pi \end{aligned} \quad (6)$$

where  $n_{\leq k+1}$  is the size of  $Y_{\leq k+1}$ .

Since the locations of observations in  $Y_{\leq k}$  were already fixed, we represent the log likelihood function in terms of a vector of future sampling points  $\tilde{q}$  at time  $t_{k+1}$  only and the hyperparameter vector  $\Psi$ :

$$\mathcal{L}(\tilde{q}, \Psi) := \ln p(Y_{\leq k+1}(\tilde{q})|\Psi)$$

Now consider the Fisher Information Matrix (FIM) that measures the information produced by measurements  $Y_{\leq k+1}$  for estimating the hyperparameter vector at time  $t_{k+1}$ . The Cramér-Rao lower bound (CRLB) theorem states that the inverse of the FIM is a lower bound of the estimation error covariance matrix [19,21]:

$$\mathbb{E} \left( (\hat{\Psi}_{k+1} - \Psi)(\hat{\Psi}_{k+1} - \Psi)^T \right) \succeq \text{FIM}^{-1}$$

where  $\hat{\Psi}_{k+1}$  represents the estimation of  $\Psi$  at time  $t_{k+1}$ . The FIM [19] is given by

$$[\text{FIM}(\tilde{q}, \Psi)]_{ij} = -\mathbb{E} \left( \frac{\partial^2 \mathcal{L}(\tilde{q}, \Psi)}{\partial \psi_i \partial \psi_j} \right) \quad (7)$$

where the expectation is taken with respect to  $p(Y_{\leq k+1}|\Psi)$ . The analytical closed-form of FIM is given by

$$[\text{FIM}(\tilde{q}, \Psi)]_{ij} = \frac{1}{2} \text{tr} \left( \Sigma_{Y_{\leq k+1}}^{-1} \frac{\partial \Sigma_{Y_{\leq k+1}}}{\partial \psi_i} \Sigma_{Y_{\leq k+1}}^{-1} \frac{\partial \Sigma_{Y_{\leq k+1}}}{\partial \psi_j} \right)$$

Since the true value of  $\Psi$  is not available, we will evaluate the FIM in Equation (7) at the currently available best estimate  $\hat{\Psi}_k$ . This has been an effective practical solution when we evaluate the FIM and estimate  $\Psi$  simultaneously [22,23]. The term due to the MAP estimation error in evaluating the FIM in Equation (7) will decrease as the number of samples increases.

We can expect that minimizing the CRLB results in a decrease of uncertainty in estimating  $\Psi$  [22]. Using the D-optimality criterion [24,25], the objective function  $J$  is given by

$$J(\tilde{q}, \hat{\Psi}_k) := \det \left( \text{FIM}^{-1}(\tilde{q}, \hat{\Psi}_k) \right) \quad (8)$$

Minimizing  $J$  in Equation (8) corresponds to minimizing the volume of the ellipsoid which represents the maximum confidence region for the estimated hyperparameters. However, if one hyperparameter has a much larger variance compared to the others, minimizing the volume may not be very useful [25]. As an alternative, the A-optimality which minimizes the sum of the variances may be used. The objective function  $J$  based on the A-optimality criterion is

$$J(\tilde{q}, \hat{\Psi}_k) := \text{tr} \left( \text{FIM}^{-1}(\tilde{q}, \hat{\Psi}_k) \right) \quad (9)$$

A control law for the mobile sensor network can be formulated as follows:

$$q(t_{k+1}) = \arg \min_{\tilde{q} \in \mathcal{Q}} J(\tilde{q}, \hat{\Psi}_k) \quad (10)$$

A gradient descent strategy can be used to find the next optimal sampling positions:

$$\tilde{q}^{(i+1)} = \tilde{q}^{(i)} - \alpha^{(i)} \nabla_{\tilde{q}^{(i)}} J(\tilde{q}^{(i)}, \hat{\Psi}_k) \quad (11)$$

where  $\alpha^{(i)}$  is a small positive number which can be obtained by using a backtracking line search. Alternatively, a control law for the mobile sensor network can be formulated as follows:

$$q(t_{k+1}) = \arg \min_{\tilde{q} \in \prod_{i=1}^{N_s} Q_i} J(\tilde{q}, \hat{\Psi}_k)$$

where

$$Q_i = q_i(t_k) + \prod_{j=1}^{n_d} [-\delta_j \delta_j]$$

where  $n_d = 2$  denotes the dimension of the surveillance region  $\mathcal{Q}$  and  $\delta_j$  is the maximum distance for each agent to move in  $x$  and  $y$  directions.

