Gaussian process regression for sensor networks under localization uncertainty

Mahdi Jadaliha, Yunfei Xu, Jongeun Choi, Nicholas S. Johnson and Weiming Li

***Abstract*—In this paper, we formulate Gaussian process regression with observations under the localization uncertainty due to the resource-constrained mobile sensor networks. In our formulation, effects of observations, measurement noise, localization uncertainty, and prior distributions are all correctly incorporated in the posterior predictive statistics. The analytically intractable posterior predictive statistics are proposed to be approximated by two techniques, viz., Monte Carlo sampling and Laplace's method. Such approximation techniques have been carefully tailored to our problems and their approximation error and complexity are analyzed. Simulation study demonstrates that the proposed approaches perform much better than approaches without considering the localization uncertainty properly. Finally, we have applied the proposed approaches on the experimentally collected real data from a dye concentration field over a section of a river and a temperature field of an outdoor swimming pool to provide proof of concept tests and evaluate the proposed schemes in real situations. In both simulation and experimental results, the proposed methods outperform the quick-and-dirty solutions often used in practice.**

I. Introduction

Recently, there has been a growing interest in wireless sensor networks due to advanced embedded network technology. Their applications include, but are not limited to, environment monitoring, building comfort control, traffic control, manufacturing and plant automation, and surveillance systems [1], [2]. Mobility in a sensor network can increase its sensing coverage both in space and time, and robustness against uncertainties in environments. Exploitation of mobile sensor networks has been increased in collecting spatio-temporal data from the environment [3]–[5].

Gaussian process regression (or Kriging in geostatistics) has been widely used to draw statistical inference from geostatistical and environmental data [6], [7]. Gaussian process modeling enables us to predict physical values, such as temperature or harmful algae bloom biomass, at any point and time with a predicted uncertainty level. For example, near-optimal static sensor placements with a mutual information criterion in Gaussian processes were proposed in [8], [9]. A distributed Kriged Kalman filter for spatial estimation based on mobile sensor networks was developed in [5]. Centralized and distributed navigation strategies for mobile sensor networks to move in order to reduce prediction error variances at points of interest were developed in [10].

Localization in sensor networks is a fundamental problem in various applications to correctly fuse the spatially collected data and estimate the process of interest. However, obtaining precise localization of robotic networks under limited resources is very challenging [11], [12]. Among the localization schemes, the range-based approach [13], [14] provides higher precision as compared to the range-free approach that could be cost-effective. The global positioning system (GPS) becomes one of the major absolute positioning systems in robotics and mobile sensor networks. Most of affordable GPSs slowly update their measurements and have resolution worse than one meter. A GPS is often augmented by the inertial navigation system (INS) for better resolution [15]. In practice, resource-constrained sensor network systems are prone to large uncertainty in localization. Most previous works on Gaussian process regression for mobile sensor networks [6], [9], [10], [16] have assumed that the exact sampling positions are available, which is not practical in real situations.

Therefore, motivated by the aforementioned issues, we consider correct (Bayesian) integration of uncertainties in sampling positions, and measurements noise for Gaussian process regression and its computation error and complexity analysis for the sensor network applications. The overall picture of our work is similar to the one in [17] in which an extended Kalman filter (EKF) was used to incorporate robot localization uncertainty and field parameter uncertainty. However, [17] relies on a parametric model, which is a radial basis function network model and EKF, while our motivation is to use more flexible non-parametric approach, viz., Gaussian process regression taking into account localization uncertainty in a Bayesian framework.

Gaussian process regression in [18] integrate uncertainty in the test input position for multiple-step ahead time series forecasting. In [18], uncertainty was not considered in the sampling positions of the training data (or observations).

However, localization uncertainty effect is potentially significant. For example, Fig. 1 shows the effect of noisy sampling positions on the results of Gaussian process regression. Note that adding noise to the sampling positions considerably increase the empirical RMS error, shown in the third row of

M. Jadaliha is with the Department of Mechanical Engineering, Michigan State University, MI 48824, USA jadaliha@msu.edu

Y. Xu is with the Department of Mechanical Engineering, Michigan State University, MI 48824, USA xuyunfei@egr.msu.edu

J. Choi is with the Departments of Mechanical Engineering and Electrical and Computer Engineering, Michigan State University, MI 48824, USA jchoi@egr.msu.edu

N. S. Johnson is with USGS, Great Lakes Science Center, Hammond Bay Biological Station, 11188 Ray Road, Millersburg, MI 49759, USA njohnson@usgs.gov

W. Li is with the Department of Fisheries and Wildlife, Michigan State University, Room 13 Natural Resources Building, East Lansing, MI 48824, USA xliweim@anr.msu.edu

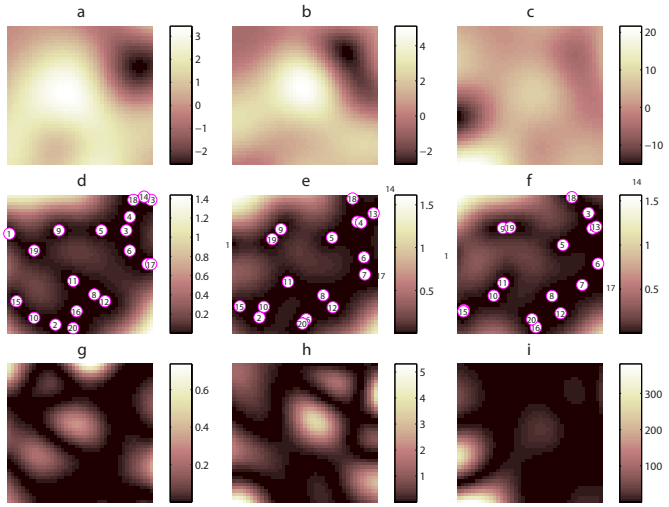


Fig. 1. The prediction results of applying Gaussian process regression on the true and noisy sampling position are shown. The first, second, and third rows correspond to the prediction, prediction error variance, and squared empirical error (between predicted and true fields) fields. The first column shows the result of applying Gaussian process regression on the true sampling positions. Second and third columns show the results of applying Gaussian process regression on the noisy sampling positions with $\Sigma_1 = 0.1I$ and $\Sigma_2 = 0.4I$ noise covariance matrices, respectively. True and noisy sampling positions are shown in circles with agent indices in the second row.

Fig. 1.

A Gaussian approximation to the intractable posterior predictive statistics obtained in [18] has been utilized for the predictive control with Gaussian models [19], [20] and Gaussian process dynamic programming [21]. In general, the length of the training data is much longer than that of the test point for sensor network applications, therefore, our problem possibly involves a high dimensional uncertainty vector for the sampling positions.

Gaussian process prediction with localization uncertainty can be obtained as a posterior predictive distribution using Bayes' rule. The main difficulty to this is that it has no analytic closed-form solution and has to be approximated either through Monte Carlo sampling [22] or other approximation techniques such as variational inference [23]. As an important analytical approximation technique, Laplace's method has been known to be useful in many such situations [24], [25]. Different Laplace approximation techniques have been analyzed in terms of approximation error and computation complexity [24]–[27].

The contribution of this paper is as follows. First, we formulate Gaussian process regression with observations under the localization uncertainty due to the resource-constrained (possibly mobile) sensor networks. Next, approximations have been obtained for analytically intractable predictive mean, and predictive variance by using the Monte Carlo sampling and Laplace's method. Such approximation methods have been carefully tailored to our problems. In particular, a modified version of the moment generating function (MGF) approximation [25] has been developed as a part of Laplace's method to reduce the computational complexity. In addition, we have analyzed and compared the approximation error and the complexity so that one can choose a tradeoff between the performance requirements and constrained resources for a particular sensor network application. Another important contribution has been to provide proof of concept tests and evaluate the proposed schemes in real situations. We have applied the proposed approaches on the experimentally collected real data from a dye concentration field over a section of a river and a temperature field of an outdoor swimming pool.

The preliminary work of this paper containing only the Laplace methods along with simulation results was reported in [28].

The remainder of this paper is organized as follows. In Section II, we review Gaussian process regression, Monte Carlo sampling and Laplace's method briefly. In Section III, Gaussian process prediction in the presence of the localization uncertainty has been formulated for a proposed sampling scheme as a Bayesian inference problem. We first present the Monte Carlo estimators for the posterior predictive mean and variance in Section IV. In Section V, using Laplace's method, we provide approximations for the posterior predictive statistics. Section VI compares the computational cost and approximation accuracy over the proposed estimators based on the Mont Carlo sampling and the Laplace's method. Finally the simulation and experimental results are provided in Sections VII and VIII, respectively.

