In this paper, we introduce a family of spatio-temporal Gaussian processes specified by a class of covariance functions. Nonparametric prediction based on truncated observations is proposed for mobile sensor networks with limited memory and computational power. We show that there is a trade-off between precision and efficiency when prediction based on truncated observations is used. Next, we propose both centralized and distributed navigation strategies for mobile sensor networks to move in order to reduce prediction error variances at positions of interest. Simulation results demonstrate the effectiveness of the proposed schemes.

1 INTRODUCTION

In recent years, due to drastic global climate changes, it is necessary to monitor the changing ecosystems over vast regions in lands, oceans, and lakes. Hence, it is necessary to develop autonomous robotic systems that can perform a series of tasks such as estimation, prediction, monitoring, tracing and removal of a scalar field of interest undergoing often complex transport phenomena\(^1\).

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 [1–5]. 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 the mobile agents can be designed in order to perform the optimal sampling of the field of interest. Mobile sensor networks that optimize ocean sampling performance defined in terms of uncertainty in a model estimate of a sampled field were developed in [6]. However, this approach optimized the collective patterns of mobile agents parameterized by a restricted number of parameters rather than optimizing individual trajectories. Swarming agents with a gradient climbing strategy for tracking peaks of a field of interest using kernel regression were proposed in [7]. In general, we design the mobility of sensing agents to find the most informative locations to make observations for a spatio-temporal phenomenon [8].

To find these locations that predict the phenomenon best, one needs a model of the spatio-temporal phenomenon. In our approaches, we rely 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 [9–12]. A Gaussian process is a natural generalization of the Gaussian probability distribution. It generalizes a 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 in [8, 13]. A distributed Kriged Kalman filter for spatial estimation based on mobile sensor networks is developed in [14]. Multi-agent systems that are versatile for various tasks by ex-

\(^1\)Common examples are diffusion, convection, and advection.
exploiting predictive posterior statistics of Gaussian processes were developed in [15, 16].

The motivation of our work is twofold. First, the main reason why the nonparametric prediction using Gaussian processes is not popular for resource-constrained multi-agent systems is the fact that the optimal prediction must use all cumulatively measured values [10, 11]. In this case, a robot needs to compute the inverse of the covariance matrix whose size grows at each iteration. The robot will then run out of memory quickly. Therefore, it is necessary to develop a class of prediction algorithms using spatio-temporal Gaussian processes under a fixed memory size. The space-time Kalman filter model proposed in [17] and utilized in [16] partially solved this problem by assuming no correlation between random values of a Gaussian process with two different time indices. However, the value of a temporal mean field at a point (realized by a stable linear system) consists of a linear sum of colored white noises, and transient responses that converge to zero values exponentially fast [16], which can not represent a wide range of spatio-temporal, physical phenomena. Hence, we need to develop a family of spatio-temporal Gaussian processes in which the prediction can be based on truncated samples (e.g. the last n observations). The second motivation is as follows. As pointed out in [8], a typical sensor placement technique [18] that puts sensors at the locations where the entropy is high tends to place sensors along the borders of the area of interest. The sensor placement algorithm proposed in [8], which maximizes the mutual information between the measurement locations and the rest of the space for the case of no measurement noise. When the observations are corrupted by noise, and so the area of interest is the entire surveillance space, this approach reduces to the sensor placement problem with the entropy criterion. Therefore, it is important to revisit the issue and develop the optimal sampling strategy for sensor networks to minimize a cost function, which has more physical implications such as the prediction error variances of a region of interest.

The contribution of this paper is as follows. First we introduce a class of covariance functions used in [19] and show that the quality of prediction based on truncated observations does not deteriorate much as compared to that of prediction based on all cumulative data. Hence, a good trade-off between precision and efficiency to predict the Gaussian process can be studied. Next, for the introduced Gaussian process, we propose both centralized and distributed navigation strategies for mobile sensor networks to move in order to reduce the prediction error variances at positions of interest. For the latter part, we start with optimal coordination for the entropy criterion to motivate the coordination by minimizing the prediction error variances at positions of interest. In this paper, positions of interest will be referred to as target positions. The introduction of target positions, which can be arbitrarily specified by a user, provides a flexible way to define a geometrical shape of a subregion of interest in a surveillance region. In simulation study, for instance, we will show how to define a circular ring shaped subregion using a set of target positions and the resulting navigation behaviors of agents for the specified target positions.