However, optimization on  $\ln p(Y_{\leq k+1} | \Psi)$  in Equation (6) and  $J(\tilde{q}, \hat{\Psi}_k)$  in Equation (8) or (9) can be numerically costly due to the increasing size of  $\Sigma_{Y_{\leq k}}$ . One way to deal with this problem is to use a truncated data set

$$Y_{k-\delta \leq \cdot \leq k} := \text{col}(y(s_j(t_m), t_m) | m \in \mathbb{Z}, j \in \mathcal{I}, k - \delta \leq m \leq k)$$

instead of using  $Y_{\leq k}$ . In addition, this approach based on the truncated observations can be viewed as a strategy to deal with a slowly time-varying parameter vector  $\Psi$ , which will further investigated in Section 6.2.

The overall protocol for the sensor network is summarized as in Table 1.

**Table 1.** An adaptive sampling strategy for mobile sensor networks.

- 
1. Learning: At time  $t_k$ , the sensor network updates  $\hat{\Psi}_k$  using a MAP estimate Equation (4) for a data set  $Y_{\leq k}$ . Start this MAP optimization with the initial point  $\hat{\Psi}_{k-1}$ .
  2. Prediction: For given  $Y_{\leq k}$  and  $\hat{\Psi}_k$ , agents can compute prediction at any point and time using Equation (3), i.e.,  $p(z(s, t) | Y_{\leq k}; \hat{\Psi}_k)$ .
  3. Sampling: Based on  $\{\hat{\Psi}_k, Y_{\leq k}\}$ , the sensor network computes the control Equation (10) in order to maximize  $J(\tilde{q}, \hat{\Psi}_k)$ . Update the positions of agents accordingly and collect measurements at time  $t_{k+1}$ .
  4. Repeat the steps 1–3 until  $\Psi$  converges.
- 

## 6. Simulation Results

In this section, we evaluate the proposed approach for a spatio-temporal Gaussian process (Section 6.1) and an advection-diffusion process (Section 6.3). For both cases, we compare the simulation results using the proposed optimal sampling strategy with results using a benchmark random sampling strategy. In this random sampling strategy, each agent was initially randomly deployed in the surveillance region. At each time step, the next sampling position for agent  $i$  is generated randomly with the same mobility constraint, viz. a random position within a square region with length 2 centered at the current position  $q_i$ . For fair comparison, the same values were used for all other conditions. In Section 6.2, our approach based on truncated observations has been applied to a Gaussian process with a time-varying covariance function to demonstrate the adaptability of the proposed scheme.

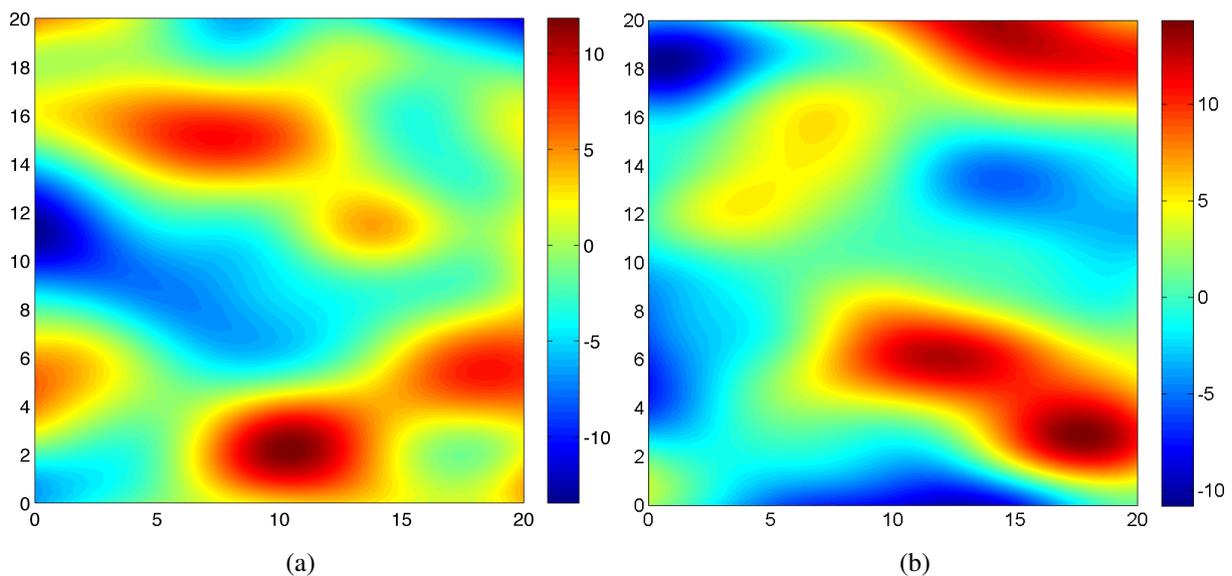
### 6.1. A Spatio-Temporal Gaussian Process