Standard notation will be used throughout the paper. Let $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, $\mathbb{R}_{>0}$, $\mathbb{Z}$, $\mathbb{Z}_{\geq 0}$, $\mathbb{Z}_{>0}$ denote, respectively, the sets of real, non-negative real, positive real, integer, non-negative integer, and positive integer numbers. I_n denotes the identity matrix of size n (I for an appropriate dimension.) For column vectors $v_a \in \mathbb{R}^a$, $v_b \in \mathbb{R}^b$, and $v_c \in \mathbb{R}^c$, $\text{col}(v_a, v_b, v_c) := \begin{bmatrix} v_a^T & v_b^T & v_c^T \end{bmatrix}^T \in \mathbb{R}^{a+b+c}$ stacks all vectors to create one column vector, and $\|v_a\|$ denotes the Euclidean norm (or the vector 2-norm) of v_a. $|A|$ denotes the determinant of a matrix $A \in \mathbb{R}^{n \times n}$, and $\text{tr}(A)$ denotes trace of a square matrix A. Let $\mathbb{E}(z)$ and $\text{Var}(z)$ denote, respectively, the expectation and the variance of a random vector z. A random vector $z \in \mathbb{R}^q$, which is distributed by a multivariate Gaussian distribution of a mean $\mu \in \mathbb{R}^q$ and a variance $\Sigma \in \mathbb{R}^{q \times q}$, is denoted by $z \sim \mathcal{N}(\mu, \Sigma)$. We define the first and the second derivative operators on $h : \mathbb{R}^m \to \mathbb{R}$ with respect to a vector $x \in \mathbb{R}^m$ as follow.

$$\nabla h(x) = \frac{\partial h(x)}{\partial x} = \left(\frac{\partial h(x)}{\partial x_1}, \cdots, \frac{\partial h(x)}{\partial x_m} \right)^T,$$

$$\nabla^2 h(x) = \frac{\partial^2 h(x)}{\partial x \partial x^T} = \begin{pmatrix} \frac{\partial^2 h(x)}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 h(x)}{\partial x_1 \partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 h(x)}{\partial x_m \partial x_1} & \cdots & \frac{\partial^2 h(x)}{\partial x_m \partial x_m} \end{pmatrix}.$$

If there exists c and $k \in \mathbb{R}_{>0}$, such that the approximation error satisfies $\|\hat{x} - x\| \leq k|\epsilon|, \forall \epsilon < c$, we say that the error between $\hat{x}$ and x is of order $O(\epsilon)$ and also write $\hat{x} - x = O(\epsilon)$. Other notation will be explained in due course.

II. Preliminaries

In this section, we review the spatio-temporal Gaussian process, Monte Carlo sampling and Laplace's method, which will be used throughout the paper.

A. Gaussian Process Regression

A Gaussian process defines a distribution over a space of functions and it is completely specified by its mean function and covariance function. Let $x \in \mathcal{Q} := \mathcal{R} \times \mathcal{T} \subset \mathbb{R}^d$ denote the index vector, where $x := [\ r^T \quad t\ ]^T$ contains the sampling location $r \in \mathcal{R} \subset \mathbb{R}^{d-1}$ and the sampling time $t \in \mathcal{T} \subset \mathbb{R}_{\geq 0}$. A Gaussian process, $y(x) \in \mathbb{R}$, is formally defined as below.

Definition 1: A Gaussian process [7] is a collection of random variables, any finite number of which have a joint Gaussian distribution.

For notational simplicity, we consider a zero-mean Gaussian process[1] $y(x) \sim \mathcal{GP}(0, \mathcal{K}(x, x')) \in \mathbb{R}$. For example, one may consider a covariance function defined as

$$\mathcal{K}(x, x') = \sigma^2 \exp\left(-\frac{\|x - x'\|^2}{2\sigma_x^2}\right), \tag{1}$$

where $x, x' \in \mathbb{R}^d$. In general, the mean and the covariance functions of a Gaussian process can be estimated a priori by maximizing the likelihood function [16].

Suppose, we have q noise corrupted observations without localization error, and $\mathcal{D}' = \{(x^{(i)}, \bar{y}^{(i)}) \,|\, i = 1, \cdots, q\}$. Assume that $\bar{y}^{(i)} = y^{(i)} + w^{(i)} \in \mathbb{R}$, where $w^{(i)}$ is an independent and identically distributed (i.i.d.) white Gaussian noise with variance σ_w^2. $\mathbf{x}$ is defined by $\mathbf{x} = \text{col}(x^{(1)}, x^{(2)}, \cdots, x^{(q)})$. The collections of the realizations $\mathbf{y} = [\ y^{(1)} \quad \cdots \quad y^{(q)}\ ]^T \in \mathbb{R}^q$ and the observations $\bar{\mathbf{y}} = [\ \bar{y}^{(1)} \quad \cdots \quad \bar{y}^{(q)}\ ]^T \in \mathbb{R}^q$ have the Gaussian distributions

$$\mathbf{y} \sim \mathcal{N}(0, K(\mathbf{x})), \quad \bar{\mathbf{y}} \sim \mathcal{N}(0, K(\mathbf{x}) + \sigma_w^2 I),$$

where $K(\mathbf{x}) \in \mathbb{R}^{q \times q}$ is the covariance matrix of $\mathbf{y}$ obtained by $K_{ij}(\mathbf{x}) = \mathcal{K}(x^{(i)}, x^{(j)})$, and I is the identity matrix with an appropriate size. We can predict the value $y_\star$ of the Gaussian process at a point $x_\star$ as [7]

$$y_\star | \mathcal{D}' \sim \mathcal{N}(\mu_\star(\mathbf{x}), \sigma_\star^2(\mathbf{x})). \tag{2}$$

In (2), the predictive mean is

$$\mu_\star(\mathbf{x}) = \mathbb{E}(y_\star | \mathcal{D}') = k(\mathbf{x})^T (K(\mathbf{x}) + \sigma_w^2 I)^{-1} \bar{\mathbf{y}}, \tag{3}$$

and the predictive variance is given by

$$\sigma_\star^2(\mathbf{x}) = \text{Var}(y_\star | \mathcal{D}') = \sigma^2 - k(\mathbf{x})^T (K(\mathbf{x}) + \sigma_w^2 I)^{-1} k(\mathbf{x}), \tag{4}$$

where $k(\mathbf{x}) \in \mathbb{R}^q$ is the covariance matrix between $\mathbf{y}$ and $y_\star$ obtained by $k_j(\mathbf{x}) = \mathcal{K}(x^{(j)}, x_\star)$, and $\sigma^2 = \mathcal{K}(x_\star, x_\star) \in \mathbb{R}$ is the variance at $x_\star$. (3) and (4) can be obtained from the fact that

$$\text{col}(y_\star, \bar{\mathbf{y}}) | x_\star, \mathbf{x} \sim \mathcal{N}\left(0, \begin{bmatrix} \sigma^2 & k(\mathbf{x})^T \\ k(\mathbf{x}) & (K(\mathbf{x}) + \sigma_w^2 I) \end{bmatrix}\right). \tag{5}$$

Note that the predictive mean in (3) and its prediction error variance in (4) require the inversion of the covariance matrix whose size depends on the number of observations q. Hence a drawback of Gaussian process regression is that its computational complexity and memory space increase as more measurements are collected, making the method prohibitive for agents with limited memory and computing power. To overcome this increase in complexity, a number of approximation methods for Gaussian process regression have been proposed. In particular, the sparse greedy approximation method [29], the Nystrom method [30], the informative vector machine [31], the likelihood approximation [32], and the Bayesian committee machine [33] have been shown to be effective for many problems.

In particular, it has been proposed that spatio-temporal Gaussian process regression can be applied to truncated observations including only measurements near the target position and time of interest for agents with limited resources [10]. To justify prediction based on only the most recent observations, a similar argument has been made in [34] in the sense that the data from the remote past do not change the predictors significantly under the exponentially decaying correlation functions.

In this paper, we also assume that at each iteration the mobile sensor networks only needs to fuse a fixed number of truncated observations, which are near the target points of interest, to limit the computational resources.

B. Monte Carlo and Importance Sampling

In what follows, we briefly review Monte Carlo and Importance sampling based on [35]. The idea of Monte Carlo simulation is to draw an i.i.d. set of random samples $\{\mathbf{x}^{(i)}\}_{i=1}^s$ from a target density $\pi(\mathbf{x})$, where $\mathbf{x} \in \mathcal{X} = \mathcal{Q}^q$. These s random samples will be used to approximate $\pi(\mathbf{x})$ by an empirical point-mass function $\pi_s(\mathbf{x}) = \frac{1}{s} \sum_{i=1}^{s} \delta(\mathbf{x}^{(i)} - \mathbf{x})$, where $\delta(\cdot)$ denotes the Dirac delta. Consequently, the integrals $I(f)$ can be approximated with tractable sums $I_s(f)$ that converge as follows.

$$I_s(f) = \frac{1}{s} \sum_{i=1}^{s} f(\mathbf{x}^{(i)}) \underset{s \to \infty}{\overset{a.s.}{\to}} I(f) = \int_{\mathcal{X}} f(\mathbf{x}) \pi(\mathbf{x}) d\mathbf{x}.$$

$I_s(f)$ is an unbiased estimator. In addition, it will almost surely asymptotically converge to $I(f)$, which can be proved by the strong law of large numbers. Considering $f : \mathcal{Q}^q \to \mathbb{R}$ for simplicity, if $f(x)$ satisfies $\sigma_f^2 = \mathbb{E}(f^2(\mathbf{x})) - I^2(f) < \infty$, then the variance of the estimator $I_s(f)$ converges to $\frac{\sigma_f^2}{s}$ as s increases. In particular, the central limit theorem provides us with convergence in distribution as follows. $\sqrt{s}\,(I_s(f) - I(f)) \underset{s \to \infty}{\Rightarrow} \mathcal{N}(0, \sigma_f^2)$, where $\Rightarrow$ denotes convergence in distribution.

