EFFICIENT SPATIAL PREDICTION USING GAUSSIAN MARKOV RANDOM FIELDS UNDER UNCERTAIN LOCALIZATION

Mahdi Jadaliha **Yunfei Xu**
Department of Mechanical Engineering
Michigan State University
East Lansing, MI 48824, USA
E-mails: {jadaliha, xuyunfei}@msu.edu

Jongeun Choi*
Department of Mechanical Engineering
Department of Electrical and Computer Engineering
Michigan State University
East Lansing, MI 48824, USA
E-mail: jchoi@egr.msu.edu

ABSTRACT

In this paper, we develop efficient spatial prediction algorithms using Gaussian Markov random fields (GMRFs) under uncertain localization and sequential observations. We first review a GMRF as a discretized Gaussian process (GP) on a lattice, and justify the usage of maximum *a posteriori* (MAP) estimates of noisy sampling positions in making inferences. We show that the proposed approximation can be viewed as a discrete version of Laplace's approximation for GP regression under localization uncertainty. We then formulate our problem of computing prediction and propose an approximate Bayesian solution, taking into account observations, measurement noise, uncertain hyperparameters, and uncertain localization in a fully Bayesian point of view. In particular, we present an efficient scalable approximation using MAP estimates of noisy sampling positions with a controllable tradeoff between approximation error and complexity. The effectiveness of the proposed algorithms is illustrated using simulated and real-world data.

1 Introduction

Lately, there has been an increasing exploitation of mobile robotic sensors in environmental monitoring [1–5]. Gaussian processes (GPs) (or Gaussian random fields) defined by mean and covariance functions over a continuum space [6,7] have been frequently used for mobile sensor networks to statistically model physical phenomena such as harmful algal blooms, pH, and temperature, e.g., [8–10].

A set of hyperparameters in the covariance function can be estimated by a point estimator such as a maximum likelihood (ML) estimator or a maximum *a posteriori* (MAP) estimator and then it can be used for the prediction [11]. However, the point estimate itself needs to be identified using certain amount of measurements and it does not take into account the uncertainty in the estimated hyperparameters in prediction.

The advantage of a fully Bayesian approach is the capability of incorporating various uncertainties in the model parameters and measurement processes in the prediction [12]. However, the solution often requires Markov Chain Monte Carlo (MCMC) methods, which greatly increases the computational complexity. In [5], a sequential Bayesian prediction algorithm and its distributed version (without resorting to MCMC methods) have been developed to deal with uncertain bandwidths by using a compactly supported covariance function and selecting a subset of collected measurements.

Recently, there have been efforts to fit a computationally efficient Gaussian Markov random field (GMRF) on a discrete lattice to a GP on a continuum space [13–16]. It has been demonstrated that GMRFs with small neighborhood can approximate Gaussian fields surprisingly well [13]. This approximated GMRF and its regression are very attractive for the resource-constrained mobile sensor networks due to its computational efficiency and scalability [17] as compared to the standard GP and its regression. In [18] the authors developed a new class of GPs that builds on a GMRF and derived the predictive statistics for known hyperparameters. On the other hand, [16] developed sequential fully Bayesian prediction algorithms for a GMRF with unknown hyperparameters.

In practice, resource-constrained mobile sensor networks are subject to localization uncertainty. In [19], GP regression under known hyperparameters and uncertain localization has been

*Address all correspondence to this author.

formulated. This has been extended in [20], where sequential fully Bayesian prediction algorithms for a GMRF that can take into account uncertain localization have been developed. However, the resulting computation time for the proposed approach in [20] is still prohibitive for real-time implementation of the robotic sensors. Hence, the objective of this paper is to design efficient prediction algorithms in order to improve the complexity of the scheme developed in [20].

The contributions of the paper are as follows. We first show that an approximated predictive distribution based on MAP estimates of noisy sampling positions can be viewed as a discrete version of Laplace's approximation for GP regression under localization uncertainty in a continuous space (Section 2). We introduce the practical models for the mobile sensor network and the spatio-temporal random field that will be used for our main problems (Section 3). We then formulate our problem of computing prediction and propose an approximate Bayesian solution, taking into account observations, measurement noise, uncertain hyperparameters, and uncertain localization in a fully Bayesian point of view (Section 4). In particular, the proposed solution is scalable and can be efficiently computed using MAP estimates of sampling positions. The effectiveness of the proposed algorithms is illustrated using simulated and real-world data (Section 5).

Standard notation will be used throughout the paper. Let $\mathbb{R}$ and $\mathbb{Z}_{>0}$ denote, respectively, the sets of real and positive integer numbers. The operator of expectation is denoted by E. A random vector x, which has a multivariate normal distribution of mean vector μ and covariance matrix Σ, is denoted by $x \sim \mathcal{N}(\mu, \Sigma)$. For given $G = \{c, d\}$ and $H = \{1, 2\}$, the multiplication between two sets is defined as $H \times G = \{(1,c), (1,d), (2,c), (2,d)\}$. Other notation will be explained in due course.

2 Gaussian Process and Gaussian Markov Random fields

In this section, we first review a GMRF as a discretized GP on a lattice, and justify the usage of MAP estimates of noisy sampling positions in computing the predictive distribution.

Consider a zero-mean GP: $z(q) \sim \mathcal{GP}(0, \Sigma(q, q'))$, where $\Sigma(\cdot,\cdot)$ is the covariance function defined in a continuum space $\mathcal{S}_c$. We discretize the compact domain $\mathcal{S}_c := [0\, x_{max}] \times [0\, y_{max}]$ into n spatial sites $\mathcal{S} := \{s^{[1]}, \cdots, s^{[n]}\} \subset \mathbb{R}^d$, where $n = hx_{max} \times hy_{max}$. h will be chosen such that $n \in \mathbb{Z}_{>0}$. Note that $n \to \infty$ as $h \to \infty$. The collection of realized values of the random field in $\mathcal{S}$ is denoted by $z := (z^{[1]}, \cdots, z^{[n]})^T \in \mathbb{R}^n$, where $z^{[i]} := z(s^{[i]})$.

The prior distribution of z is given by $z \sim \mathcal{N}(0, \Sigma_0)$, and so we have

$$\pi(z) \propto \exp\left(-\frac{1}{2} z^T \Sigma_0^{-1} z\right),$$

where $\Sigma_0 \in \mathbb{R}^{n \times n}$ is the covariance matrix. The i, j-th element of Σ_0 is defined as $\Sigma_0^{[ij]} = \mathrm{Cov}(z^{[i]}, z^{[j]}) = \Sigma(z^{[i]}, z^{[j]})$. The prior distribution of z can be written by a precision matrix $Q_0 = \Sigma_0^{-1}$, i.e., $z \sim \mathcal{N}(0, Q_0^{-1})$. This can be viewed as a discretized version of the GP (or a GMRF) with a precision matrix Q_0 on $\mathcal{S}$. Note that Q_0 of this GMRF is not sparse. However, a sparse version of Q_0, i.e., $\hat{Q}_0$ with local neighborhood that can represent the original GP can be found, for example, making $\hat{Q}_0$ close to Q_0 in some norm [13–15]. This approximate GMRF will be computationally efficient due to the sparsity of $\hat{Q}_0$. For our main problems, we will use a GMRF with a sparse precision matrix that represents a GP precisely (see Section 3.2).