This paper is organized as follows. In Section 2, we briefly review the mobile sensor network model and the notation related to a graph. A nonparametric approach to predict a field of interest based on truncated observations is presented and analyzed in Section 3. In Section 4, we introduce both centralized and distributed navigation strategies to improve the prediction quality at target positions. Simulation results are shown in Section 5 to illustrate the usefulness of our approaches.

The standard notation will be used in the paper. Let \( \mathbb{R}, \mathbb{R}_{\geq 0} \), denote, respectively, the set of real, non-negative real numbers. The positive definiteness of a matrix A is denoted by \( A > 0 \). \( \mathbb{E} \) denotes the expectation operator. Other notation will be explained in due course.

2 MOBILE SENSOR NETWORKS

First, we explain the mobile sensing network and sensor models used in this paper. Let \( N \) be the number of sensing agents distributed over the surveillance region \( Q \in \mathbb{R}^2 \). Assume that \( Q \) is a compact set. The identity of each agent is indexed by \( i := \{1, 2, \cdots, N\} \). Let \( q_i(t) \in Q \) be the position 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 the sensor noise \( w_i(t) \), at its position \( q_i(t) \) and some measurement time \( t \in T := \{t_1, t_2, \cdots\} \), i.e.,

\[
y(q_i(t), t) := z(q_i(t), t) + w_i(t). \]

The communication network of mobile agents can be represented by a graph with edges. Let \( G(t) := (I, E(t)) \) be an undirected communication graph such that an edge \( (i, j) \in E(t) \) if and only if agent \( i \) can communicate with agent \( j \neq i \). We define the neighborhood of agent \( i \) at time \( t \) by \( N_i(t) := \{j : (i, j) \in E(t), i \in I\} \). 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 PREDICTION BASED ON TRUNCATED OBSERVATIONS

In our previous work [19], for a spatio-temporal Gaussian process \( z(s, t) \), we introduced the following generalized covariance function \( \mathcal{K}(s, t, s_3, t_3; \nabla) \) with a hyperparameter vector \( \nabla := [\sigma_f \sigma_x \sigma_y \sigma_t]^T \):

\[
\mathcal{K}(s, t, s_3, t_3; \nabla) = \sigma_f^2 \exp \left( -\frac{1}{2} \sum_{l \notin \{x, y\}} \frac{(s - s_3)^2}{\sigma^2_l} \right) \exp \left( -\frac{1}{2} \frac{(t - t_3)^2}{\sigma^2_t} \right), \tag{1}
\]

where \( s, s_3 \in Q, t, t_3 \in T \). \( \{\sigma_x, \sigma_y\} \) and \( \sigma_t \) are kernel bandwidths for space and time, respectively. (1) shows that points close in the measurement space and time indices are strongly correlated and produce similar values. A spatially isotropic version of the covariance function in (1) has been used in [6].

\[\]
Suppose at time $t_k \in T$, agent $i$ collects measurements \( \{y(q_j(t_k), t_k) | j \in \bar{N}_i(t_k)\} \) from itself and its neighbors, denoted as $Y_k$. The measurements $y(q_j(t_k), t_k) = z(q_j(t_k), t_k) + w_j(t_k)$ are taken at different positions $q_j(t_k) \in Q$ and time $t_k \in T$. The measurements are corrupted with communication and measurement noises represented by the Gaussian white noise $w_j \sim \mathcal{N}(0, \sigma^2_w)$.

We now consider the truncation (or windowing) of the observed time series data to limit the size of the covariance matrix for reducing the computational cost. Assume that up to time $t_k$, agent $i$ keeps cumulative measurements that have been collected from time $t_{k-\eta_1}$ to $t_k$, i.e., $Y_k^\eta := [Y_{t_{k-\eta_1}}^T \cdots Y_{t_k}^T]^T$, where $\eta \in \Omega := \{1, \cdots, k\}$ is the truncation size. The conditional distribution of $z(s,t)$ given $Y_k^\eta$ is obtained by

\[
z(s,t|k,\eta) := z(s,t)|Y_k^\eta \sim \mathcal{N}(\hat{z}(s,t|k,\eta), \sigma^2(s,t|k,\eta)),
\]

where $\hat{z}(s,t|k,\eta) = E\{z(s,t|k,\eta)\}$ is the prediction of $z(s,t)$ at location $s$ and time $t$ given

\[
\hat{z}(s,t|k,\eta) = \Sigma_{zY_k^\eta} \Sigma_{y_k^\eta} \eta_k^\eta,
\]

and its prediction error variance is

\[
\sigma^2(s,t|k,\eta) = \Sigma_{z} - \Sigma_{zY_k^\eta} \Sigma_{y_k^\eta} \eta_k^\eta. \tag{3}
\]