Importance sampling is a special case of Monte Carlo implementation, having random samples generated from an available distribution rather than the distribution of interest. Let us introduce an arbitrary importance proposal distribution $\rho(\mathbf{x})$ whose support includes the support of $\pi(\mathbf{x})$. We can then rewrite $I(f)$ as $I(f) = \int_{\mathcal{X}} f(\mathbf{x}) \omega(\mathbf{x}) \rho(\mathbf{x}) d\mathbf{x}$, where $\omega(\mathbf{x}) = \frac{\pi(\mathbf{x})}{\rho(\mathbf{x})}$ is known as the *importance weight*. Simulating s i.i.d. samples $\{\mathbf{x}^{(i)}\}_{i=1}^s$ according to $\rho(\mathbf{x})$ and evaluating $\omega(\mathbf{x}^{(i)})$, a possible unbiased Monte Carlo estimate of $I(f)$ is given by $\hat{I}_s(f) = \sum_{i=1}^s f(\mathbf{x}^{(i)}) \omega(\mathbf{x}^{(i)}) / \sum_{i=1}^s \omega(\mathbf{x}^{(i)})$.

Under weak assumptions, the strong law of large numbers applies, yielding $\hat{I}_s(f) \underset{s \to \infty}{\overset{a.s.}{\to}} I(f)$. This integration method

[1] A Gaussian process with a nonzero-mean can be treated by a change of variables. Even without a change of variables, this is not a drastic limitation, since the mean of the posterior process is not confined to zero [7].

can also be interpreted as a sampling method where the posterior density $\pi(\mathbf{x})$ is approximated by $\hat{\pi}_s(\mathbf{x}) = \sum_{i=1}^{s} \omega(\mathbf{x}^{(i)})\delta(\mathbf{x}-\mathbf{x}^{(i)}) / \sum_{i=1}^{s} \omega(\mathbf{x}^{(i)})$, and $\hat{I}_s(f)$ is the integration of $f(\mathbf{x})$ with respect to the empirical measure $\hat{\pi}_s(\mathbf{x})$.

In this paper, we shall compute the ratio of two integrals in the form of

$$G(f) = \frac{\int_{\mathcal{X}} f(\mathbf{x})\omega(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}}{\int_{\mathcal{X}} \omega(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}}, \tag{6}$$

where $\rho(\mathbf{x})$, $f(\mathbf{x})$ and $\omega(\mathbf{x})$ are defined on a space $\mathcal{X}$. To compute $G(f)$ in (6), we use the following approximation as proposed in [36].

$$\hat{G}_s(f) = \frac{\sum_{i=1}^{s} f(\mathbf{x}^{(i)})\omega(\mathbf{x}^{(i)})}{\sum_{i=1}^{s} \omega(\mathbf{x}^{(i)})}, \tag{7}$$

where $\{\mathbf{x}^{(i)} \,|\, i = 1, \cdots, s\}$ is a sequence of i.i.d. random vectors, which is drawn from $\rho(\mathbf{x})$ distribution.

In the following theorem, which can be shown to be a special case of the results from [36], we show the convergence properties of the approximation in (7) as functions of s, the number of random samples [36].

Theorem 2: (Theorems 1 and 2 from [36]) Consider the approximation $\hat{G}_s(f)$ given in (7) to the ratio given $G(f)$ in (6). If $\omega(\mathbf{x})\rho(\mathbf{x})$ is proportional to a proper probability density function defined on $\mathcal{X}$, and $G(f)$, $\int_{\mathcal{X}} \omega(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}$ and $\int_{\mathcal{X}} f^2(x)w^2(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}$ are finite, we have

$$\hat{G}_s(f) \underset{s\to\infty}{\overset{a.s.}{\to}} G(f), \text{ and } \sqrt{s}\left(\hat{G}_s(f) - G(f)\right) \underset{s\to\infty}{\Rightarrow} \mathcal{N}(0, \sigma_g^2),$$

where

$$\sigma_g^2 = \int_{\mathcal{X}} (f(\mathbf{x}) - G(f))^2 w^2(\mathbf{x})\rho(\mathbf{x})d\mathbf{x}.$$

Proof: It is a straightforward application of the results from Theorems 1 and 2 in [36] to (6) and (7). ∎

C. Laplace Approximations

The Laplace method is a family of asymptotic approximations that approximate an integral of a function, i.e., $\int_{\mathcal{X}} f(\mathbf{x})d\mathbf{x}$, where $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^m$, and $m = q \times d$. Let the function $f(\mathbf{x})$ be in a form $f(\mathbf{x}) = e^{-nh(\mathbf{x})}$, where $1 \ll n \in \mathbb{Z}_{>0}$, and $h : \mathcal{X} \to \mathbb{R}$ is a two times continuously differentiable real function on $\mathcal{X}$. Let $\hat{\mathbf{x}}$ denote the exact mode of $-h$, i.e.,

$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}\in\mathcal{X}} -h(\mathbf{x}).$$

Then Laplace's method produces the approximation [24]:

$$\int_{\mathcal{X}} f(\mathbf{x})d\mathbf{x} = \left(\frac{2\pi}{n}\right)^{\frac{m}{2}} |\phi(\hat{\mathbf{x}})|^{\frac{1}{2}} e^{-nh(\hat{\mathbf{x}})} + O(n^{-1}), \tag{8}$$

where $\phi(\hat{\mathbf{x}}) = [\nabla^2 h(\hat{\mathbf{x}})]^{-1}$. The Laplace approximation in (8) will produce reasonable results as long as the $-h$ is unimodal or at least dominated by a single mode.

In practice it might be difficult to find the exact mode of $-h$. A concept of an asymptotic mode is introduced to gauge the approximation error when the exact mode is not used [26].

Definition 3: $\hat{\mathbf{x}}_a$ is called an asymptotic mode of order $O(n^{-k})$ for $-h$ if $\|\hat{\mathbf{x}}_a - \hat{\mathbf{x}}\| \to 0 \quad \text{as} \quad n \to \infty$, and $\nabla h(\hat{\mathbf{x}}_a) = O(n^{-k})$.

Suppose that $\hat{\mathbf{x}}_a$ is an asymptotic mode of order $O(n^{-1})$ for $-h$ and $\{h, \hat{\mathbf{x}}_a\}$ satisfies the regularity conditions given in Appendix A. Then it follows that we have

$$\int_{\mathcal{X}} f(\mathbf{x})d\mathbf{x} = \left(\frac{2\pi}{n}\right)^{\frac{m}{2}} |\phi(\hat{\mathbf{x}}_a)|^{\frac{1}{2}} e^{-nh(\hat{\mathbf{x}}_a)} C_h(\hat{\mathbf{x}}_a) + O(n^{-1}), \tag{9}$$

where $C_h(\hat{\mathbf{x}}_a)$ is given by

$$C_h(\hat{\mathbf{x}}_a) = e^{\frac{n}{2}\nabla h(\hat{\mathbf{x}}_a)^T \phi(\hat{\mathbf{x}}_a) \nabla h(\hat{\mathbf{x}}_a)}.$$

More precise form with the asymptotic mode of order $O(n^{-2})$ is computed for an approximation of order $O(n^{-3})$ in [27].

In many Bayesian inference applications and as in our problem, we need to compute the ratio of two integrals. To this end, *fully exponential Laplace approximations* has been developed by [24] to compute Laplace approximations of the ratio of two integrals, i.e.,

$$\mathcal{M} = \frac{\int_{\mathcal{X}} e^{-nu(\mathbf{x})}d\mathbf{x}}{\int_{\mathcal{X}} e^{-nh(\mathbf{x})}d\mathbf{x}}. \tag{10}$$

If each of $-u$ and $-h$ has a dominant peak at its maximum, then Laplace's method may be directly applied to both the numerator and denominator of (10) separately. If the regularity conditions are satisfied, using (8) for the denominator approximation and (9) for the numerator approximation, Miyata [26] obtained the following approximation and its error order ,

$$\begin{aligned} \hat{\mathcal{M}} &= \frac{|\nabla^2 h(\hat{\mathbf{x}})|^{\frac{1}{2}} e^{nh(\hat{\mathbf{x}})}}{|\nabla^2 u(\hat{\mathbf{x}}_a)|^{\frac{1}{2}} e^{nu(\hat{\mathbf{x}}_a)}} \times C_u(\hat{\mathbf{x}}_a), \\ \mathcal{M} &= \hat{\mathcal{M}} + O(n^{-2}), \end{aligned} \tag{11}$$

where $\hat{\mathbf{x}}$ is the exact mode of $-h$, and $\hat{\mathbf{x}}_a$ is the asymptotic mode of $-u$, and

$$\begin{aligned} C_u(\hat{\mathbf{x}}_a) &= e^{\left(\frac{n}{2}\nabla u(\hat{\mathbf{x}}_a)^T [\nabla^2 u(\hat{\mathbf{x}}_a)]^{-1} \nabla u(\hat{\mathbf{x}}_a)\right)}, \\ \hat{\mathbf{x}}_a &= \hat{\mathbf{x}} - [\nabla^2 u(\hat{\mathbf{x}})]^{-1} \nabla u(\hat{\mathbf{x}}). \end{aligned} \tag{12}$$

Laplace approximations typically has an error of order $O(n^{-1})$ as shown in (8) and (9). On the other hand, fully exponential Laplace approximations for the ratio form yield an error of order $O(n^{-2})$ as shown in (11) since the error terms of order $O(n^{-1})$ in the numerator and denominator cancel each other [24].