We apply our approach to a spatio-temporal Gaussian process. The Gaussian process was numerically generated for the simulation [11]. The hyperparameters used in the simulation were chosen such that  $\Psi = [\sigma_f \ \sigma_x \ \sigma_y \ \sigma_t \ \sigma_w]^T = [5 \ 4 \ 2 \ 8 \ 0.5]^T$ . Two snap shots of the realized Gaussian random field at time  $t = 1$  and  $t = 20$  are shown in Figure 1. In this case,  $N_s = 5$  mobile sensing agents were initialized at random positions in a surveillance region  $\mathcal{Q} = [0 \ 20] \times [0 \ 20]$ . The initial values for the algorithm were given to be  $\Psi^{(0)} = [1 \ 10 \ 10 \ 1 \ 0.1]^T$ . A prior of the hyperparameter vector has been selected as

$$p(\Psi) = p(\sigma_f)p(\sigma_x)p(\sigma_y)p(\sigma_t)p(\sigma_w)$$

where  $p(\sigma_f) = p(\sigma_x) = p(\sigma_y) = p(\sigma_t) = \Gamma(5, 2)$ , and  $p(\sigma_w) = \Gamma(5, 0.2)$ .  $\Gamma(a, b)$  is a Gamma distribution with mean  $ab$  and variance  $ab^2$  in which all possible values are positive. The gradient method was used to find the MAP estimate of the hyperparameter vector.

**Figure 1.** Snap shots of the realized Gaussian process at (a)  $t = 1$  and (b)  $t = 20$ .



For simplicity, we assumed that the global basis is the same as the model basis. We considered a situation where at each time, measurements of agents are transmitted to a leader (or a central station) that uses our Gaussian learning algorithm and sends optimal control back to individual agents for next iteration to improve the quality of the estimated covariance function. The maximum distance for agents to move in one time step was chosen to be 1 for both  $x$  and  $y$  directions. The A-optimality criterion was used for optimal sampling.

For both proposed and random strategies, Monte Carlo simulations were run for 100 times and the statistical results are shown in Figure 2. The estimates of the hyperparameters (shown in circles and error bars) tend to converge to the true values (shown in dotted lines) for both strategies. As can be seen, the proposed scheme (Figure 2(a)) outperforms the random strategy (Figure 2(b)) in terms of the A-optimality criterion.

**Figure 2.** Monte Carlo simulation results (100 runs) for a spatio-temporal Gaussian process using (a) the random sampling strategy, and (b) the adaptive sampling strategy. The estimated hyperparameters are shown in blue circles with error-bars. The true hyperparameters that used for generating the process are shown in red dashed lines.