$\Sigma_{z}$ is the covariance of $z(s,t)$, obtained by $\Sigma_{z} = \mathcal{K}(s,t,s,t;\Psi)$. \(\Sigma_{zY_k^\eta} \Sigma_{y_k^\eta} \eta_k^\eta\) is the covariance matrix between $z$ and $Y_k^\eta$, obtained by \(\Sigma_{zY_k^\eta} \Sigma_{y_k^\eta} \eta_k^\eta = \mathcal{K}(s,t,s',t';\Psi)\), where $s' \in Q$, and $t' \in T$ are the position and the time corresponding to the $j$-th element of $Y_k^\eta$. \(\Sigma_{y_k^\eta} \eta_k^\eta\) is the covariance matrix of $Y_k^\eta$ obtained by \(\Sigma_{y_k^\eta} \eta_k^\eta = \mathcal{K}(s,t',s',t';\Psi) + \sigma^2_\eta \delta_{tj}\), where $\delta_{tj}$ is the Dirac delta function.

The following two propositions justify the usage of prediction based on truncated observations for the spatio-temporal Gaussian process with the covariance function in (1).

**Proposition 1.** Let $\eta$, and $k$ be the window size, and the current time index, respectively. \(\Omega\) denotes the set given by \(\Omega := \{1, \cdots, k\}\). Consider the function of prediction error variance $f : \Omega \rightarrow \mathbb{R}_{\geq 0}$ as a function of $\eta$ defined by $f(\eta) := \sigma^2(s,t|k,\eta)$. Then, $f(\eta_1) > f(\eta_2)$ for any $\eta_1 > \eta_2$, i.e., $f$ is a monotonically decreasing function with respect to $\eta \in \Omega$.

**Proof.** From (3), we have

\[
f(\eta_2) - f(\eta_1) = \sigma^2(s,t|k,\eta_2) - \sigma^2(s,t|k,\eta_1)
\]

\[
= \Sigma_{zY_k^\eta} \Sigma_{y_k^\eta} \eta_k^\eta \Sigma_{y_k^\eta} \eta_k^\eta - \Sigma_{zY_k^\eta} \Sigma_{y_k^\eta} \eta_k^\eta \Sigma_{y_k^\eta} \eta_k^\eta.
\]

Let $Y_e = Y_{t_{k-\eta_2}}^\eta$ and $Y_d = Y_{t_{k-\eta_1}}^\eta := [Y_{t_{k-\eta_2}}^T \cdots Y_{t_{k-\eta_1}}^T]^T$. Notice that $Y_k^\eta = [Y_e^T Y_d^T]^T$ since $\eta_1 > \eta_2$. The covariance matrix of $Y_k^\eta$ is then defined as

\[
\Sigma_{y_k^\eta} \eta_k^\eta = \begin{bmatrix}
\Sigma_{Y_e} \Sigma_{Y_d} \\
\Sigma_{Y_d} \Sigma_{Y_e}
\end{bmatrix},
\]

where $\Sigma_{Y_e}, \Sigma_{Y_d},$ and $\Sigma_{Y_e}$ are defined as in (3). Let $\Sigma$ be the positive definite prediction error covariance matrix when using $Y_e$ to predict $Y_d$, i.e., $\Sigma := \Sigma_{Y_d} - \Sigma_{Y_e} \Sigma_{Y_d}^\dagger \Sigma_{Y_e} > 0$. The inverse of the covariance matrix $\Sigma_{y_k^\eta} \eta_k^\eta$ then becomes

\[
\Sigma_{y_k^\eta} \eta_k^\eta = \begin{bmatrix}
\Sigma_{Y_e}^{-1} + \Sigma_{Y_d}^{-1} \Sigma_{Y_e} \Sigma_{Y_d}^\dagger \Sigma_{Y_d}^{-1} - \Sigma_{Y_e}^{-1} \Sigma_{Y_e} \Sigma_{Y_d}^\dagger \\
-\Sigma_{Y_d}^{-1} \Sigma_{Y_d} \Sigma_{Y_e}^\dagger \\
\end{bmatrix}.
\]