Fully exponential Laplace approximations which are presented in (11) is limited for positive functions. Then, Tierney et al. [25] proposed a second-order approximation to the expectation of a general function $g(\mathbf{x})$ (not necessarily positive) by applying the fully exponential method to approximate $\mathcal{M}(\tau) = \mathbb{E}(e^{\tau g(\mathbf{x})}|\mathcal{D})$ and then differentiating the approximated $\mathcal{M}(\tau)$. Consider $\mathcal{M}(\tau)$ is defined as follow

$$\mathcal{M}(\tau) = \frac{\int_{\mathcal{X}} e^{-nu(\mathbf{x})}d\mathbf{x}}{\int_{\mathcal{X}} e^{-nh(\mathbf{x})}d\mathbf{x}} = \frac{\int_{\mathcal{X}} e^{\tau g(\mathbf{x})} e^{-nh(\mathbf{x})}d\mathbf{x}}{\int_{\mathcal{X}} e^{-nh(\mathbf{x})}d\mathbf{x}},$$

where $u(\mathbf{x}) = -\frac{\tau}{n} g(\mathbf{x}) + h(\mathbf{x})$, and $e^{\tau g(\mathbf{x})}$ is positive, while $g(\mathbf{x})$ could be positive or negative. In particular, $\frac{d}{d\tau}\hat{\mathcal{M}}(\tau)$

evaluated at $\tau = 0$ yields a second-order approximation to $\mathbb{E}(g(\mathbf{x}))$ and its order of the error as follow.

$$\mathbb{E}(g(\mathbf{x})) = \frac{d}{d\tau}\mathcal{M}(\tau)\Big|_{\tau=0} = \frac{d}{d\tau}\hat{\mathcal{M}}(\tau)\Big|_{\tau=0} + O(n^{-2}). \quad (13)$$

This method, which is called moment generating function (MGF) Laplace approximation, gives a standard-form approximation using the exact mode of $-h(\mathbf{x})$ [25].

Miyata [26], [27] extended the MGF method for one without computing the exact mode of $-h(\mathbf{x})$. Let $\hat{\mathbf{x}}$ be an asymptotic mode of order $O(n^{-1})$ for $-h(\mathbf{x})$. Suppose that $\{h, \hat{\mathbf{x}}\}$ satisfies the regularity conditions for the asymptotic-mode Laplace method. By using Theorem 5 in [26], the approximation of $\hat{\mathbb{E}}(g(\mathbf{x}))$ and its error order are given as

$$\begin{aligned}\hat{\mathbb{E}}(g(\mathbf{x})) &= g(\hat{\mathbf{x}}) + \frac{1}{2n}\text{tr}\left(\nabla^2 g(\hat{\mathbf{x}})\phi(\hat{\mathbf{x}})\right) \\ &\quad - \frac{1}{2n}\sum_{ijk\ell=1}^{m}\frac{\partial^3 h(\hat{\mathbf{x}})}{\partial\mathbf{x}_i\partial\mathbf{x}_j\partial\mathbf{x}_k}\phi_{i\ell}\phi_{jk}\frac{\partial g(\hat{\mathbf{x}})}{\partial\mathbf{x}_\ell} \\ &\quad - \nabla g(\hat{\mathbf{x}})^T\phi(\hat{\mathbf{x}})\nabla h(\hat{\mathbf{x}}), \\ \mathbb{E}(g(\mathbf{x})) &= \hat{\mathbb{E}}(g(\mathbf{x})) + O(n^{-2}),\end{aligned} \quad (14)$$

where ϕ_{ij} is the i-th row, j-th column element of the matrix $\phi(\hat{\mathbf{x}})$. Furthermore, if $\hat{\mathbf{x}}$ is the exact mode of $-h(\mathbf{x})$, then the approximation has a simpler form because the terms that include $\nabla h(\hat{\mathbf{x}})$ vanish

$$\begin{aligned}\hat{\mathbb{E}}(g(\mathbf{x})) =& g(\hat{\mathbf{x}}) + \frac{1}{2n}\text{tr}\left(\nabla^2 g(\hat{\mathbf{x}})\phi(\hat{\mathbf{x}})\right) \\ &- \frac{1}{2n}\sum_{ijk\ell=1}^{m}\frac{\partial^3 h(\hat{\mathbf{x}})}{\partial\mathbf{x}_i\partial\mathbf{x}_j\partial\mathbf{x}_k}\phi_{i\ell}\phi_{jk}\frac{\partial}{\partial\mathbf{x}_\ell}g(\hat{\mathbf{x}}).\end{aligned} \quad (15)$$

III. The Problem Statement

In practice, $\mathcal{D}'$ (data with perfect localization) is not available due to localization uncertainty, and instead its exact sampling points will be replaced with noise corrupted sampling points. To average out measurement and localization noises, in this paper, we propose to use a sampling scheme in which multiple measurements are taken repeatedly at a set of sampling points of a sensor network. For robotic sensors or mobile sensor networks, this sampling strategy could be efficient and inexpensive since the large energy consumption is usually due to the mobility of the sensor network. Let q sensing agents be indexed by $\mathcal{J} = \{1, \cdots, q\}$. From the proposed sampling scheme, we assume that each agent takes multiple data pairs $\{(\bar{x}^{(i)}, \bar{y}^{(i)}) \,|\, i \in \mathcal{I}\}$, which are indexed by $\mathcal{I} = \{1, \cdots, n\}$ at a set of sampling points by the sensor network $\{x^{(j)} \,|\, j \in \mathcal{J}\}$. We then define the map $\psi : \mathcal{I} \to \mathcal{J}$ that takes the index of the data pair in $\mathcal{I}$ as the input and returns the index of the sensor that produced the data pair as the output. Consider the following realizations using the sampling scheme and the notation just introduced.

$$\begin{aligned}\bar{x}^{(i)} &= x^{(\psi(i))} + v^{(i)} \in \mathbb{R}^d, \quad \forall i \in \mathcal{I} \\ \bar{y}^{(i)} &= y^{(\psi(i))} + w^{(i)} \in \mathbb{R}, \quad \forall i \in \mathcal{I},\end{aligned}$$

where $w^{(i)}$ is an i.i.d. white Gaussian noise with a zero mean and a variance of σ_w^2, i.e., $w^{(i)} \sim \mathcal{N}(0, \sigma_w^2)$ and $v^{(i)}$ is a localization error which has a multivariate normal distribution with a zero mean and a covariance matrix $\Sigma_v \in \mathbb{R}^{d\times d}$, i.e., $v^{(i)} \sim \mathcal{N}(0, \Sigma_v)$. For instance, the distribution of the localization error may be a result of the fusion of GPS and INS measurements [15], or landmark observations and robot's kinematics [37].

To simplify the notation, $\mathcal{D}$ is introduced to denote the data with the measurement noise and localization error as follows.

$$\mathcal{D} = \left\{(\bar{x}^{(i)}, \bar{y}^{(i)}) \,|\, i \in \mathcal{I}\right\}. \quad (16)$$

We also define the collective sampling point vectors with and without localization uncertainty, and the cumulative localization noise vector, respectively by

$$\begin{aligned}\mathbf{x} &= \text{col}(x^{(1)}, x^{(2)}, \cdots, x^{(q)}) \in \mathbb{R}^{dq}, \\ \bar{\mathbf{x}} &= \text{col}(\bar{x}^{(1)}, \bar{x}^{(2)}, \cdots, \bar{x}^{(n)}) \in \mathbb{R}^{dn}, \\ \mathbf{v} &= \text{col}(v^{(1)}, v^{(2)}, \cdots, v^{(n)}) \in \mathbb{R}^{dn}.\end{aligned} \quad (17)$$

From the proposed sampling scheme, to average out the measurement and localization uncertainties, the number of measurements n can be increased without increasing the number of sensors q, and consequently without increasing the dimension of $\mathbf{x} \in \mathbb{R}^{dq}$. Hence, this approach may be efficient for the resource-constrained (mobile) sensor network at the cost of taking more measurements. Using collective sampling point vectors in (17), we have the following relationship.

$$\bar{\mathbf{x}} = L\mathbf{x} + \mathbf{v}, \quad (18)$$

where $L = \bar{L} \otimes I_d \in \mathbb{R}^{dn\times dq}$, $\bar{L} \in \mathbb{R}^{n\times q}$ and $\bar{L}_{ij} = 1$ if $\psi(i) = j$, otherwise $\bar{L}_{ij} = 0$.