Let us assume that we take N noisy measurements $y = (y^{[1]}, \cdots, y^{[N]})^T \in \mathbb{R}^N$ from corresponding sampling locations $q_c = (q_c^{[1]T}, \cdots, q_c^{[N]T})^T \in \mathcal{S}_c^N$. Here, the measurement model is given by

$$y^{[i]} := y(q_c^{[i]}) = z(q_c^{[i]}) + \varepsilon^{[i]}, \forall i = 1, \cdots, N$$

where $\varepsilon^{[i]} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$ is the measurement noise and is assumed to be independent and identically distributed (i.i.d.).

Using GP regression, the posterior distribution for $z \in \mathbb{R}^n$ is given by

$$z | q_c, y \sim \mathcal{N}(\mu, \Sigma). \tag{1}$$

The predictive mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ can by obtained by

$$\mu = K^T C^{-1} y, \quad \Sigma = \Sigma_0 - K^T C^{-1} K, \tag{2}$$

where the covariance matrices are defined as $K := \mathrm{Cov}(y, z) \in \mathbb{R}^{N \times n}$, $C := \mathrm{Cov}(y, y) \in \mathbb{R}^{N \times N}$, and $\Sigma_0 := \mathrm{Cov}(z, z) \in \mathbb{R}^{n \times n}$.

Now, consider localization uncertainty. The prior distribution for sampling location $q_c^{[i]}$ is given by $\pi(q_c^{[i]} | \tilde{q}_c^{[i]})$, possibly with a compact support in $\mathcal{S}_c$. Then the predictive distribution of z given the measured locations $\tilde{q}_c = (\tilde{q}_c^{[1]T}, \cdots, \tilde{q}_c^{[N]T})^T$ is thus given by

$$\pi(z | \tilde{q}_c, y) = \int_{q \in \mathcal{S}_c} \pi(z | q, y) \pi(q | \tilde{q}_c, y) dq, \tag{3}$$

where $\pi(z | q, y)$ can be obtained in (1). However, the predictive distribution in (3) does not have a closed-form solution and needs to be computed either by MCMC methods or approximation techniques [19].

The following MAP estimate of q_c, i.e.,

$$\hat{q} = \arg\max_{q \in \mathcal{S}_c} \pi(q | \tilde{q}_c, y), \tag{4}$$

can be used to approximate the predictive distribution as $\pi(z | \hat{q}, y)$. In our previous work in [19], we have the following result.

Theorem 2.1. *(see Theorem 4 in [19]) Let $\hat{q}$ be an asymptotic mode of order $O(g^{-1})$ for $\frac{1}{g}\log(\pi(y|q)\pi(q|\tilde{q}_c))$. This mode is the MAP estimator given in (4). Consider the following Laplace approximations by plugging $\hat{q}$ for q_c into (2):*

$$\hat{\mu} = K(\hat{q})^T C(\hat{q})^{-1} y, \quad \hat{\Sigma} = \Sigma_0 - K(\hat{q})^T C(\hat{q})^{-1} K(\hat{q}).$$

We then have the following order of errors.

$$\hat{\mu}^{[i]} = \mu^{[i]} + O(g^{-1}), \quad \hat{\Sigma}^{[ii]} = \Sigma^{[ii]} + O(g^{-1}).$$

Now we consider the discretized version of the GP, i.e., (GMRF) with a precision matrix Q_0 on $\mathcal{S}$. Since the sampling points of GP regression are not necessarily on $\mathcal{S}$, we use the nearest grid point of a given sampling point q_c in $\mathcal{S}_c$

$$q^{[i]} = \arg\min_{q \in \mathcal{S}} \|q_c^{[i]} - q\|.$$

The sampling positions for the GMRF are then exactly on the lattice, i.e., $q^{[i]} \in \mathcal{S}$. The posterior distribution of $z \in \mathbb{R}^n$ on $\mathcal{S}$ given by measurements in $y \in \mathbb{R}^N$ and sampling positions in $q = (q^{[1]T}, \cdots, q^{[N]T})^T \in \mathcal{S}^N$, is then obtained by

$$z|q,y \sim \mathcal{N}(Q^{-1}b, Q^{-1}), \tag{5}$$

where $Q = Q_0 + HP^{-1}H^T$, $b = HP^{-1}y$, with $P = \sigma_\varepsilon^2 I \in \mathbb{R}^{N\times N}$ and $H \in \mathbb{R}^{n\times N}$ defined as

$$H^{[ij]} = \begin{cases} 1, \text{ if } s^{[i]} = q^{[j]}, \\ 0, \text{ otherwise.} \end{cases}$$

We consider again localization uncertainty for this GMRF. Let the measured noisy location $\tilde{q}^{[i]}$ be the nearest grid point of the measured noisy sampling point $\tilde{q}_c^{[i]}$ of the GP. Now we obtain a set of discretized probabilities in $\mathcal{S}$ induced by the continuous prior distribution defined in $\mathcal{S}_c$. The discrete prior distribution for the sampling location $q^{[i]}$ is given by

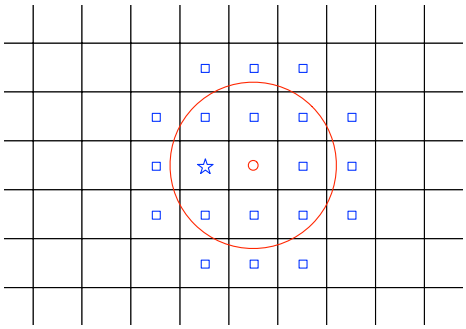
$$\pi(q^{[i]} = s^{[j]}|\tilde{q}^{[i]}) = \int_{s\subset\mathcal{V}_j} \pi(s|\tilde{q}^{[i]})ds,$$

where $\pi(s|\tilde{q}^{[i]})$ is the continuous prior as in GP regression and $\mathcal{V}_j$ is the Voronoi cell of the j-th grid point $s^{[j]}$ given by

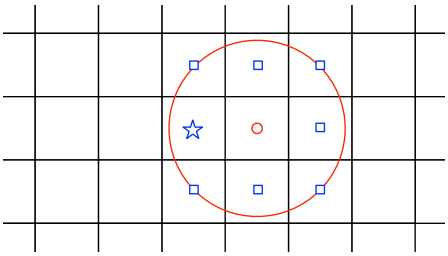
$$\mathcal{V}_j := \{s \in S \,|\, \|s - s^{[j]}\| \le \|s - s^{[i]}\|, \forall i \neq j\}.$$

The predictive distribution of z given y and $\tilde{q}$ is thus given by

$$\pi(z|\tilde{q},y) = \sum_{q\in\mathcal{S}} \pi(z|q,y)\pi(q|\tilde{q},y), \tag{6}$$



(a) h_1



(b) h_2

Figure 1. Example of localization uncertainty for $q^{[i]}$. The measured sampling location $\tilde{q}^{[i]}$ is indicated in a small red circle. The small red circle along with the blue squares and the blue star show the possible locations of the true sampling point $q^{[i]}$ according to the prior distribution $\pi(q^{[i]}|\tilde{q}^{[i]})$ with a compact support as shown in the big red circle. The blue star indicates the location $\hat{q}^{[i]}$ which has the maximum posterior distribution of $\pi(q^{[i]}|\tilde{q},y)$. The predictive distribution of the field z is then approximated by $\pi(z|\hat{q},y)$.

where $\pi(z|q,y)$ can be obtained by (5) and the summation is over all possible locations in $\mathcal{S}$. However, the summation in (6) requires a significant computation time. Hence, we look for a similar approximation to the one in Theorem 2.1.