Similarly, the covariance matrix between the random field $z(s,t)$ and $Y_k^\eta$ is defined as $\Sigma_{zY_k^\eta} \eta_k^\eta = \Sigma_{zY_e} \Sigma_{Y_d}^\dagger$, where $\Sigma_{zY_e}$ and $\Sigma_{zY_d}$ are defined accordingly as in (3). Since $\Sigma^{-1}$ is a positive definite matrix, we have

\[
f(\eta_2) - f(\eta_1)
\]

\[
= [\Sigma_{zY_e} \Sigma_{Y_d}] \begin{bmatrix}
\Sigma_{Y_e}^{-1} + \Sigma_{Y_d}^{-1} \Sigma_{Y_e} \Sigma_{Y_d}^\dagger \Sigma_{Y_d}^{-1} - \Sigma_{Y_e}^{-1} \Sigma_{Y_e} \Sigma_{Y_d}^\dagger \\
-\Sigma_{Y_d}^{-1} \Sigma_{Y_d} \Sigma_{Y_e}^\dagger \\
\end{bmatrix} \begin{bmatrix}
\Sigma_{Y_d} \\
\Sigma_{zY_e}
\end{bmatrix}
\]

\[
= (\Sigma_{zY_d} - \Sigma_{zY_e} \Sigma_{Y_d} \Sigma_{Y_e}^\dagger) \Sigma^{-1} (\Sigma_{Y_e} - \Sigma_{zY_e} \Sigma_{Y_d} \Sigma_{Y_e}^\dagger)^T > 0.
\]

This completes the proof. QED.

For given collective measurements up to time $t_k$, consider the difference between prediction error variances obtained by using truncated measurements and by using all measurements $e(\eta) := f(\eta) - f(k)$. Based on Lemma 1, for a given time series of data and a given level of $\epsilon^*$, we can always find a minimal $\eta^+$ such that $e(\eta^+) < \epsilon^*$. An example is shown in Fig. 1 with a given level of $\epsilon^* = 0.001$ (in a green dashed line). In this case, we find a minimal truncation size ($\eta^+ = 8$ for $\sigma_t = 5$ and $\eta^+ = 14$ for $\sigma_t = 10$) such that $e(\eta^+) < \epsilon^*$. Notice that for a fixed $\epsilon^* = 0.001$ we can take a smaller truncation size $\eta^+ = 8$ for a smaller $\sigma_t = 5$ (Fig. 1) due to the covariance function in (1).

The following proposition shows properties of the introduced Gaussian process and motivates agents to discard the oldest set of observations to maintain a finite number of cumulative observations.

**Proposition 2.** Consider the following simple sampling strategy for a Gaussian process with the covariance function defined in (1). A sensor network is sampling a finite number of noisy measurements at a fixed set of spatially distributed positions in $Q$, at times $t_i$ and $t_j$ with $j > i$. Then we have $\sigma^2(s,t|i,1) > \sigma^2(s,t|j,1), \quad \forall s \in Q, t > \frac{\sigma_t}{2}$. Copyright © 2009 by ASME
Figure 1. The difference in prediction error variances of the truncated and all observations \( e(\eta) := f(\eta) - f(20) \) as a function of the truncation size \( \eta \). The blue solid line represents the error function with \( \sigma_0 = 5 \) and the red dashed line represents the error function with \( \sigma_0 = 10 \).

**Proof.** According to the specific structure of the covariance function in (1), notice that the closer two measurements are made in time, the higher the correlation is. Hence, we have \( \Sigma_{t,j} = c \Sigma_{s,j} \), where \( c > 1 \) and \( z := z(s,t), \forall s \in Q, t > \frac{k+1}{2} \). Also, we have \( \Sigma_{t} = \Sigma_{s} \) by the sampling strategy. From (3), we then obtain

\[
\sigma^2(s,t|j,1) - \sigma^2(s,t|j,1) = \sigma_{t} \Sigma_{s,t}^{-1} \Sigma_{s,t} - \sigma_{t} \Sigma_{s,t}^{-1} \Sigma_{s,t} = (c-1)\Sigma_{t} \Sigma_{s,t}^{-1} \Sigma_{s,t} > 0,
\]

which completes the proof. QED.

4 NAVIGATION STRATEGIES

The goal of the navigation of agents is to improve the quality of prediction at target positions \( \{p_1, p_2, \ldots, p_M\} \). In particular, we allow that the number of target positions \( M \) can be greater than that of agents \( N \), which is often the case in practice.