The conditional probabilities $\pi(\bar{\mathbf{x}}|\mathbf{x})$ and $\pi(\mathbf{x}|\bar{\mathbf{x}})$ can be expressed as

$$\begin{aligned}\pi(\bar{\mathbf{x}}|\mathbf{x}) =& \frac{1}{|2\pi\Sigma_{\mathbf{v}}|^{\frac{1}{2}}}e^{-\frac{1}{2}(\bar{\mathbf{x}}-L\mathbf{x})^T\Sigma_{\mathbf{v}}^{-1}(\bar{\mathbf{x}}-L\mathbf{x})}, \\ \pi(\mathbf{x}|\bar{\mathbf{x}}) =& \frac{\pi(\bar{\mathbf{x}}|\mathbf{x})\pi(\mathbf{x})}{\int_{\mathcal{X}}\pi(\bar{\mathbf{x}}|\mathbf{x})\pi(\mathbf{x})d\mathbf{x}},\end{aligned}$$

where $\Sigma_{\mathbf{v}} = I_n \otimes \Sigma_v \in \mathbb{R}^{dn\times dn}$. From a Bayesian perspective, we can treat $\mathbf{x}$ as a random vector to incorporate a prior distribution on $\mathbf{x}$. For example, if we assign a prior distribution on $\mathbf{x}$ such that $\mathbf{x} \sim \mathcal{N}(0, \Sigma_{\mathbf{x}})$ then we have

$$\pi(\mathbf{x}|\bar{\mathbf{x}}) = \frac{1}{|2\pi\Sigma_Z|^{\frac{1}{2}}}e^{-\frac{1}{2}(\mathbf{x}-H\bar{\mathbf{x}})^T\Sigma_Z^{-1}(\mathbf{x}-H\bar{\mathbf{x}})}. \quad (19)$$

where $H = \Sigma_Z L^T\Sigma_{\mathbf{v}}^{-1}$ and $\Sigma_Z^{-1} = \Sigma_{\mathbf{x}}^{-1} + L^T\Sigma_{\mathbf{v}}^{-1}L$.

Evaluating posterior predictive statistics such as the density, the mean, and the variance are of critical importance for the sensor network applications.

Therefore, given the data $\mathcal{D}$ in (16), our goal is to compute the posterior predictive statistics. In particular we focused on the following two quantities given in detail.

- The predictive mean estimator (PME) is given by $\mathbb{E}(y_\star|\mathcal{D}) = \int_y y_\star\pi(y_\star|\mathcal{D})dy_\star$.

$$\begin{aligned}\mathbb{E}(y_\star|\mathcal{D}) &= \frac{\int_y y_\star\int_{\mathcal{X}}\pi(y_\star|\mathbf{x},\bar{\mathbf{y}})\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}dy_\star}{\int_{\mathcal{X}}\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}} \\ &= \frac{\int_{\mathcal{X}}\mu_\star(\mathbf{x})\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}}{\int_{\mathcal{X}}\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}},\end{aligned} \quad (20)$$

where $\mu_\star(\mathbf{x})$ is given by (3).

- The predictive variance estimator (PVE) is obtained similarly. From the following equation,

$$\mathbb{E}(y_\star^2|\mathcal{D}) = \frac{\int_{\mathcal{X}} \left(\int_y y_\star^2 \pi(y_\star|\mathbf{x},\bar{\mathbf{y}})dy_\star\right)\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}}{\int_{\mathcal{X}}\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}},$$

where $\int_y y_\star^2\pi(y_\star|\mathbf{x},\bar{\mathbf{y}})dy_\star = \sigma_\star^2(\mathbf{x}) + \mu_\star^2(\mathbf{x})$, we obtain $\mathrm{Var}(y_\star|\mathcal{D})$ given as the following formula.

$$\begin{aligned}\mathrm{Var}(y_\star|\mathcal{D}) =& \frac{\int_{\mathcal{X}}(\sigma_\star^2(\mathbf{x}) + \mu_\star^2(\mathbf{x}))\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}}{\int_{\mathcal{X}}\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}} \\ &- \mathbb{E}^2(y_\star|\mathcal{D}),\end{aligned} \tag{21}$$

where $\sigma_\star^2(\mathbf{x})$ is given by (4).

The main challenge to our problems is the fact that there are no closed-form formulas for the posterior predictive statistics listed in (20), and (21). Therefore, in this paper, approximation techniques will be carefully applied to obtain approximate solutions. In addition, the tradeoffs between the computational complexity and the precision will be investigated for the sensor networks with limited resources.

From (19), one might be tempted to use the best estimate of $\mathbf{x}$, e.g., the conditional expectation of $\mathbf{x}$ for given measured locations $\bar{\mathbf{x}}$, i.e., $\mathbb{E}(\mathbf{x}|\bar{\mathbf{x}})$ for Gaussian process regression. Comparison between these type of quick-and-dirty solutions and the proposed Bayesian approaches will be evaluated and discussed with simulated and experimental data presented in Sections VII and VIII, respectively.

IV. The Monte Carlo method

In this section, we propose importance sampling to compute the posterior predictive statistics. We also summarize the convergence results of the MC estimators based on Theorem 2.

A. Predictive mean

The predictive mean estimator is given by (20). Using importance sampling to approximate (20) leads to the following equation,

$$\hat{\mathbb{E}}(y_\star|\mathcal{D}) = \frac{\sum_{i=1}^s \pi(\bar{\mathbf{y}}|\mathbf{x}^{(i)})\mu_\star(\mathbf{x}^{(i)})}{\sum_{i=1}^s \pi(\bar{\mathbf{y}}|\mathbf{x}^{(i)})}. \tag{22}$$

Theorem 2 and (22) lead to

$$\begin{aligned}&\hat{\mathbb{E}}(y_\star|\mathcal{D}) \overset{a.s.}{\underset{s\to\infty}{\to}} \mathbb{E}(y_\star|\mathcal{D}),\\ &\left(\mathbb{E}(y_\star|\mathcal{D}) - \hat{\mathbb{E}}(y_\star|\mathcal{D})\right) \underset{s\to\infty}{\Rightarrow} \mathcal{N}\left(0, \frac{\sigma_m^2}{s}\right),\end{aligned}$$

where $\sigma_m^2 = \int_{\mathcal{X}}(\mu_\star(\mathbf{x}) - \mathbb{E}(y_\star|\mathcal{D}))^2\pi^2(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}$, $\mu_\star(\mathbf{x})$ is calculated from (3), and $\mathbf{x}^{(i)}$ has been sampled from $\pi(\mathbf{x}|\bar{\mathbf{x}})$ given by (19).

B. Predictive variance

For the prediction error variance given by (21), we have

$$\mathrm{Var}(y_\star|\mathcal{D}) = \mathbb{E}\left(\sigma_\star^2(\mathbf{x})\right) + \mathrm{Var}\left(\mu_\star(\mathbf{x})\right).$$

Thus, we propose the following estimator.

$$\begin{aligned}\widehat{\mathrm{Var}}(y_\star|\mathcal{D}) =& \frac{\sum_{i=1}^s \pi(\bar{\mathbf{y}}|\mathbf{x}^{(i)})\sigma_\star^2(\mathbf{x}^{(i)})}{\sum_{i=1}^s \pi(\bar{\mathbf{y}}|\mathbf{x}^{(i)})} \\ &+ \frac{\sum_{i=1}^s \pi(\bar{\mathbf{y}}|\mathbf{x}^{(i)})(\mu_\star(\mathbf{x}^{(i)}) - \hat{\mathbb{E}}(y_\star|\mathcal{D}))^2}{\sum_{i=1}^s \pi(\bar{\mathbf{y}}|\mathbf{x}^{(i)})},\end{aligned} \tag{23}$$

where $\sigma_\star^2(\mathbf{x}^{(i)}) = \mathrm{Var}(y_\star|\mathbf{x}^{(i)},\bar{\mathbf{y}})$, $\mu_\star(\mathbf{x}^{(i)}) = \mathbb{E}(y_\star|\mathbf{x}^{(i)},\bar{\mathbf{y}})$ are obtained from the formulas similar to (4) and (3), and $\hat{\mathbb{E}}(y_\star|\mathcal{D})$ is given by (22).