Let $\hat{q}$ be the MAP estimator that maximizes the posterior distribution $\pi(q|\tilde{q},y)$. As in the continuous case, we can approximate the predictive distribution with $\pi(z|\hat{q},y)$ to reduce the computational complexity by removing the summation in (6). Fig. 1 shows two examples of using this approximation approach with $h_1 > h_2$. When $h \to \infty$, $\tilde{q} \to \tilde{q}_c$, and the predictive distribution approximated by $\pi(z|\hat{q},y)$ converges to the one using Laplace's method as shown in Theorem 2.1. This shows that an approximation based on MAP estimates of the sampling positions can be viewed as a discrete version of Laplace's approximation. Leveraging this idea, we will develop efficient sequential spatial prediction algorithms using a GMRF under uncertain localization in the rest of this paper.

3 Practical Models for Main Problems

In this section, we introduce specific models for the mobile sensor network and the spatio-temporal random field.

3.1 Mobile Sensor Networks

Suppose that the sampling time $t \in \mathbb{Z}_{>0}$ is discrete. Let $z_t := ({z_t}^{[1]}, \cdots, {z_t}^{[n]})^T \in \mathbb{R}^n$ be the corresponding values of the scalar field at n special sites and time t.

Consider N spatially distributed mobile sensing agents indexed by $j \in \mathcal{J} := \{1, \cdots, N\}$ sampling at time $t \in \mathbb{Z}_{>0}$. At time t, agent j takes a noise corrupted measurement at it's current location ${q_t}^{[j]} = s^{[i]} \in \mathcal{S}$, i.e.,

$$y_t^{[j]} = z_t^{[i]} + \varepsilon_t^{[j]}, \quad \varepsilon_t^{[j]} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2), \tag{7}$$

where the measurement errors $\{\varepsilon_t^{[j]}\}$ are assumed to be i.i.d. The measurement noise level $\sigma_\varepsilon^2 > 0$ is assumed to be known. For notational simplicity, we denote all agents' locations at time t by $q_t = \left({q_t}^{[1]T}, \cdots, {q_t}^{[N]T}\right)^T \in \mathcal{S}^N$ and the observations made by all agents at time t by $y_t = \left(y_t^{[1]}, \cdots, y_t^{[N]}\right)^T \in \mathbb{R}^N$. Furthermore, we denote the collection of agents' locations and the collective observations from time 1 to t by $q_{1:t} = \left({q_1}^T, \cdots, {q_t}^T\right)^T \in \mathcal{S}^{Nt}$ and $y_{1:t} = (y_1, \cdots, y_t)^T \in \mathbb{R}^{Nt}$, respectively. In addition, let us define $z_t = ({z_t}^{[1]}, \cdots, {z_t}^{[n]})^T \in \mathbb{R}^n$ on $\mathcal{S}$, and $\varepsilon_t = (\varepsilon_t^{[1]}, \cdots, \varepsilon_t^{[N]})^T \in \mathbb{R}^N$. We then have the following collective notation.

$$y_t = H_t^T z_t + \varepsilon_t, \tag{8}$$

where $H_\tau \in \mathbb{R}^{n \times N}$ is defined by

$$H_\tau^{[ij]} = \begin{cases} 1, \text{ if } s^{[i]} = q_\tau^{[j]}, \\ 0, \text{ otherwise.} \end{cases}$$

3.2 Spatio-Temporal Field Model

The value of the scalar field at space $s^{[i]}$ and time t is denoted by $z_t^{[i]}$ and is modeled by a sum of a time-varying mean function and a GMRF

$${z_t}^{[i]} = \lambda_t^{[i]} + \eta_t^{[i]}, \forall i \in \{1, \cdots, n\}, t \in \mathbb{Z}_{>0}. \tag{9}$$

Here the mean function $\lambda_t^{[i]} : \mathcal{S} \times \mathbb{Z}_{>0} \to \mathbb{R}$ is defined as

$$\lambda_t^{[i]} = f(s^{[i]})^T \beta_t,$$

where $f(s^{[i]}) = (f_1(s^{[i]}), \cdots, f_p(s^{[i]}))^T \in \mathbb{R}^p$ is a known regression function and $\beta_t = (\beta_t^{[1]}, \cdots, \beta_t^{[p]})^T \in \mathbb{R}^p$ is an unknown vector of regression coefficients. The time evolution of $\beta_t \in \mathbb{R}^p$ is modeled by a linear time-invariant system given by

$$\beta_{t+1} = A_t \beta_t + B_t \omega_t, \tag{10}$$

where $\omega_t \sim \mathcal{N}(0, W)$, $\beta_0 \sim \mathcal{N}\left(\mu_{\beta_0}, \Sigma_{\beta_0}\right)$, and A_t and B_t are known system parameters.

In addition, we consider a zero-mean GMRF [21] $\eta_t = \left(\eta_t^{[1]}, \cdots, \eta_t^{[n]}\right)^T \in \mathbb{R}^n$ whose covariance matrix is given by

$$\mathrm{E}(\eta_t \eta_k^T | \theta) = Q_\theta^{-1} \delta(t-k), \tag{11}$$

where $\delta(\cdot)$ is the Kronecker delta defined by

$$\delta(k) = \begin{cases} 1, \, k = 0, \\ 0, \text{ otherwise,} \end{cases}$$

and the inverse covariance matrix (or precision matrix) $Q_\theta \in \mathbb{R}^{n \times n}$ is a function of the hyperparameter vector θ.

There are different parameterizations of the GMRF (i.e., the precision matrix Q_θ) [21]. Our Bayesian approach does not depend on the choice of the parameterization for the precision matrix. However, for a concrete and useful exposition, we describe a specific parameterization used in this paper. The precision matrix is parameterized with the full conditionals as follows.