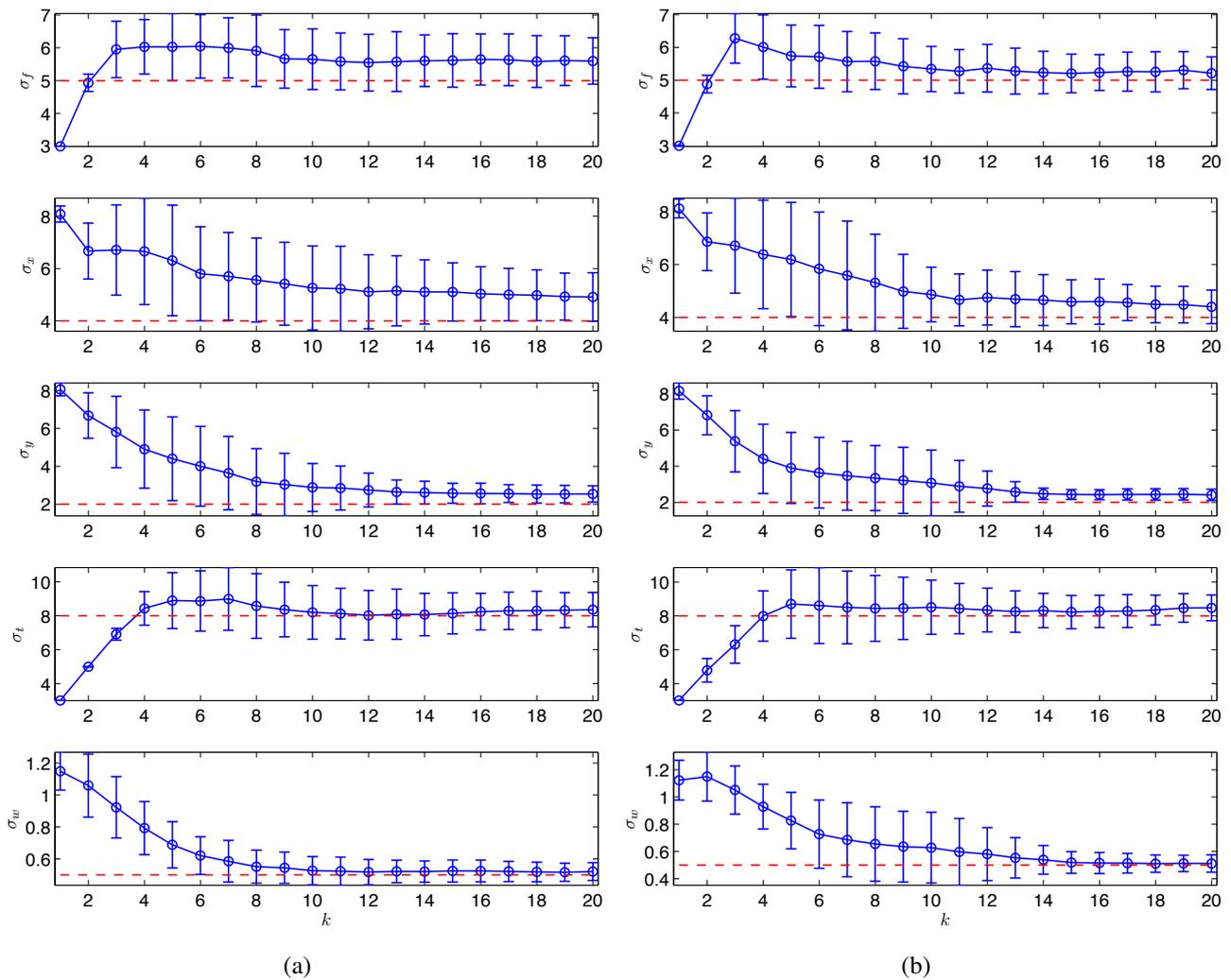
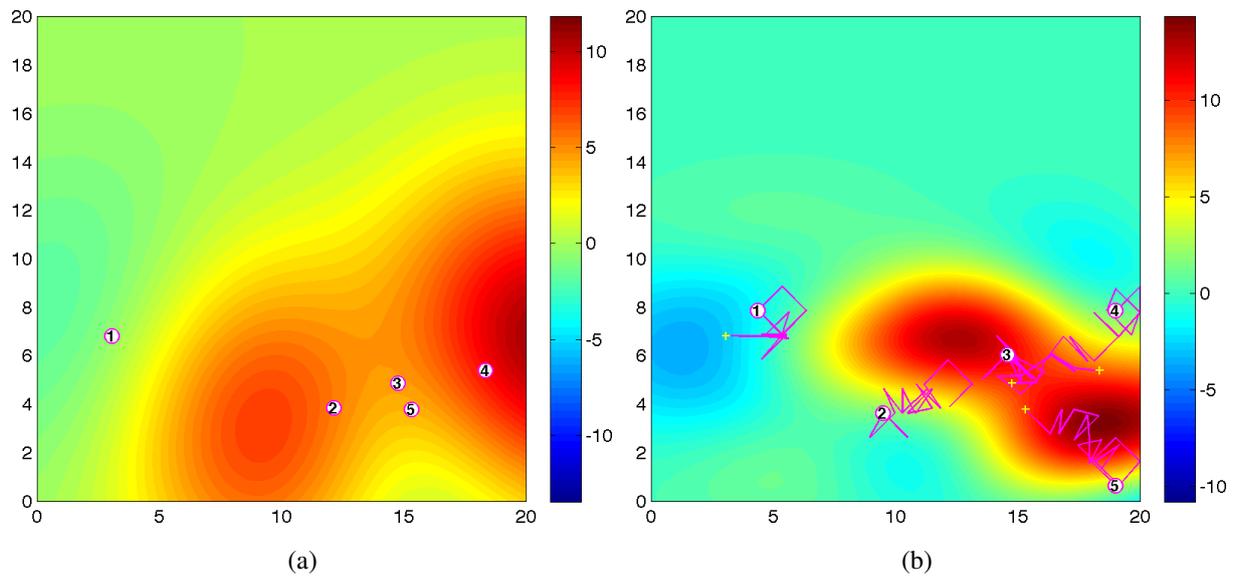


Figure 3 shows the predicted field along with agents' trajectories at time  $t = 1$  and  $t = 20$  for one trial. As shown in Figure 1(a) and Figure 3(a), at time  $t = 1$ , the predicted field is far from the true field due to the inaccurate hyperparameters estimation and small number of observations. As time increases, the predicted field will be closer to the true field due to the improved quality of the estimated the covariance function and the cumulative observations. As expected, at time  $t = 20$ , the quality of the predicted field is very well near the sampled positions as shown in Figure 3(b). With 100 observations, the running time is around 30s using Matlab, R2008a (MathWorks) in a PC (2.4 GHz Dual-Core Processor). No attempt has been made to optimize the code. After converging to a good estimate of  $\Psi$ , agents can switch to a decentralized configuration and collect samples for other goals such as peak tracking and prediction of the process [6,16,17].