To use prediction based on truncated observations, motivated by Proposition 2, we propose that each agent discards the oldest set of measurements \( Y_{k-\eta+1} \) taken at time \( t_{k-\eta+1} \) after making the prediction at time \( t_k \). At the next time index \( t_{k+1} \), using the remained observations \( Y_{k-1} \) in the memory along with new measurements at time \( t_{k+1} \), agents will predict \( z(s,t) \) evaluated at target positions \( \{p_1, p_2, \ldots, p_M\} \) and time \( t_{k+1} \). Hence, agents should move to the most informative locations for taking measurements at time \( t_{k+1} \) [8].

Let \( \tilde{Y} \) be the union of the truncated measurements \( Y_{k-\eta+1} \) and the measurements that will be taken at time \( t_{k+1} \), i.e., \( \tilde{Y} := Y_{k+1}^\eta \). A common measure of uncertainty is the conditional entropy [8, 20] of the target positions \( p := \text{col}(p_1, \ldots, p_M) \in Q^M \) at time \( t_{k+1} \) after placing sensors at locations \( \tilde{q} := q(t_{k+1}) = \text{col}(q_1(t_{k+1}), \ldots, q_N(t_{k+1})) \in Q^N \), i.e.,

\[
H(Z(p,t_{k+1})|\bar{Y}(\tilde{q})) := -\int \int f(Z(p,t_{k+1}), \bar{Y}(\tilde{q})) \log_2 f(Z(p,t_{k+1}), \bar{Y}(\tilde{q})) dz d\bar{Y},
\]

where \( Z(p,t_{k+1}) \) is the random field at \( p \) and \( t_{k+1} \). \( f \) is the joint probability density function. The conditional entropy of the multivariate normal distribution can be obtained by

\[
H(Z(q,t_{k+1})|\bar{Y}(\tilde{q})) = \frac{1}{2} \log_2 (2\pi e)^M |\Sigma_e(\tilde{q})|,
\]

where \( \Sigma_e(\tilde{q}) \in \mathbb{R}^{M \times M} \) is the prediction error covariance matrix calculated by

\[
\Sigma_e(\tilde{q}) = \Sigma - \Sigma Z_{\tilde{Y}}(\tilde{q}) \Sigma^{-1}_{\tilde{Y}} \Sigma_{\tilde{Y}} Z_{\tilde{Y}}^T.
\]

A set of good sampling positions at time \( t_{k+1} \) will minimize the conditional entropy [8, 20] in (4) which is equivalent to minimizing the determinant of the prediction error covariance matrix in (5). For instance, we may seek to find

\[
q(t_{k+1}) = \arg \min_{\tilde{q}} H(Z(p,t_{k+1})|\bar{Y}(\tilde{q}))
\]

\[
= \arg \min_{\tilde{q}} |\Sigma_e(\tilde{q})|.
\]

The major drawback of this approach is that, if any eigenvalue of the covariance matrix \( \Sigma_e \) becomes small, the determinant of \( \Sigma_e \) may become small as well. Thus, the outcome of the optimization does not carry a physical sense in practice. In addition, the sensor placement may not be optimal in minimizing prediction error variances at target positions.

4.1 A Centralized Strategy

In this section, we consider the minimization of an upper-bound of the prediction error variances at target positions \( p \). First notice that an upper-bound of the prediction error variance \( \sigma_j^2 \) at \( p_j \) for all \( j \in \{1, \cdots, M\} \) can be obtained by the trace of \( \Sigma_e \), i.e.,

\[
\sigma_j^2(\tilde{q}) < \sum_{j=1}^{M} \sigma_j^2(\tilde{q}) = \text{tr}(\Sigma_e(\tilde{q})),
\]

where \( \text{tr}(A) \) represents the trace of the matrix \( A \). Hence, we consider the trace of \( \Sigma_e \) as the cost function

\[
J_e(\tilde{q}) := \text{tr}(\Sigma_e(\tilde{q})).
\]
In order to reduce the upper-bound of the prediction error variances at target positions \( p \), the sensor network solves the following optimization problem:

\[
q(t_{k+1}) = \arg \min_{\tilde{q}} J_c(\tilde{q}).
\]

A gradient descent algorithm can be used to find a local minimum of \( J_c \) for the prediction at time \( t_{k+1} \).