Let η be a GMRF on a regular two-dimensional lattice. The associated Gaussian full conditional mean is

$$\mathrm{E}(\eta_t^{[i]} | \eta_t^{[-i]}, \theta) = -\frac{1}{Q_\theta^{[ii]}} \sum_{j=1}^{n} Q_\theta^{[ij]} \eta_t^{[j]}, \tag{12}$$

where $Q_\theta^{[ij]}$ is the i-th row and j-th column element of $\kappa^{-1} Q_\theta$. Here, $\eta_t^{[-i]}$ is the collection of η_t values everywhere except $s^{[i]}$. The hyperparameter vector is defined as $\theta = (\kappa, \alpha)^T \in \mathbb{R}^2_{>0}$, where $\alpha = a - 4$. Fig. 2 shows the value of $Q_\theta^{[ij]}$ for one point along with its neighbors, graphically. The value of $Q_\theta^{[ii]}$ is $4 + a^2$ as denoted at the center node of the graph in Fig. 2. That of $Q_\theta^{[ij]}$ is $-2a$ if j is one of the four closest neighbors of i in the vector 1-norm sense as illustrated by the graph in Fig. 2. Thus, the value of $Q_\theta^{[ij]}$ is zero if j is not one of the twelve closest neighbors of i (or twelve neighbors whose 1-norm distance to the i-th location is less than or equal to 2). The equation in (12) states that the conditional expectation of $\eta_t^{[i]}$ given the value of η_t everywhere else (i.e., $\eta_t^{[-i]}$) can be determined just by knowing the value of η_t on the twelve closest neighbors (see more details in [22]). The resulting GMRF accurately represents a Gaussian random field with the Matérn covariance function as shown in [22]

$$G(r) = \sigma_f^2 \frac{2^{1-\rho}}{\Gamma(\rho)} \left(\frac{\sqrt{2\rho} r}{\ell}\right)^\rho K_\rho \left(\frac{\sqrt{2\rho} r}{\ell}\right), \tag{13}$$

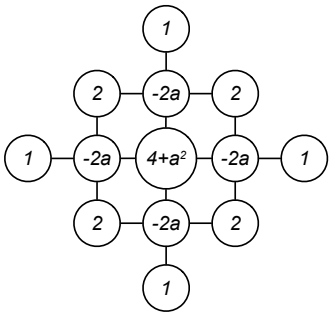


Figure 2. Elements of the precision matrix Q related to a given location.

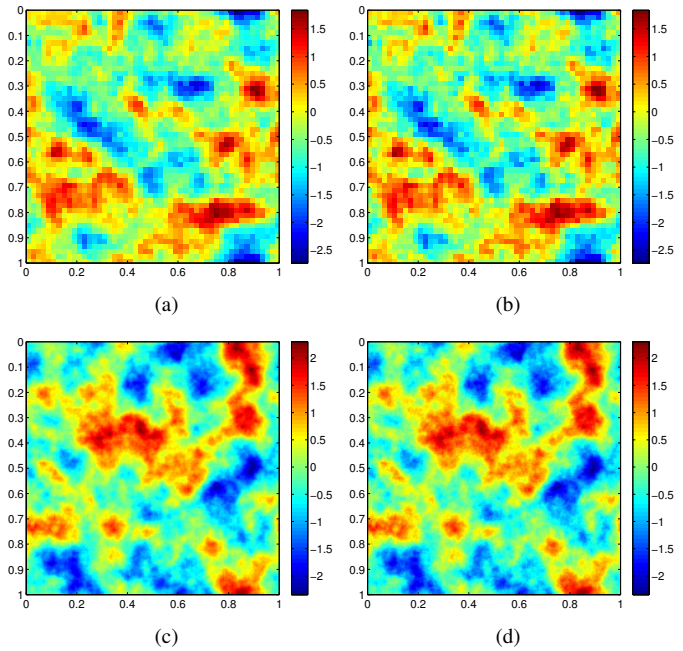


Figure 3. GP vs. GMRF. The left column shows the Gaussian processes with the Matérn covariance function. The right column shows the approximated Gaussian processes by the Gaussian Markov random field. In first, and second rows, the continuous fields are discretized by 50×50, and 200×200 grid points, respectively.

where $K_\rho(\cdot)$ is a modified Bessel function [7], with order $\rho = 1$, a bandwidth $\ell = 1/h\sqrt{\frac{\alpha}{2}}$, and vertical scale $\sigma_f^2 = 1/4\pi\alpha\kappa$. The capability of this GMRF model with a sparse precision matrix to represent a discretized Gaussian process with the same Matérn covariance function is illustrated in Fig. 3.

The hyperparameter $\alpha > 0$ guarantees the positive definiteness of the precision matrix Q_θ. In the case where $\alpha = 0$, the resulting GMRF is a second-order polynomial intrinsic GMRF [21, 23].

From the presented model in (9), (10), and (11), the distribution of z_t given β_t and θ is Gaussian, i.e.,

$$z_t|\beta_t,\theta \sim \mathcal{N}\left(F_s\beta_t, Q_\theta^{-1}\right), \tag{14}$$

where $F_s := (f(s^{[1]}),\cdots,f(s^{[n]}))^T \in \mathbb{R}^{n\times p}$.

In other words, $z_t|\beta_t,\theta \sim \mathcal{GP}(F_s\beta_t,\Sigma_\theta) \in \mathbb{R}^n$ is a non-zero mean Gaussian process. Here, the covariance matrix Σ_θ is defined as inverse of the precision matrix (i.e., $\Sigma_\theta = Q_\theta^{-1}$). Note that the precision matrix is a positive definite matrix and invertible, and $\Sigma_\theta^{[ij]} = \mathrm{Cov}(z_t^{[i]}, z_t^{[j]})$, where $\Sigma_\theta^{[ij]}$ is the i,j-th element of the covariance matrix.

For simplicity, let us define $\mathcal{B}_t = \{\beta_t, q_t, y_t, \theta\}$. Using Gaussian process regression, the posterior distribution for $z_t|\mathcal{B}_t \in \mathbb{R}^n$ is given by

$$\begin{aligned}\mu_{z_t|\mathcal{B}_t} &= H_t^T F_s\beta_t + \Sigma_\theta H_t\left(H_t^T\Sigma_\theta H_t + \sigma_\varepsilon^2 I\right)^{-1}\left(y_t - H_t^T F_s\beta_t\right),\\ \Sigma_{z_t|\mathcal{B}_t} &= \Sigma_\theta - \Sigma_\theta H_t\left(H_t^T\Sigma_\theta H_t + \sigma_\varepsilon^2 I\right)^{-1} H_t^T\Sigma_\theta.\end{aligned} \tag{15}$$

The basic idea behind the model introduced in (9), (10), and (11) stems from the space-time Kalman filter model proposed in [24]. The advantage of this spatio-temporal model with known hyperparameters is to make inferences in a recursive manner as the number of observations increases. The zero-mean Gaussian process represents a spatial structure by assuming that the difference between the parametric mean function and the dynamical environmental process is governed by a relatively large time scale. This formulation in turn makes the optimal prediction recursive in time.

In this paper, however, uncertainties in the precision matrix and sampling positions are considered in a fully Bayesian manner. In addition, in contrast to [4, 24], the GMRF with a sparse precision matrix is used to increase the computational efficiency.

4 Bayesian Predictive Inference

In this section, we present efficient and scalable prediction algorithms for cases with perfect and uncertain localization.

4.1 Uncertain hyperparameters and exact localization

In this section, we brifly state the result of [20] to make predictive inferences of the spatio-temporal random field $z_t \in \mathbb{R}^n$ for the case with uncertain hyperparameters and the exact localization. To this end, we use the following assumptions A.1-A.5.

A.1 The spatio-temporal random field is generated by (9), (10), and (11).

A.2 The precision matrix Q_θ is a given function of an uncertain hyperparameter vector θ.

A.3 The noisy measurements $\{y_t\}$, as in (8), are continuously collected by robotic sensors in time t.