**Figure 3.** The predicted fields along with agents' trajectories at (a)  $t = 1$  and (b)  $t = 20$ .

### 6.2. Time-Varying Covariance Functions

To illustrate the adaptability of the proposed strategy to time-varying covariance functions, we introduce a Gaussian process defined by the following covariance function. The time-varying covariance function is modeled by a time-varying weighted sum of two known covariance functions  $\mathcal{K}_1(\cdot)$  and  $\mathcal{K}_2(\cdot)$  such as

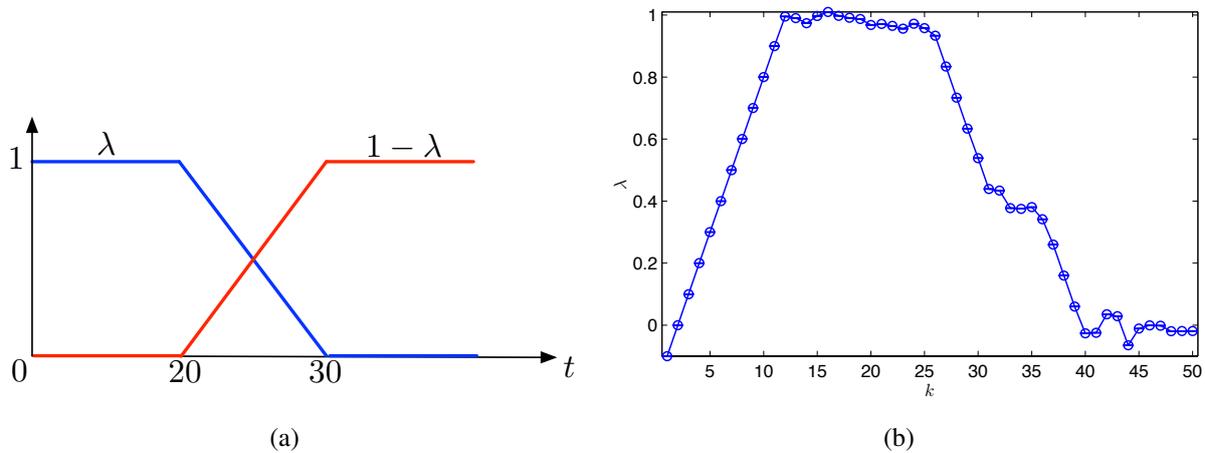
$$\mathcal{K}(\cdot) = \lambda(t)\mathcal{K}_1(\cdot) + (1 - \lambda(t))\mathcal{K}_2(\cdot) \quad (12)$$

where  $\lambda(t) \in [0, 1]$  is a time-varying weight factor that needs to be estimated. In the simulation study,  $\mathcal{K}_1(\cdot)$  is constructed with  $\sigma_f = 1$ ,  $\sigma_x = 0.2$ ,  $\sigma_y = 0.1$ ,  $\sigma_t = 8$ , and  $\sigma_w = 0.1$ ; and  $\mathcal{K}_2(\cdot)$  is with  $\sigma_f = 1$ ,  $\sigma_x = 0.1$ ,  $\sigma_y = 0.2$ ,  $\sigma_t = 8$ , and  $\sigma_w = 0.1$ . This Gaussian process defined in (12) with these particular  $\mathcal{K}_1$  and  $\mathcal{K}_2$  effectively models hyperparameter changes in  $x$  and  $y$  directions.

To improve the adaptability, the mobile sensor network uses only observations sampled during the last 20 iterations for estimating  $\lambda(t)$  online. The true  $\lambda(t)$  and the estimated  $\lambda(t)$  are shown in Figure 4(a,b), respectively. From Figure 4, it is clear that the weighting factor  $\lambda(t)$  can be estimated accurately after some delay about 5–8 iterations. The delay is due to using the truncated observations that contain past observations since the time-varying covariance function changes continuously in time.

### 6.3. Fitting a Gaussian Process to an Advection-Diffusion Process

We apply our approach to a spatio-temporal process generated by physical phenomena (advection and diffusion). This work can be viewed as a statistical modeling of a physical process, *i.e.*, as an effort to fit a Gaussian process to a physical advection-diffusion process in practice. The advection-diffusion model developed in [26] was used to generate the experimental data numerically. An instantaneous release of *Qkg* of gas occurs at a location  $(x_0, y_0, z_0)$ . This is then spread by the wind with mean velocity  $u = [u_x \ 0 \ 0]^T$ . Assuming that all measurements are recorded at a level  $z = 0$ , and the release