\[
\frac{dg(t)}{dt} = -k_c \nabla J_c \big|_{\tilde{q} = q(t)} \quad \text{for} \quad k_c > 0 \quad \text{a gain factor and} \quad \nabla g(x) \quad \text{the partial derivative of} \quad g(x) \quad \text{with respect to} \quad x.
\]

where \( \frac{\partial}{\partial \bar{q}_i} \text{tr}(\Sigma) = \sum_{j=1}^{M} \bar{q}_j \frac{\partial \text{vec}(\Sigma)}{\partial \bar{q}_i} = \text{vec}^T(I_M) \text{vec} \left( \frac{\partial \Sigma}{\partial \bar{q}_i} \right) \]

where \( \text{vec}(A) \) is the vectorization of the matrix \( A \), and \( I_M \in \mathbb{R}^{M \times M} \) is the identity matrix.

The size of the truncation size \( \eta \) should be carefully selected according to a trade-off relationship between accuracy and efficiency.

A centralized navigation strategy for the sensor network with the cost function \( J_c \) in (6) is summarized as follows.

1. At time \( t_1 \), each agent samples initial measurements and transmits them to the leader. Also set \( k = 1 \).
2. At time \( t_k \), the leader computes the prediction using (2) based on truncated collective measurements \( Y_k^\eta \).
3. If \( k > \eta \), the leader discards the oldest set of measurements taken at time \( t_{k-\eta+1} \).
4. The leader computes the optimal control using (7) with the remained data \( Y_k^{\eta-1} \) and sends the control commands to other agents so that the sensor network approaches a critical point of \( J_c(\tilde{q}) \) in (6) before time \( t_{k+1} \).
5. Each agent collects new measurements from itself and its neighbors at time \( t_{k+1} \) and transmits them to the leader.
6. Increment \( k \) by 1 and go to step 2.

4.2 A Distributed Strategy

In this section, we propose a distributed strategy. Another upper-bound of \( \sigma_c \) for agent \( i \) can be obtained by pretending that there will be no measurements from neighbors of agent \( i \) at time \( t_{k+1} \).

\[
\sigma_i^2(\tilde{q}_i) < \sum_{j=1}^{M} \sigma_{ij}^2(\tilde{q}_i) = \text{tr}(\Sigma_c(\tilde{q}_i)) = J_c(\tilde{q}_i), \quad \text{for} \quad k > \eta.
\]

where \( \Sigma_c(\tilde{q}_i) = \Sigma_Z - \Sigma_{Z\tilde{q}}^{\eta-1} \Sigma_{\tilde{q}Z}^{\eta-1} \tilde{Y}(\tilde{q}_i) \) and \( \tilde{Y}(\tilde{q}_i) := [Y(\tilde{q}_i, t_{k+1}) (Y_k^{\eta-1})^T] \). For agent \( i \), a gradient descent algorithm can be used to find a local minimum of \( J_d \) for the prediction at time \( t_{k+1} \).

\[
\frac{dg_i(t)}{dt} = -k_d \nabla J_d \big|_{\tilde{q} = q(t)} \quad \text{for} \quad k_d > 0 \quad \text{is a gain factor.}
\]

A distributed navigation strategy for agent \( i \) with the cost function \( J_d \) in (8) is summarized as follows.

1. At time \( t_1 \), agent \( i \) samples initial measurements and transmits them to neighbors. Also set \( k = 1 \).
2. At time \( t_k \), agent \( i \) computes the prediction using (2) based on truncated collective measurements \( Y_k^\eta \).
3. If \( k > \eta \), agent \( i \) discards the oldest set of measurements taken at time \( t_{k-\eta+1} \).
4. Using the control (9) with the remained data \( Y_k^{\eta-1} \), agent \( i \) approaches a critical point of \( J_d(\tilde{q}_i) \) in (8) before time \( t_{k+1} \).
5. Agent \( i \) collects new measurements and receives measurements transmitted from neighbors at time \( t_{k+1} \).
6. Increment \( k \) by 1 and go to step 2.

5 SIMULATION

In this section, we apply our approach to a spatio-temporal Gaussian process with a covariance function in (1). The Gaussian process was numerically generated [12] for the simulation. The hyperparameters used in the simulation was chosen to be \( \Psi = [ \alpha_0, \sigma_y, \sigma_z]^T = [10, 2, 3, 5]^T \).