A.4 The sample positions $\{q_t\}$ are measured precisely by robotic sensors in time t.

A.5 The prior distribution of the hyperparameter vector θ is discrete with a support $\Theta = \{\theta^{(1)},\cdots,\theta^{(L)}\}$.

For notational simplicity, we denote the full latent field of dimension $n+p$ by $x_t = ({z_t}^T, \beta_t^T)^T$. Let's define $\mathcal{D}_{k:r} := \{\mathcal{P}_{k-1}, q_{k:r}, y_{k:r}\}$, where $\mathcal{P}_k = \{\mu_{x_k|\mathcal{D}_{1:k}}, \Sigma_{x_k|\mathcal{D}_{1:k}}\} \cup \{\pi(\theta|\mathcal{D}_{1:k}) | \theta \in \Theta\}$, and $\mathcal{P}_0$ is assumed to be known.

We formulate the first problem as follows.

Problem 4.1. *Consider the assumptions A.1-A.5. Our problem is to find the predictive distribution, mean, and variance of x_t conditional on $\mathcal{D}_{t-m+1:t}$.*

In what follows, we summarize the intermediate steps to obtain the solution to Problem 4.1. Under assumptions A.1 and A.2, the predictive distribution of x_t conditional on the hyperparameter vector θ and the measurements $\mathcal{D}_{t-m+1:t-1}$ is Gaussian with the following mean and precision matrix

$$\begin{aligned}\mu_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}} &= \begin{pmatrix} F_s \mu_{\beta_t|\theta,\mathcal{D}_{t-m+1:t-1}} \\ \mu_{\beta_t|\theta,\mathcal{D}_{t-m+1:t-1}} \end{pmatrix}, \\ Q_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}} &= \begin{pmatrix} Q_\theta & -Q_\theta F_s \\ -F_s^T Q_\theta & F_s^T Q_\theta F_s + \Sigma^{-1}_{\beta_t|\theta,\mathcal{D}_{t-m+1:t-1}} \end{pmatrix},\end{aligned} \tag{16}$$

where $\mu_{\beta_t|\theta,\mathcal{D}_{t-m+1:t-1}}$ denotes the expectation of β_t conditional on θ and $\mathcal{D}_{t-m+1:t-1}$ obtained by

$$\mu_{\beta_t|\theta,\mathcal{D}_{t-m+1:t-1}} = A_t \mu_{\beta_{t-1}|\theta,\mathcal{D}_{t-m+1:t-1}}$$

and the associated estimation error covariance matrix can be obtained by

$$\Sigma_{\beta_t|\theta,\mathcal{D}_{t-m+1:t-1}} = A_t \Sigma_{\beta_{t-1}|\theta,\mathcal{D}_{t-m+1:t-1}} A_t^T + B_t W B_t^T.$$

For a given hyperparameter vector θ, (16) provides the optimal prediction of the spatio-temporal field in time t using data up to time $t-1$.

Under assumptions A.3 and A.4, the posterior distribution of the hyperparameter vector θ can be obtained recursively via

$$\pi(\theta|\mathcal{D}_{t-m+1:t}) \propto \pi(y_t|\theta, \mathcal{D}_{t-m+1:t-1}, q_t)\pi(\theta|\mathcal{D}_{t-m+1:t-1}), \tag{17}$$

where the distribution of y_t given $\{\theta, \mathcal{D}_{t-m+1:t-1}, q_t\}$ is Gaussian with the following mean and variance

$$\begin{aligned}\mu_{y_t|\theta,\mathcal{D}_{t-m+1:t-1},q_t} &= \Gamma_{q_t}^T \mu_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}}, \\ \Sigma_{y_t|\theta,\mathcal{D}_{t-m+1:t-1},q_t} &= \Gamma_{q_t}^T \Sigma_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}} \Gamma_{q_t} + \sigma_\varepsilon^2 I,\end{aligned} \tag{18}$$

here $\Gamma_{q_t}^T = [H_t^T \ 0] \in \mathbb{R}^{N\times(n+p)}$.

Algorithm 1 Sequential Bayesian predictive inference.

Initialization:

1: initialize F_s
2: for $\theta \in \Theta$, initialize Q_θ, and compute Q_θ^{-1}

At time $t \in \mathbb{Z}_{>0}$, do:

1: obtain new observations y_t collected at current locations q_t
2: find the map Γ_{q_t} from q_t to spacial sites $\mathcal{S}$, and compute radial basis values F_{q_t} in q_t.
3: **for** $\theta \in \Theta$ **do**
4: predict $\mu_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}}$ and $Q_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}}$ using measurements up to time $t-1$, given by (16).
5: compute $\mu_{x_t|\theta,\mathcal{D}_{t-m+1:t}}$ and $Q_{x_t|\theta,\mathcal{D}_{t-m+1:t}}$ given by (20).
6: compute $\mu_{y_t|\theta,\mathcal{D}_{t-m+1:t-1},q_t}$ and $\Sigma_{y_t|\theta,\mathcal{D}_{t-m+1:t-1},q_t}$ given by (18).
7: calculate $\pi(\theta|\mathcal{D}_{t-m+1:t})$ given by (17).
8: **end for**
9: compute the predictive mean and variance using (22).

Under assumptions A.1-A.4, the full conditional distribution of x_t for a given hyperparameter vector and data up to time t is also Gaussian, i.e.,

$$x_t|\theta, \mathcal{D}_{t-m+1:t} \sim \mathcal{N}(\mu_{x_t|\theta,\mathcal{D}_{t-m+1:t}}, Q^{-1}_{x_t|\theta,\mathcal{D}_{t-m+1:t}}), \tag{19}$$

where

$$\begin{aligned}Q_{x_t|\theta,\mathcal{D}_{t-m+1:t}} =& Q_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}} + \sigma_\varepsilon^{-2}\Gamma_{q_t}\Gamma_{q_t}^T, \\ \mu_{x_t|\theta,\mathcal{D}_{t-m+1:t}} =& \mu_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}} + \sigma_\varepsilon^{-2} Q^{-1}_{x_t|\theta,\mathcal{D}_{t-m+1:t}} \Gamma_{q_t} \\ & (y_t - \Gamma_{q_t}^T \mu_{x_t|\theta,\mathcal{D}_{t-m+1:t-1}}).\end{aligned} \tag{20}$$

Under assumption A.5, the predictive distribution of $x_t|\mathcal{D}_{t-m+1:t}$ is given by

$$\pi(x_t|\mathcal{D}_{t-m+1:t}) = \sum_{\theta\in\Theta} \pi(x_t|\theta, \mathcal{D}_{t-m+1:t})\pi(\theta|\mathcal{D}_{t-m+1:t}), \tag{21}$$

where $\pi(\theta|\mathcal{D}_{t-m+1:t})$ and $\pi(x_t|\theta, \mathcal{D}_{t-m+1:t})$ are given by (17) and (19), respectively. The predictive mean and variance are as follows.

$$\begin{aligned}\mu_{x_t|\mathcal{D}_{t-m+1:t}} &= \sum_{\theta\in\Theta} \mu_{x_t|\theta,\mathcal{D}_{t-m+1:t}} \pi(\theta|\mathcal{D}_{t-m+1:t}), \\ \Sigma_{x_t|\mathcal{D}_{t-m+1:t}} &= \\ &\sum_{\theta\in\Theta} \big[\Sigma_{x_t|\theta,\mathcal{D}_{t-m+1:t}} + (\mu_{x_t|\theta,\mathcal{D}_{t-m+1:t}} - \mu_{x_t|\mathcal{D}_{t-m+1:t}}) \\ &\quad (\mu_{x_t|\theta,\mathcal{D}_{t-m+1:t}} - \mu_{x_t|\mathcal{D}_{t-m+1:t}})^T\big] \pi(\theta|\mathcal{D}_{t-m+1:t}).\end{aligned} \tag{22}$$

The proposed solution to the formulated problem is summarized by Algorithm 1.