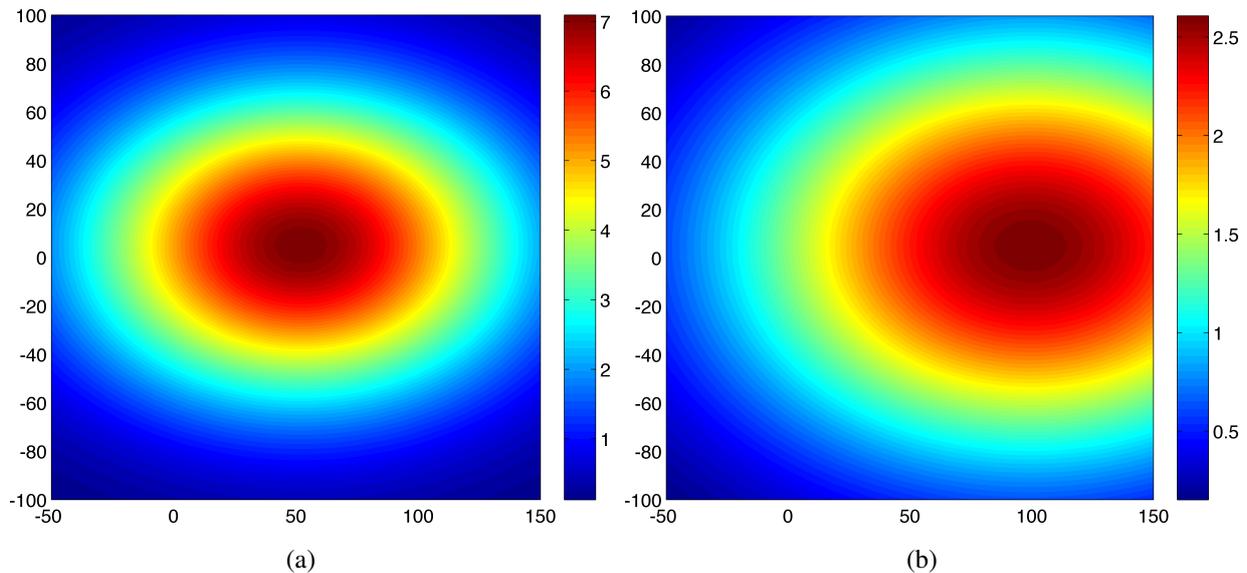
**Figure 4.** (a) The weighting factor  $\lambda(t)$  and (b) the estimated  $\lambda(t)$ .**Table 2.** Parameters used in simulation.

Parameter	Notation	Unit	Value
Number of agents	$N_s$	-	5
Sampling time	$t_s$	min	5
Initial time	$t_0$	min	100
Gas release mass	$Q$	kg	$10^6$
Wind velocity in $x$ axis	$u_x$	m/min	0.5
Eddy diffusivity in $x$ axis	$K_x$	$m^2/min$	20
Eddy diffusivity in $y$ axis	$K_y$	$m^2/min$	10
Eddy diffusivity in $z$ axis	$K_z$	$m^2/min$	0.2
Location of explosion	$x_0$	m	2
Location of explosion	$y_0$	m	5
Location of explosion	$z_0$	m	0
Sensor noise level	$\sigma_w$	$kg/m^3$	0.1

occurs at a ground level (*i.e.*,  $z_0 = 0$ ), the concentration  $C$  at an arbitrary location  $(x, y, 0)$  and time  $t$  is described by the following analytical solution [27]:

$$C(x, y, 0, t) = \frac{Q \exp\left(-\frac{(\Delta x - u\Delta t)^2}{4K_x\Delta t} - \frac{\Delta y^2}{4K_y\Delta t}\right)}{4\pi^{\frac{3}{2}}(K_x K_y K_z)^{\frac{1}{2}}(\Delta t)^{\frac{3}{2}}} \quad (13)$$

where  $\Delta x = x - x_0$ ,  $\Delta y = y - y_0$ , and  $\Delta t = 5(t - 1) + t_0$ . The parameters used in the simulation study are shown in Table 2. Notice that this process generates an anisotropic concentration field with parameters  $K_x = 20m^2/min$  and  $K_y = 10m^2/min$  as in Table 2. The fields at time  $t = 1$  and  $t = 20$  are shown in Figure 5. Notice the center of the concentration moved. In this case,  $N_s = 5$  mobile sensing agents were initialized at random positions in a surveillance region  $\mathcal{Q} = [-50 \ 150] \times [-100 \ 100]$ .

**Figure 5.** Snap shots of the advection-diffusion process at (a)  $t = 1$  and (b)  $t = 20$ .

The initial values for the algorithm was chosen to be  $\Psi^{(0)} = [100 \ 100 \ 100]^T$  where we assumed  $\sigma_f = 1$  and  $\sigma_w = 0.1$ . For this application, we did not assume any prior knowledge about the covariance function. Hence, the MAP estimator was the same as the ML estimator. The gradient method was used to find the ML estimate.

We again assumed that the global basis is the same as the model basis and assumed all agents have the same level of measurement noises for simplicity. In our simulation study, agents start sampling at  $t_0 = 100min$  and take measurements at time  $t_k$  with a sampling time of  $t_s = 5min$  as in Table 2.