Applying Theorem 2 for $f(x) = \sigma_\star^2(\mathbf{x}) + (\mu_\star(\mathbf{x}) - \hat{\mathbb{E}}(y_\star|\mathcal{D}))^2$, $\omega(\mathbf{x}) = \pi(\bar{\mathbf{y}}|\mathbf{x})$, and $\rho(\mathbf{x}) = \pi(\mathbf{x}|\bar{\mathbf{x}})$, we obtain the following results.

$$\widehat{\mathrm{Var}}(y_\star|\mathcal{D}) \overset{a.s.}{\underset{s\to\infty}{\to}} \overline{\mathrm{Var}}(y_\star|\mathcal{D}),$$

where $\overline{\mathrm{Var}}(y_\star|\mathcal{D}) = \mathbb{E}(y_\star^2|\mathcal{D}) - 2\mathbb{E}(y_\star|\mathcal{D})\hat{\mathbb{E}}(y_\star|\mathcal{D}) + \hat{\mathbb{E}}^2(y_\star|\mathcal{D})$, $\left(\overline{\mathrm{Var}}(y_\star|\mathcal{D}) - \widehat{\mathrm{Var}}(y_\star|\mathcal{D})\right) \underset{s\to\infty}{\Rightarrow} \mathcal{N}\left(0, \sigma_z^2 s^{-1}\right)$, and

$$\begin{aligned}\sigma_z^2 = &\int_{\mathcal{X}} (\sigma_\star^2(\mathbf{x}) + (\mu_\star(\mathbf{x}) - \hat{\mathbb{E}}(y_\star|\mathcal{D}))^2 - \overline{\mathrm{Var}}(y_\star|\mathcal{D}))^2 \\ &\times \pi^2(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})d\mathbf{x}.\end{aligned}$$

Remark 4: Since $\mathbb{E}(y_\star|\mathcal{D})$ is not available in (23), $\hat{\mathbb{E}}(y_\star|\mathcal{D})$ is used instead. Subsequently, the convergence results for $\widehat{\mathrm{Var}}(y_\star|\mathcal{D})$ are given with respect to $\overline{\mathrm{Var}}(y_\star|\mathcal{D})$ which converges to $\mathrm{Var}(y_\star|\mathcal{D})$ as $s \to \infty$ by the definition.

V. Fully Exponential Laplace Approximations

In this section, we propose fully exponential Laplace approximations to compute the posterior predictive statistics. In the process of applying Laplace approximations, we also obtain the estimation of the sampling points given $\mathcal{D}$ as a by-product. From the observations $\mathcal{D}$ given by (16), we can update the estimates of the sampling points $\mathbf{x}$. To this end, we use the maximum a posteriori probability (MAP) estimate of $\mathbf{x}$ given by

$$\hat{\mathbf{x}}_{MAP} = \arg\max_{\mathbf{x}\in\mathcal{X}} \pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}}). \tag{24}$$

A. Perdictive mean

The predictive mean estimator, given by (20), can be approximated using MGF method (14). To compute $\mathbb{E}(y_\star|\mathcal{D})$, let $g(\mathbf{x}) = \mu_\star(\mathbf{x})$. Using

$$h(\mathbf{x}) = -\frac{1}{n}\ln\left(\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}})\right), \tag{25}$$

the predictive mean estimation using MGF method (MGF-PME), $\hat{\mathbb{E}}(y_\star|\mathcal{D})$ and its error order are given as

$$\begin{aligned}&\hat{\mathbb{E}}(y_\star|\mathcal{D}) = \hat{\mathbb{E}}(g(\mathbf{x})),\\ &\mathbb{E}(y_\star|\mathcal{D}) = \hat{\mathbb{E}}(y_\star|\mathcal{D}) + O(n^{-2}),\end{aligned} \tag{26}$$

where $\hat{\mathbb{E}}(g(\mathbf{x}))$ is given in (14) or (15). However, the MGF-PME given by (26) needs the computation of the third derivative of h, which increases the complexity of the algorithm.

In this paper, another MGF method has been developed in order not to use the third derivative of h. To approximate the derivative of $\mathcal{M}(\cdot)$ at a point τ in (13), we utilize a three-point estimation, which is the slope of a nearby secant line through the points $(\tau-\zeta, \mathcal{M}(\tau-\zeta))$ and $(\tau+\zeta, \mathcal{M}(\tau+\zeta))$. Approximating the derivative in (13) with the three-point estimation, we can avoid the third derivative in (14) or (15). We summarize our results in the following theorem.

Theorem 5: Let $\hat{\mathbf{x}}$ be the exact mode of $-h(\mathbf{x})$. The three-point predictive mean estimator (TP-PME) and its order of the error are given by

$$\begin{aligned}\hat{\mathbb{E}}(y_\star|\mathcal{D}) &= \frac{1}{2}n^{\frac{3}{4}}\left|\nabla^2 h(\hat{\mathbf{x}})\right|^{\frac{1}{2}} e^{nh(\hat{\mathbf{x}})} \\ &\quad\times\Big\{\left|\nabla^2 b(\hat{\mathbf{x}}_b)\right|^{-\frac{1}{2}} C_b(\hat{\mathbf{x}}_b)e^{-nb(\hat{\mathbf{x}}_b)} \\ &\quad-\left|\nabla^2 p(\hat{\mathbf{x}}_p)\right|^{-\frac{1}{2}} C_p(\hat{\mathbf{x}}_p)e^{-np(\hat{\mathbf{x}}_p)}\Big\}, \\ \hat{\mathbb{E}}(y_\star|\mathcal{D}) &= \mathbb{E}(y_\star|\mathcal{D}) + O(n^{-3/2}),\end{aligned} \tag{27}$$

where $h(\mathbf{x})$ is given by (25), and we have used the following definitions

$$\begin{aligned} b(\mathbf{x}) &= h(\mathbf{x}) - n^{-\frac{7}{4}}\mu_\star(\mathbf{x}), \\ p(\mathbf{x}) &= h(\mathbf{x}) + n^{-\frac{7}{4}}\mu_\star(\mathbf{x}), \\ \hat{\mathbf{x}}_b &= \hat{\mathbf{x}} - [\nabla^2 b(\hat{\mathbf{x}})]^{-1}\nabla b(\hat{\mathbf{x}}), \\ \hat{\mathbf{x}}_p &= \hat{\mathbf{x}} - [\nabla^2 p(\hat{\mathbf{x}})]^{-1}\nabla p(\hat{\mathbf{x}}), \\ C_b(\hat{\mathbf{x}}_b) &= e^{\frac{n}{2}\nabla b(\hat{\mathbf{x}}_b)^T[\nabla^2 b(\hat{\mathbf{x}}_b)]^{-1}\nabla b(\hat{\mathbf{x}}_b)}, \\ C_p(\hat{\mathbf{x}}_p) &= e^{\frac{n}{2}\nabla p(\hat{\mathbf{x}}_p)^T[\nabla^2 p(\hat{\mathbf{x}}_p)]^{-1}\nabla p(\hat{\mathbf{x}}_p)}. \end{aligned}$$

Proof: First we use the three-point method to approximate derivative such that

$$\left.\frac{d}{d\tau}\mathcal{M}(\tau)\right|_{\tau=0} = \frac{\mathcal{M}(\zeta)-\mathcal{M}(-\zeta)}{2\zeta} + O(\zeta^2).$$

Plugging $\mathcal{M}(\zeta) = \hat{\mathcal{M}}(\zeta) + O(n^{-2})$ into the above equation and using (13), we obtain

$$\hat{\mathbb{E}}(y_\star|\mathcal{D}) = \frac{\hat{\mathcal{M}}(\zeta)-\hat{\mathcal{M}}(-\zeta)}{2\zeta} + O(n^{-2})O(\zeta^{-1}) + O(\zeta^2). \tag{28}$$

By selecting $\zeta = n^{-3/4}$, we recover the order of the error $O(n^{-2})O(\zeta^{-1}) + O(\zeta^2) = O(n^{-3/2})$. By computing the estimates $\hat{\mathcal{M}}(\zeta)$ and $\hat{\mathcal{M}}(-\zeta)$ in (28) using (11), (27) is obtained. ∎

Remark 6: The complexity of either (14) or (15) is $O(n^4)$ while the complexity of (27) is $O(n^3)$. In return, the error of (15) is of order $O(n^{-2})$ and the error of (27) is of order $O(n^{-3/2})$.

B. Predictive variance

We now apply Laplace's method to approximate the prediction error variance in a similar way. The prediction error variance is given by (21). In this case, $h(\mathbf{x})$ is given by (25) and $u(\mathbf{x}) = -\frac{1}{n}\left(\ln(\sigma_\star^2(\mathbf{x})+\mu_\star^2(\mathbf{x})) + \ln(\pi(\bar{\mathbf{y}}|\mathbf{x})\pi(\mathbf{x}|\bar{\mathbf{x}}))\right)$. Applying (11) to this case, the approximate of $\text{Var}(y_\star|\mathcal{D})$ and its order of the error are given by

$$\begin{aligned}\widehat{\text{Var}}(y_\star|\mathcal{D}) &= \frac{|\nabla^2 h(\hat{\mathbf{x}})|^{\frac{1}{2}}e^{nh(\hat{\mathbf{x}})}}{|\nabla^2 u(\hat{\mathbf{x}}_a)|^{\frac{1}{2}}e^{nu(\hat{\mathbf{x}}_a)}}C_u(\hat{\mathbf{x}}_a) - \hat{\mathbb{E}}^2(y_\star|\mathcal{D}), \\ \widehat{\text{Var}}(y_\star|\mathcal{D}) &= \text{Var}(y_\star|\mathcal{D}) + O(n^{-3/2}),\end{aligned} \tag{29}$$

where $\hat{\mathbf{x}}$ is the exact mode of $-h$, and $\hat{\mathbf{x}}_a$ is the asymptotic mode of $-u$. $C_u(\hat{\mathbf{x}}_a)$ and $\hat{\mathbf{x}}_a$ are given by (12) and $\hat{\mathbb{E}}(y_\star|\mathcal{D})$ is given by (27).