4.2 Uncertain hyperparameters and localization

In the previous section, we assumed that the localization data $q_{1:t}$ is exactly known. However, in practice positions of sensor networks cannot be measured without noise. Instead, for example, there could be several probable possibilities inferred from the measured position. In [20], we provided the exact solution and a successive approximation to this problem. In this section, a new approximation will be proposed to reduce the computation complexity significantly.

In order to take into account the uncertainty in the sampling positions, we replace assumption A.4 with the following assumption A.6.

A.6 The prior distribution $\pi(q_t)$ is discrete with a support $\Omega(t) = \{q_t^{(k)} | k \in I(t)\}$, which is given at time t along with the corresponding measurement y_t. Here, $I(t) = \{1, \cdots, \gamma(t)\}$ denotes the index in the support and $\gamma(t)$ is the number of the probable possibilities for q_t.

An straightforward consequence of the assumption A.6 is that the prior distribution $\pi(q_{k:r})$ is discrete with a support $\Omega(k:r) := \prod_{g=k}^{r} \Omega(g)$. In addition, $I(k:r) := \prod_{g=k}^{r} I(g)$ denotes the index in the support $\Omega(k:r)$, and $\gamma(k:r) := \prod_{g=k}^{r} \gamma(g)$ is the number of the probable possibilities for $q_{k:r}$. Now we state the problem as follows.

Problem 4.2. *Consider assumptions A.1, A.2, A.3, A.5, and A.6. Our problem is to find the predictive distribution, mean and variance of x_t conditional on the prior $\mathcal{P}_0$ and the measurements $y_{1:t}$.*

For the sake of conciseness, let us define $\mathcal{R}_{k:k} := \{\mathcal{P}_{r-1}, y_{r:k}\}$. We then have that $\mathcal{R}_{k:k} \subset \mathcal{D}_{r:k}$, where we recall that $\mathcal{D}_{r:k} := \{\mathcal{P}_{r-1}, q_{r:k}, y_{r:k}\}$.

Here, we proposed an approximation, with a controllable tradeoff between approximation error and complexity. The idea is based on the fact that the estimation of x_t is more susceptible to the uncertainties in recently sampled positions as compared to old ones. In addition, to avoid the summation over all probable sampling positions we compute the MAP estimates of sampling positions and we plug them into Algorithm 1.

Using prior distribution of x_{t-m} and measured data $y_{t-m+1:t}$, where $m \in \mathbb{Z}_{>0}$, the posterior distribution of $q_{t-m+1:t}$ can be obtained recursively via

$$\begin{aligned}\pi\left(q_{t-m+1:t-1}^{(j)}, q_t^{(k)} | \mathcal{R}_{t-m+1:t}\right) \propto \\ \pi(q_{t-m+1:t-1}^{(j)} | \mathcal{R}_{t-m+1:t-1}) \pi(y_t | \mathcal{D}_{t-m+1:t-1}^{(j)}, q_t^{(k)}) \pi(q_t^{(k)}).\end{aligned} \tag{23}$$

where $j \in I(t-m+1:t-1)$, and $k \in I(t)$.

We consider the following conditions.

C.1 For $1 \ll m \leq t$, we have that

$$\pi(x_t | \mathcal{R}_{1:t}) \approx \pi(x_t | \mathcal{R}_{t-m+1:t}) \tag{24}$$

Algorithm 2 Sequential Bayesian predictive inference approximation with uncertain localization.

At time $t \in \mathbb{Z}_{>0}$, do:

1: obtain new observations y_t along with the probabilities for locations $\pi(q_t)$
2: **for** $q_t^{(h)} \in \Omega(t-m+1:t)$ **do**
3: compute $\pi\left(q_{t-m+1:t}^{(h)} | \mathcal{R}_{t-m+1:t}\right)$ using (23).
4: **end for**
5: find MAP estimation $\hat{q}_{t-m+1:t}$ using (27)
6: use following approximation to update estimations

$$\mu_{x_t|\mathcal{R}_{1:t}} \approx \mu_{x_t|\hat{\mathcal{D}}_{t-m+1:t}}, \Sigma_{x_t|\mathcal{R}_{1:t}} \approx \Sigma_{x_t|\hat{\mathcal{D}}_{t-m+1:t}},$$
$$\mathcal{P}_t \approx \{\mu_{x_t|\hat{\mathcal{D}}_{t-m+1:t}}, \Sigma_{x_t|\hat{\mathcal{D}}_{t-m+1:t}}\} \cup \{\pi(\theta|\hat{\mathcal{D}}_{t-m+1:t}) | \theta \in \Theta\}$$

C.2 For $1 \ll m \leq t$, $\mathcal{P}_t$ can be approximated by

$$\mathcal{P}_t \approx \{\mu_{x_t|\mathcal{R}_{t-m+1:t}}, \Sigma_{x_t|\mathcal{R}_{t-m+1:t}}\} \cup \{\pi(\theta|\mathcal{R}_{t-m+1:t}) | \theta \in \Theta\}. \tag{25}$$

Under conditions C.1 and C.2, [20] proposed the following approximations (see Theorem 3.13 in [20]).

$$\mu_{x_t|\mathcal{R}_{1:t}} \approx \mu_{x_t|\mathcal{R}_{t-m+1:t}}, \quad \Sigma_{x_t|\mathcal{R}_{1:t}} \approx \Sigma_{x_t|\mathcal{R}_{t-m+1:t}}. \tag{26}$$

However, the proposed approximation in (26) has constant complexity in time, still the number of possibilities for sampling positions can increase the computational time considerably. To minimize the computational complexity, one may prefer a simpler approximation. As we discussed in Section 2, we propose efficient approximation by using MAP estimation of sampling positions. The MAP estimator is given by

$$\hat{q}_{t-m+1:t} = \arg \max_{q_{t-m+1:t}^{(h)} \in \Omega_{t-m+1:t}} \pi\left(q_{t-m:t}^{(h)} | \mathcal{R}_{t-m+1:t}\right), \tag{27}$$

where the distribution of $q_{t-m+1:t}^{(h)} | \mathcal{R}_{t-m+1:t}$ is computed by (23), recursively. Let us define $\hat{\mathcal{D}}_{t-m+1:t} = \{\hat{q}_{t-m+1:t}\} \cup \mathcal{R}_{t-m+1:t}$. We propose the following approximations

$$\mu_{x_t|\mathcal{R}_{1:t}} \approx \mu_{x_t|\hat{\mathcal{D}}_{t-m+1:t}}, \quad \Sigma_{x_t|\mathcal{R}_{1:t}} \approx \Sigma_{x_t|\hat{\mathcal{D}}_{t-m+1:t}}. \tag{28}$$

4.3 Complexity of Algorithms

In this section, we discuss complexity aspects of the proposed algorithms. The complexity of Algorithm 1 is studied in [20]. The complexity of the approximation proposed in [20] is

$O(\gamma(t-m+1:t))$ times the complexity of Algorithm 1 for m time steps. In contrast to the one proposed in [20], the complexity of the new scheme, summarized by Algorithm 2, is the complexity of Algorithm 1 for m time steps plus the complexity of finding $\hat{q}$, which significantly improves in computation time.