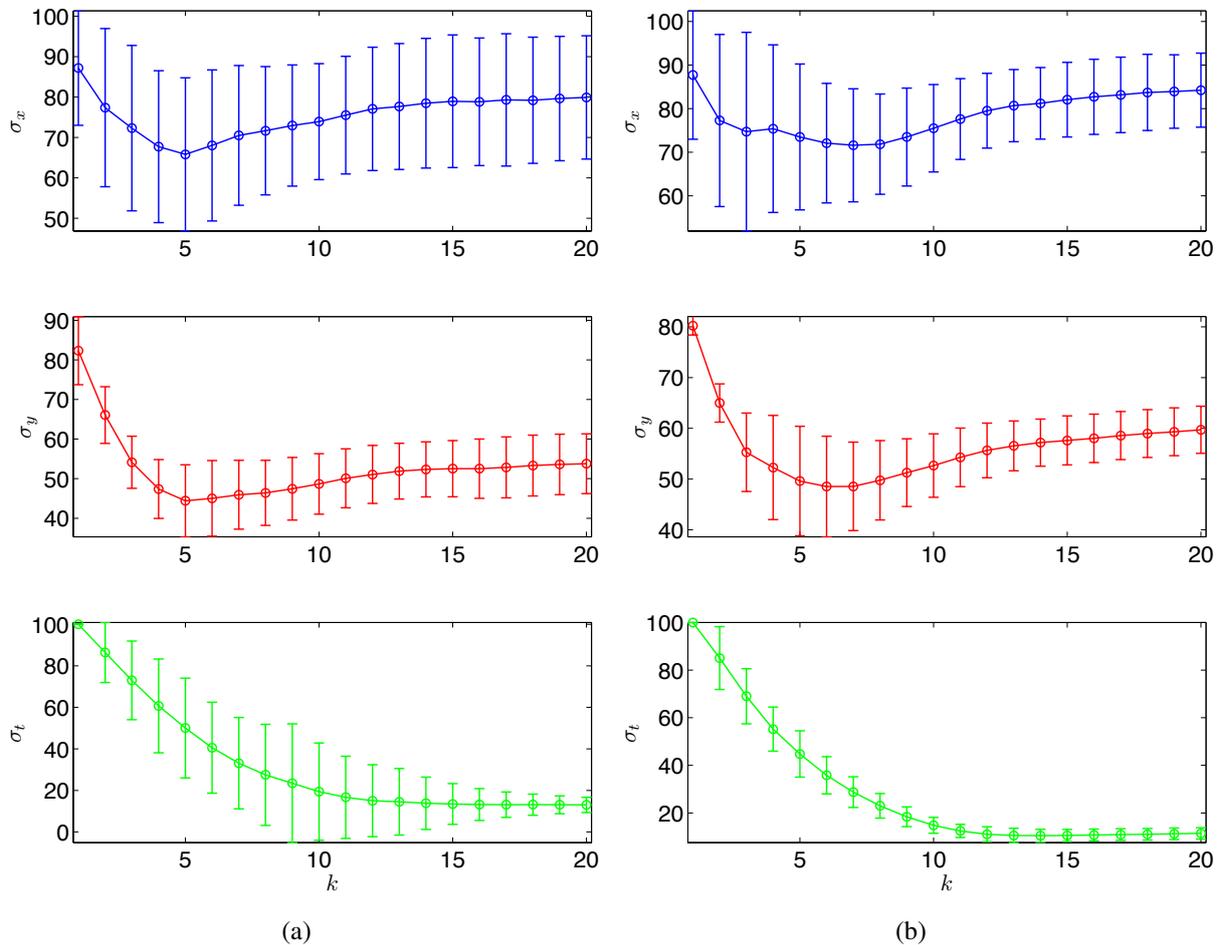
Monte Carlo simulations were run for 100 times, and Figure 6 shows the estimated  $\sigma_x$ ,  $\sigma_y$ , and  $\sigma_t$  with (a) the random sampling strategy and (b) the optimal sampling strategy, respectively. With 100 observations, the running time at each time step is around 20s using Matlab, R2008a (MathWorks) in a PC (2.4 GHz Dual-Core Processor). No attempt has been made to optimize the code. As can be seen in Figure 6, the estimates of the hyperparameters tend to converge to similar values for both strategies. Clearly, the proposed strategy outperforms the random sampling strategy in terms of the estimation error variance.

## 7. Summary

In this paper, we presented a novel class of self-organizing sensing agents that learn an anisotropic, spatio-temporal Gaussian process using noisy measurements and move in order to improve the quality of the estimated covariance function. The MAP estimator was used to estimate the hyperparameters in the unknown covariance function and the prediction of the field of interest was obtained based on the MAP estimates. An optimal navigation strategy was proposed to minimize the information-theoretic cost function of the Fisher Information Matrix for the estimated hyperparameters. The proposed scheme was applied to both a spatio-temporal Gaussian process and a true advection-diffusion field. Simulation study indicated the effectiveness of the proposed scheme and the adaptability to time-varying covariance

functions. The trade-off between a precise estimation and computational efficiency using truncated observations will be studied in the future work.

**Figure 6.** Simulation results (100 runs) for a advection-diffusion process. The estimated hyperparameters with (a) random sampling and (b) optimal sampling.