C. Simple Laplace approximations

To minimize the computational complexity, one may prefer a simpler approximation. In this paper, we propose such a simple approximation at the cost of precision, which is summarized in the following theorem.

Theorem 7: Let $\hat{\mathbf{x}}$ be an asymptotic mode of order $O(n^{-1})$ for $-h$ given by (25). Assume that $\{h, \hat{\mathbf{x}}\}$ satisfies the regularity conditions. Consider the following approximations for $\mathbb{E}(y_\star|\mathcal{D})$ and $\text{Var}(y_\star|\mathcal{D})$

$$\hat{\mathbb{E}}(y_\star|\mathcal{D}) = k^T(\hat{\mathbf{x}})(K(\hat{\mathbf{x}}) + \sigma_w^2 I)^{-1}\bar{\mathbf{y}}, \tag{30}$$

$$\widehat{\text{Var}}(y_\star|\mathcal{D}) = \sigma^2 - k^T(\hat{\mathbf{x}})(K(\hat{\mathbf{x}}) + \sigma_w^2 I)^{-1}k(\hat{\mathbf{x}}), \tag{31}$$

where $K(\hat{\mathbf{x}})$ and $k(\hat{\mathbf{x}})$ are covariance matrices as in (3) but obtained with $\hat{\mathbf{x}}$.

We have then the following order of errors.

$$\begin{aligned}\hat{\mathbb{E}}(y_\star|\mathcal{D}) &= \mathbb{E}(y_\star|\mathcal{D}) + O(n^{-1}), \\ \widehat{\text{Var}}(y_\star|\mathcal{D}) &= \text{Var}(y_\star|\mathcal{D}) + O(n^{-1}).\end{aligned}$$

Proof: The approximation for $\mathbb{E}(y_\star|\mathcal{D})$ given by (30) is the first term of (14) neglecting high order terms. The second and third terms in (14) have n^{-1} inside. $\hat{\mathbf{x}}$ is an asymptotic mode of order $O(n^{-1})$ for $-h$. By Definition 3, the last term of (14) that contains $\nabla h(\hat{\mathbf{x}})$ is $O(n^{-1})$. Hence, without these high order terms, we obtain a simpler approximation of order $O(n^{-1})$.

The approximation for $\text{Var}(y_\star|\mathcal{D})$ given by (31) can be obtained by approximating (21) with the first term of the MGF approximation. Let $g(\mathbf{x}) = \sigma_\star^2(\mathbf{x}) + \mu_\star^2(\mathbf{x})$, then $\text{Var}(y_\star|\mathcal{D}) = \mathbb{E}(g(\mathbf{x})|\mathcal{D}) - \mathbb{E}^2(y_\star|\mathcal{D})$. Using the first term of MGF $\mathbb{E}(g(\mathbf{x})|\mathcal{D}) = g(\hat{\mathbf{x}}) + O(n^{-1}) = \sigma_\star^2(\hat{\mathbf{x}}) + \mu_\star^2(\hat{\mathbf{x}}) + O(n^{-1})$, and using $\mathbb{E}^2(y_\star|\mathcal{D}) = \mu_\star^2(\hat{\mathbf{x}}) + O(n^{-1})$, we obtain $\text{Var}(y_\star|\mathcal{D}) = \sigma_\star^2(\hat{\mathbf{x}}) + O(n^{-1})$. The procedure of showing the order of the approximation error is the same as what was shown for the approximation in (30). ∎

Remark 8: Note that the simple Laplace predictive mean and variance estimators in (30) and (31) take the same forms of the original predictive mean and variance without the localization error, respectively given in (3) and (4), evaluated at the MAP estimator of $\mathbf{x}$. As we previously mentioned, $\hat{\mathbf{x}}$ is the mode of $-h$ given by (25) and is the MAP estimator of $\mathbf{x}$, i.e., $\hat{\mathbf{x}} = \hat{\mathbf{x}}_{MAP}$ as defined in (24). Therefore, the difference in the simple Laplace approximations with respect to a quick-and-dirty solution in which the measured location vector $\bar{\mathbf{x}}$ is

TABLE I
ERROR AND COMPLEXITY ANALYSIS

Estimator	Method	Error	Complexity
MC-PME in (22)	Monte Carlo	$O_p\left(\frac{\sigma_m}{\sqrt{s}}\right)$	$O(n^3 \times s)$
MGF-PME in (14)	Laplace MGF	$O(n^{-2})$	$O(n^4)$
TP-PME in (27)	Laplace MGF	$O(n^{-3/2})$	$O(n^3)$
S-PME in (30)	Laplace MGF	$O(n^{-1})$	$O(n^3)$

used is that the simple Laplace approximations use $\hat{\mathbf{x}}$ instead of $\bar{\mathbf{x}}$.

In applying Laplace's method, using the one step Newton gradient method to compute asymptotic modes, e.g., $\hat{\mathbf{x}}_a$ required in (29) or $\hat{\mathbf{x}}_b$ and $\hat{\mathbf{x}}_p$ required in (27) may not satisfy the regularity conditions. In this case, one needs to continue the Newton gradient optimization until the regularity conditions are satisfied.

VI. Error and Complexity Analysis

In this section, we analyze the order of the error and the computational complexity for the proposed approximation methods, which are summarized in Table I. A tradeoff between the approximation error and complexity can be chosen taking into account the performance requirements and constrained resources for a specific sensor network application.

For Laplace's method, the order of the error ranges from $O(n^{-1})$ to $O(n^{-2})$ at the cost of complexity from $O(n^3)$ to $O(n^4)$ as shown in Table I.

For the Monte Carlo estimators, we introduce $O_p(\sigma_s)$, which is the probabilistic error order and it implies that the estimation error converges to $\mathcal{N}(0, \sigma_s^2)$ in distribution as the number of random samples s increases. Therefore it is not appropriate to compare the error bound between Monte Carlo and Laplace's method exactly. Monte Carlo algorithms are used for multivariate integration of dimension $d \times q$. The probabilistic error order of Monte Carlo algorithms that use s sample evaluations is of order $s^{-1/2}$ for a given n. We may assume that the order of the error for Monte Carlo methods do not depend on the number of measurements n for a large n [36]. With this assumption, the number of samples needed for Monte Carlo algorithms to reduce the initial probabilistic error by n^{-2} is of order $s = n^4$. The complexity of the Monte Carlo methods, for the investigated problems in this paper, is $O(n^3 \times s)$. To achieve the probabilistic error order $O_p(n^{-2})$, the complexity of Monte Carlo methods has to be $O(n^7)$. If we want to keep the error order at the level of $O(n^{-2})$ and $O_p(n^{-2})$ for Laplace's and Monte Carlo methods, respectively, the Monte Carlo methods are slightly more expensive than Laplace's method.

VII. Simulation Results

In this section, we provide simulation results to evaluate the performances of different estimation methods. To this end, a realization of a Gaussian process that will serve as ground truth is shown in Fig. 2. The Gaussian process is generated for $\sigma = \sqrt{2}$ and $\sigma_x = \sqrt{2}$ in (1). The measurement noise and the sampling position uncertainty variance are $\sigma_w = 0.01$ and $\Sigma_v = \sqrt{0.1} \times I$, respectively. $q = 20$ and $n = 40$ imply that

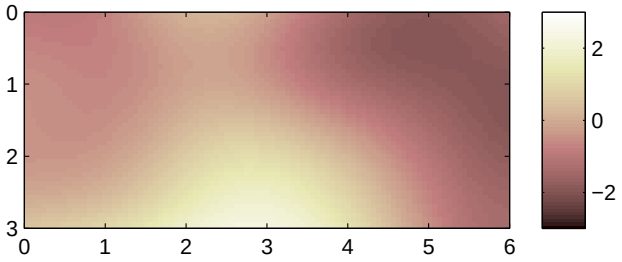


Fig. 2. A realization of the Gaussian process-ground truth.

20 robot takes measurements twice at each sampling position. In Figs. 2-7, the predicted fields and predicted error variance fields are shown with color bars.

The results from Gaussian process regression using the noiseless positions $\mathbf{x}$ and the noisy measurement $\bar{\mathbf{y}}$ are shown in Fig. 3.