For a fixed number of radial basis functions (i.e., p) and a fixed number of the special sites (i.e., n), the computational complexity of finding $\hat{q}$ in each time step is $O(mLN^2)$, where L is the number of possible hyperparameter vectors and N is the number of agents. Note that the number of radial basis functions affect the complexity of finding $\hat{q}$ as well. For a fixed set of n, L and N, the complexity of finding $\hat{q}$ with respect to p is $O\left(p^3\right)$.

5 Simulation and Experimental Results

In this section, we demonstrate the effectiveness of the proposed sequential Bayesian inference algorithms using simulations and experiments. We apply the proposed prediction algorithms to real experimental data. The light intensity fields can be easily realized for experimental setups [25,26]. Fig. 4-(a) shows our experimental setup in which a light intensity field was generated by inserting opaque materials under an acrylic plate. A CCD camera on the top of the plate captured the true light field from which noisy measurements are sampled at random sampling positions by simulated robotic sensors. The true field is used for evaluating the performance of the proposed schemes. The true scalar field, shown in Fig. 4-(b), was realized by inserting crumpled news papers between the plate and light sources. The objective is to predict the light intensity field, using the scalar field model proposed in Section 3.2.

In this study, we use the spatio-temporal field introduced in Section 3.2. The spacial sites in $\mathcal{S}$ consist of 51×51 grid points, i.e., $n = 2601$, uniformly distributed over the surveillance region $[-25,25] \times [-25,25]$. The time-varying mean function μ_t consists of ten radial basis functions (i.e., $p = 10$) given by $f_j(s^{[i]}) = \exp\left(-\frac{\|s^{[i]}-\xi_j\|^2}{2\sigma_j^2}\right), j \in \{1,\cdots,p\}$, where σ_j is the bandwidth and ξ_j is the center location of the j-th radial basis function. The first radial basis function has an infinity bandwidth (i.e., $\sigma_1 = \infty$) to represent the average of the field, and the others have a bandwidth equal to $\sigma_j = 15$. The centers of radial basis functions are $\{(0,0)\} \cup \{-15,0,15\} \times \{-15,0,15\}$. The prior distribution of β_0 is chosen to be $\beta_0 \sim \mathcal{N}(0,5I)$. The time evolution of β_t is modeled by (10), where the state matrix A_t and the input matrix B_t are given by $0.95I_{10\times10}$ and $0.5I_{10\times10}$, respectively. The input disturbance variance and the measurement noise variance are known to be $W = I_{10\times10}$ and $\sigma_\varepsilon^2 = 0.1$, respectively. Regarding the GRMF part in the model, the prior distribution of the hyperparameter vector θ is chosen to be discrete with a support $\Theta = \{\frac{50}{9}, \frac{50}{3}, 50, 150, 450\} \times \{\frac{0.005}{9}, \frac{0.005}{3}, 0.005, 0.015, 0.045\}$ with the associated uniform probabilities.

To evaluate the effectiveness of the proposed approximation in dealing with uncertain sampling positions, thirty (virtual) robotic sensors (i.e, $N = 30$) take measurements at time $t = 1$, where all agents know their sampling positions exactly except six of them (20% uncertain sampling positions). In Figs. 5-(g), (h), and (i), true, noisy, and probable sampling positions are shown in circles, stars, and corners of squares, respectively.

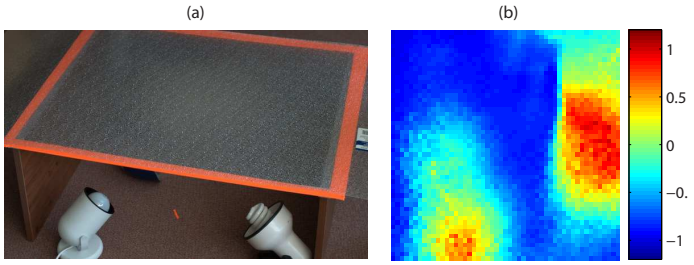


Figure 4. (a) The experimental setup to generate the light intensity field using three light bulbs. (b) The true light intensity field.

The prediction results are summarized for three methods of prediction as follows.

Case 1: Figs. 5-(a), (d), and (g) show the prediction, prediction error variance, and squared (empirical) error fields, using Algorithm 1 with exact sampling positions. With the true sampling positions, the best prediction quality is expected for this case.
Case 2: Figs. 5-(b), (e), and (h) show the resulting fields, by applying Algorithm 1 naively to measured sampling positions including noisy locations. The results clearly illustrate that naively applying Algorithm 1 to noisy sampling positions can potentially distort prediction at a significant level as shown in Figs. 5-(b) and (h).
Case 3: Figs. 5-(c), (f), and (i) show the resulting fields, by applying Algorithm 2 with $m = 1$. The resulting prediction quality is much improved as compared to Case 2 and is even comparable to the result for Case 1.

The results confirm that the quality of the prediction in Case 3 is not much compromised as compared to Case 1 and demonstrate the capability of our proposed algorithm to deal with uncertain sampling positions. The average of the squared empirical errors in Case 3 is 42% smaller than that of Case 2.

In this study, the fixed running time using Matlab R2009b (MathWorks) on a PC (2.4 GHz Intel Core 2 Duo Processor) is about 10 minutes for the proposed approximation ($m = 1$) which is fast enough for real world implementation.

Table 1 summarizes results of Case1, Case 2, Case 3 (our proposed approach), and another previously proposed approach in [20] for this experimental data and the simulated realization in [20] in term of the prediction precision and the computational time. The exact configuration of the simulation study can be found in [20]. The level of precision between the proposed approach (Case 3) based on the MAP estimator and another one proposed in [20] are comparable. The major contribution of this paper is the improvement in the computation time of the pro-