## Acknowledgment

This work has been supported by the National Science Foundation through CAREER Award CMMI-0846547. This support is gratefully acknowledged.

## References

1. Lynch, K.M.; Schwartz, I.B.; Yang, P.; Freeman, R.A. Decentralized environmental modeling by mobile sensor networks. *IEEE Trans. Robot.* **2008**, *24*, 710–724.
2. Tanner, H.G.; Jadbabaie, A.; Pappas, G.J. *Stability of Flocking Motion*; Technical Report MS-CIS-03-03, The GRASP Laboratory, University Pennsylvania, Philadelphia, PA, USA, May 2003.
3. Olfati-Saber, R. Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Trans. Automat. Contr.* **2006**, *51*, 401–420.

4. Ren, W.; Beard, R.W. Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Trans. Automat. Contr.* **2005**, *50*, 655–661.
5. Leonard, N.E.; Paley, D.A.; Lekien, F.; Sepulchre, R.; Fratantoni, D.M.; Davis, R.E. Collective motion, sensor networks, and ocean sampling. *Proc. IEEE* **2007**, *95*, 48–74.
6. Choi, J.; Oh, S.; Horowitz, R. Distributed learning and cooperative control for multi-agent systems. *Automatica* **2009**, *45*, 2802–2814.
7. Cressie, N. Kriging nonstationary data. *J. Ame. Statist. Assn.* **1986**, *81*, 625–634.
8. Cressie, N. *Statistics for Spatial Data*; A Wiley-Interscience Publication, John Wiley and Sons, Inc.: Chichester, West Sussex, UK, 1991.
9. Gibbs, M.; MacKay, D.J.C. Efficient implementation of Gaussian processes. Available online: <http://www.cs.toronto.edu/mackay/gpros.ps.gz.Preprint/> (accessed on 3 March 2011).
10. Mackay, D.J.C. Introduction to Gaussian Processes. In *NATO ASI Series. Series F: Computer and System Sciences*; Springer-Verlag: Heidelberg, Germany, 1998; pp. 133–165.
11. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2006.
12. Krause, A.; Guestrin, C.; Gupta, A.; Kleinberg, J. Near-optimal sensor placements: Maximizing information while minimizing communication cost. In *Proceedings of the 5th International Symposium on Information Processing in Sensor Networks (IPSN)*, Nashville, TN, USA, April 2006.
13. Krause, A.; Singh, A.; Guestrin, C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.* **2008**, *9*, 235–284.
14. Cortés, J. Distributed Kriged Kalman filter for spatial estimation. *IEEE Trans. Automat. Contr.* **2010**, *54*, 2816–2827.
15. Graham, R.; Cortés, J. Cooperative adaptive sampling of random fields with partially known covariance. *Int. J. Robust Nonlinear Contr.* **2009**, *1*, 1–2.
16. Choi, J.; Lee, J.; Oh, S. Swarm intelligence for achieving the global maximum using spatio-temporal Gaussian processes. In *Proceedings of the 27th American Control Conference (ACC)*, Seattle, WA, USA, June 2008.
17. Choi, J.; Lee, J.; Oh, S. Biologically-inspired navigation strategies for swarm intelligence using spatial Gaussian processes. In *Proceedings of the 17th International Federation of Automatic Control (IFAC) World Congress*, Seoul, South Korea, July 2008.
18. Nott, D.J.; Dunsmuir, W.T.M. Estimation of nonstationary spatial covariance structure. *Biometrika* **2002**, *89*, 819–829.
19. Kay, S.M. *Fundamentals of Statistical Signal Processing: Estimation Theory*; Prentice Hall, Inc.: Upper Saddle River, NJ, USA, 1993.
20. Lagarias, J.; Reeds, J.; Wright, M.; Wright, P. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optimization* **1999**, *9*, 112–147.
21. Mandic, M.; Franzoli, E. Efficient sensor coverage for acoustic localization. In *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, December 2007.
22. Martínez, S.; Bullo, F. Optimal sensor placement and motion coordination for target tracking. *Automatica* **2006**, *42*, 661–668.

23. Nagamune, R.; Choi, J. Parameter reduction in estimated model sets for robust control. *J. Dynam. Syst. Meas. Contr.* **2010**, *132*, 021002.
24. Pukelsheimi, F. *Optimal Design of Experiments*; Wiley: New York, NY, USA, 1993.
25. Emery, A.F.; Nenarokomov, A.V. Optimal experiment design. *Meas. Sci. Technol.* **1998**, *9*, 864–76.
26. Kathirgamanathan, P.; McKibbin, R. Source term estimation of pollution from an instantaneous point source. *Res. Lett. Infor. Math. Sci.* **2002**, *3*, 59–67.
27. Christopoulos, V.N.; Roumeliotis, S. Adaptive sensing for instantaneous gas release parameter estimation. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, Barcelona, Spain, 2005.

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>.)