The surveillance region \( Q \) is given by \( Q = [-5, 5]^2 \). The truncation size \( \eta \) is selected to be 10. The noise level \( \sigma_e = 0.5 \) is used throughout the simulation. In our simulation, nine agents start sampling at \( t_1 = 1s \) and take measurements at time \( t_k \) with a sampling time of 1s. The initial positions of the agents are randomly selected.

5.1 A Centralized Scheme

We consider a situation where a leading 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. At each time, measurements sampled by agents are transmitted to the leader (or the central station) that uses the centralized navigation strategy and sends control commands back to individual agents.

Case 1: First, \( 5 \times 5 \) grid points on \( Q \) are considered as the target positions. Fig. 2 shows that the minimum of the cost function \( J_c \) (in blue stars). At each time iteration, the cost function \( J_c \) which provides an upper-bound of the prediction error variances at target positions is minimized. The worst prediction error variance over the target positions at each iteration is also shown in red triangles in Fig. 2.
Fig. 2. The optimal values of the cost function $J_c$ (in blue stars) and the worst case prediction error variances (in red triangles) achieved by the centralized scheme for case 1.

Fig. 3, and 4 show (a) the true field $z(s,t)$, (b) the predicted field $\hat{z}(s,t|k,\eta)$, and (c) the prediction error variance $\sigma^2(s,t|k,\eta)$ along with the trajectories of agents at different times $t = 1s$ and $t = 20s$. In (c), the target positions, the initial positions, and the current positions are represented by white stars, yellow crosses, and white circles, respectively. At time $t_1$, agents have little information about the field and hence the prediction is far away from the true field, which produces a large prediction error variance in general. As time increases, the prediction becomes close to the true field and the prediction error variances are reduced due to the optimal navigation strategy.

**Case 2:** Now, we consider another case in which 36 target positions (plotted in Fig. 6 as white stars) are evenly distributed on three concentric circles to form a ring shaped subregion of interest. The minimum of the cost function $J_c$ and the worst prediction error variance over these target positions at each iteration are shown in Fig. 5. The prediction error variance along with the trajectories of agents at time $t = 20s$ are shown in Fig. 6. It is shown that four agents are dynamically covering the ring shaped region to minimize an upper-bound for the prediction error variances at the target positions.

### 5.2 A Distributed Scheme

Now, we consider a situation in which the sensor network has a fixed communication graph as shown in Fig. 7. For instance, $\tilde{\mathcal{N}}_i(t) = \{8,9,1,2,3\} \forall t$. At each iteration time $t_k$, agent $i$ collects measurements from itself and its neighbors $\mathcal{N}_i(t_k)$ and makes predictions in a distributed fashion. The distributed strategy is used to navigate itself to move to the next sampling position. To be comparable with the centralized scheme, the same target positions defined by $5 \times 5$ grid points on $Q$ are considered.

Fig. 8 shows the achieved cost function $J_d$ (in blue stars) along with the worst prediction error variance over the target positions (in red triangles) for each agent. The cost function $J_d$ which provides an upper-bound for the prediction error variance is minimized quickly to certain levels. As compared to Fig. 2, the minimum of the cost function $J_d$ is larger than that of the cost function $J_c$ at each iteration due to the distributed scheme.

Fig. 9 shows (a) the true field $z(s,t)$, (b) the predicted field by agent 1 $\hat{z}(s,t|k,\eta)$, and (c) the prediction error variance $\sigma^2(s,t|k,\eta)$ by agent 1 along with the trajectories of agents at time $t = 20s$. In (c), the target positions, the initial positions, and the current positions are represented by white stars, yellow crosses, and white circles, respectively. As compared to Fig. 4, the prediction made by agent 1 is worse and the variance is larger. Moreover, agents using the distributed strategy tend to move further since they pretend not to receive measurements from neighbors at the next iteration.

### 6 CONCLUSION

Based on the covariance function in [19], this paper justified the prediction based on truncated observations by showing that a minimal truncation size can be selected for a given level of truncation error. Centralized and distributed navigation strategies were proposed to minimize upper-bounds for the prediction error variances at target positions that can be arbitrarily chosen by a user. Simulation study showed the behaviors of sensing agents given different target positions under both centralized and distributed schemes. Future work will consider the optimal coordination of the mobile sensor networks subject to energy con-
ACKNOWLEDGMENT

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

REFERENCES

Figure 9. Simulation results at $t = 20s$ obtained by the distributed scheme. (a) True field; (b) Predicted field; (c) Prediction error variance and agents’ trajectories.