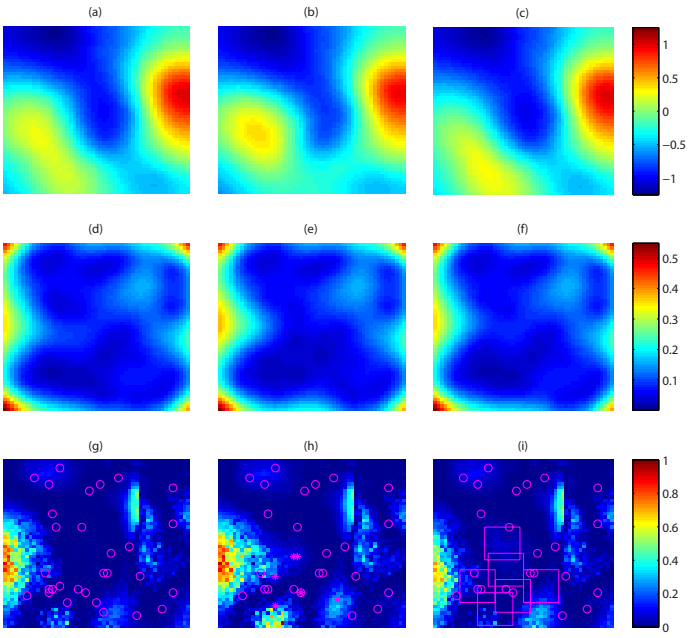


Figure 5. The prediction results of Cases 1, 2, and 3 at time $t = 1$ are shown in the first, second, third columns, respectively. The first, second, and third rows correspond to the predictions, prediction error variance, and squared empirical error fields between predicted and true fields. True, noisy, and probable sampling positions are shown in circles, stars, and corners of squares, respectively.

posed solution. The proposed solution improves the computation time by 200% as shown in Table 1.

6 Conclusion

We have discussed the problem of predicting a spatio-temporal field using successive noisy measurements obtained by robotic sensors, some of which have uncertain localization. We modeled the spatio-temporal field of interest using a GMRF and designed an efficient and sequential prediction algorithm for computing the approximated predictive inference from a Bayesian point of view. The proposed algorithm is computationally efficient and scalable as the number of measurements increases. Application of our proposed scheme to simulated and real-world data shows that our approach outperforms the previously developed approximated and exact Bayesian solutions in computation time while keeping the comparable level of precision. The future work is to evaluate the proposed algorithms in a realistic experimental setup.

Acknowledgment

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

REFERENCES

[1] K. M. Lynch, I. B. Schwartz, P. Yang, and R. A. Freeman, "Decentralized environmental modeling by mobile sensor networks," *IEEE Transactions on Robotics*, vol. 24, no. 3, pp. 710–724, June 2008.

[2] N. E. Leonard, D. A. Paley, F. Lekien, R. Sepulchre, D. M. Fratantoni, and R. Davis, "Collective motion, sensor networks, and ocean sampling," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 48–74, January 2007.

[3] J. Choi, S. Oh, and R. Horowitz, "Distributed learning and cooperative control for multi-agent systems," *Automatica*, vol. 45, no. 12, pp. 2802–2814, December 2009.

[4] J. Cortés, "Distributed Kriged Kalman filter for spatial estimation," *IEEE Transactions on Automatic Control*, vol. 54, no. 12, pp. 2816–2827, December 2009.

[5] Y. Xu, J. Choi, S. Dass, and T. Maiti, "Sequential Bayesian prediction and adaptive sampling algorithms for mobile sensor networks," *IEEE Transactions on Automatic Control*, 2012, to appear.

[6] N. Cressie, "Kriging nonstationary data," *Journal of the American Statistical Association*, vol. 81, no. 395, pp. 625–634, September 1986.

[7] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. The MIT Press, Cambridge, Massachusetts, London, England, 2006.

[8] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies," *The Journal of Machine Learning Research*, vol. 9, pp. 235–284, June 2008.

[9] Y. Xu, J. Choi, and S. Oh, "Mobile sensor network navigation using Gaussian processes with truncated observations," *IEEE Transactions on Robotics*, vol. 27, no. 6, pp. 1118 – 1131, December 2011.

[10] R. Graham and J. Cortés, "Cooperative adaptive sampling of random fields with partially known covariance," *International Journal of Robust and Nonlinear Control*, vol. 22, no. 5, pp. 504–534, March 2012.

[11] Y. Xu and J. Choi, "Adaptive sampling for learning Gaussian processes using mobile sensor networks," *Sensors*, vol. 11, no. 3, pp. 3051–3066, March 2011.

[12] C. M. Bishop, *Pattern recognition and machine learning*. Springer, New York, 2006.

[13] H. Rue and H. Tjelmeland, "Fitting Gaussian Markov random fields to Gaussian fields," *Scandinavian Journal of Statistics*, vol. 29, no. 1, pp. 31–49, March 2002.

[14] N. Cressie and N. Verzelen, "Conditional-mean least-squares fitting of Gaussian Markov random fields to Gaussian fields," *Computational Statistics & Data Analysis*, vol. 52, no. 5, pp. 2794–2807, January 2008.

[15] L. Hartman and O. Hössjer, "Fast kriging of large data sets with Gaussian Markov random fields," *Computational Statistics & Data Analysis*, vol. 52, no. 5, pp. 2331–2349, January 2008.

[16] Y. Xu, J. Choi, S. Dass, and T. Maiti, "Efficient Bayesian

Table 1. Comparison of precision and computation time.

Experiment	Case 1	Case 2 (Naive approach)	Case 3 (Proposed approximation)	Approximation in [20]
Squared empirical error	0.077	0.097	0.068	0.080
Comp. time (minute)	1	1	10	30
Simulation	Case 1	Case 2 (Naive approach)	Case 3 (Proposed approximation)	Approximation in [20]
Squared empirical error	1.173	3.276	1.242	1.204
Comp. time (minute)	0.5	0.5	3	10

spatial prediction with mobile sensor networks using Gaussian Markov random fields," 2012, submitted to Automatica.

[17] J. Le Ny and G. J. Pappas, "On trajectory optimization for active sensing in Gaussian process models," in *Proceedings of the 48th IEEE Conference on Decision and Control*, December 2009, pp. 6286–6292.

[18] Y. Xu and J. Choi, "Spatial prediction with mobile sensor networks using Gaussian Markov random fields," in *Proceedings of the 2011 ASME Dynamic Systems and Control Conference*, Arlington, VA, 2011.

[19] M. Jadaliha, Y. Xu, and J. Choi, "Gaussian process regression using Laplace approximations under localization uncertainty," in *Proceedings of the American Control Conference*, 2012, to appear.

[20] ——, "Sequential Bayesian prediction for mobile sensor networks with uncertain localization using GMRFs," 2012, submitted to the 51st IEEE Conference on Decision and Control.

[21] H. Rue and L. Held, *Gaussian Markov random fields: theory and applications*. Chapman & Hall, 2005.

[22] F. Lindgren, H. Rue, and J. Lindström, "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *Journal of the Royal Statistical Society: Series B*, vol. 73, no. 4, pp. 423–498, September 2011.

[23] H. Rue, S. Martino, and N. Chopin, "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 2, pp. 319–392, April 2009.

[24] N. Cressie and C. Wikle, "Space–time Kalman filter," *Encyclopedia of environmetrics*, vol. 4, pp. 2045–2049, 2002.

[25] M. Schwager, J. McLurkin, J. Slotine, and D. Rus, "From theory to practice: Distributed coverage control experiments with groups of robots," *Experimental Robotics*, vol. 54, pp. 127–136, 2009.

[26] H. Hamann, *Space-Time Continuous Models of Swarm Robotic Systems: Supporting Global-to-Local Programming*. Springer-Verlag New York Inc, 2010, vol. 9.