To compare with typical quick-and-dirty solutions (QDS) to deal with noisy locations $\bar{\mathbf{x}}$ in practice, we define two solutions: QDS 1 and 2 approaches. QDS 1 is to applying Gaussian process regression given by (3) and (4) by simply ignoring noises in the locations and taking $\bar{\mathbf{x}}$ as the true positions, i.e.,

$$\mu_\star(\bar{\mathbf{x}}) = k^T(\bar{\mathbf{x}})(K(\bar{\mathbf{x}}) + \sigma_w^2 I)^{-1}\bar{\mathbf{y}}. \tag{32}$$

When the measurements are taken repeatedly as suggested in Section III, QDS 1 could be improved. In this regard, QDS 2 is to use the conditional expectation of sampling positions $\mathbf{x}$ given $\bar{\mathbf{x}}$ as in (19) and the least squares solution of $\mathbf{y}$ for given $\bar{\mathbf{y}}$, which shall be plugged into Gaussian process regression, i.e.,

$$\mu_\star(H\bar{\mathbf{x}}) = k^T(H\bar{\mathbf{x}})(K(H\bar{\mathbf{x}}) + \sigma_w^2 I)^{-1}[(\bar{L}^T\bar{L})^{-1}\bar{L}^T\bar{\mathbf{y}}], \tag{33}$$

where H is from (19) and $\bar{L}$ is from (18). QDS 2 might be an improved version of QDS 1 when there are many repeated measurements for a set of fixed sampling positions. Figs. 4 and 5 show the results of applying QDS 1 and QDS 2 on $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$, respectively. This averaging helps the QDS approach to generate smoother predictions, and it shows improvement with respect to QDS 1.

The results of the Monte Carlo method with $s = 1000$ samples are shown in Fig. 6. The results of Laplace's method are shown in Fig. 7, and they look very similar to those of the Monte Carlo methods in Fig. 6. Figs. 4-7 clearly shows that both the Monte Carlo and Laplace's method outperform QDS 1 and 2 with respect to the true field in Fig. 3.

To numerically quantify the performance of each approach, we have computed the root mean square (RMS) error between the predicted and true fields for all methods, which are summarized in Table II. This RMS error analysis could be done since we know the true realization of the Gaussian process exactly in this simulation study. As expected, Gaussian process regression with the true locations perform best. The proposed approaches, i.e., Monte Carlo and Laplace's method outperform QDS 1 and 2 in terms of RMS errors as well.

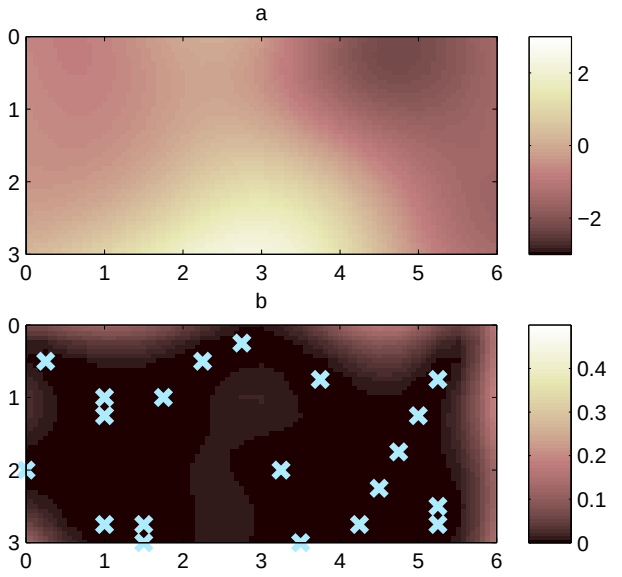


Fig. 3. Gaussian process regression using the true positions $\mathbf{x}$. a) The predictive mean estimation. b) The predictive variance estimation, and the true sampling positions $\mathbf{x}$ in aquamarine crosses.

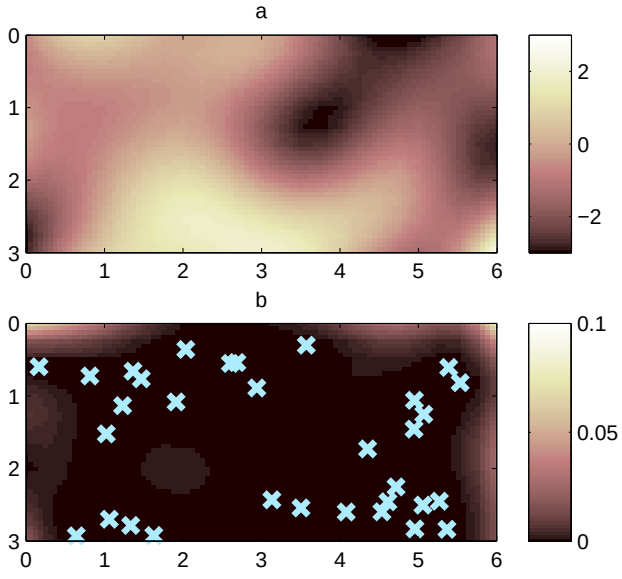


Fig. 4. The results of the QDS 1 approximations using $\bar{\mathbf{x}}$. a) The predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

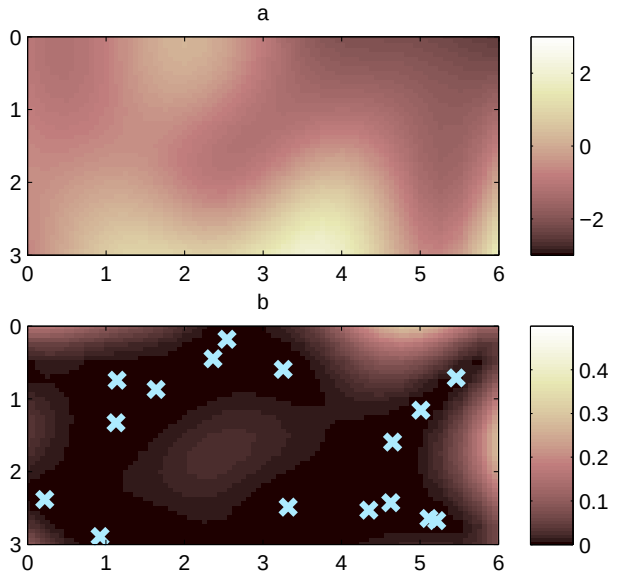


Fig. 5. The results of the QDS 2 approximations using $\mathbb{E}(\mathbf{x}|\bar{\mathbf{x}})$. a) The predictive mean estimation. b) The predictive variance estimation, and $\mathbb{E}(\mathbf{x}|\bar{\mathbf{x}})$ shown with aquamarine crosses.

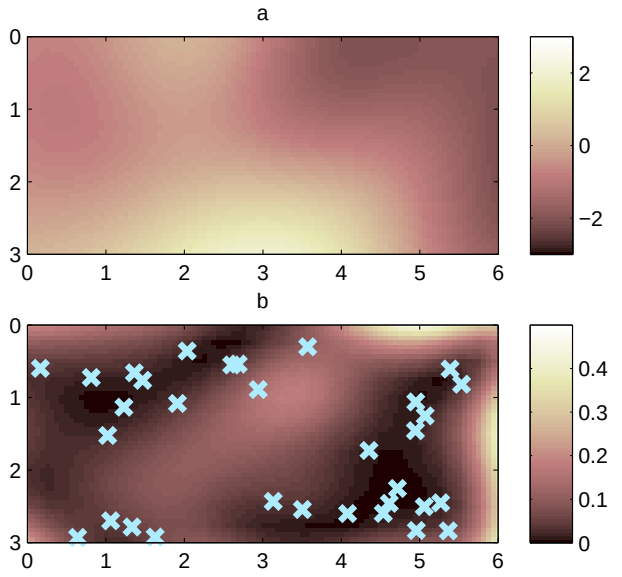


Fig. 6. The results of the Monte Carlo approximations with 1000 samples using $\bar{\mathbf{x}}$. a) The predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses..

VIII. Experimental Results

A. Experiment with simulated noisy sampling positions

In this section, we apply proposed prediction algorithms to a real experimental data set to model the concentration of 7α, 12α, 24-trihydroxy-5α-cholan-3-one 24-sulfate (3kPZS), a synthesized sea lamprey (*Petromyzon marinus*) mating pheromone, in the Ocqueoc River, MI, USA which the authors of [38] provided. The sea lamprey is an ecologically damaging vertebrate invasive fish invader of the Laurentian Great Lakes [39] and a sea lamprey management program has been established [40]. A recent study by Johnson et al. [38] showed that synthesized 3kPZS, a synthesized component of the male mating pheromone, when released into a stream to reach concentrations of $10^{-14} - 10^{-10}$ M (molar) or mol/L, triggers robust upstream movement in ovulated females drawing $\approx$ 50% into baited traps. The ability to predict 3kPZS concentration at any location and time with a predicted uncertainty level would allow for fine-scale evaluations of hypothesized sea lamprey chemo-orientation mechanisms such as odor-conditioned rheotaxis [38]. Here 3kPZS was added to the stream to produce pulsing excitation to sea lampreys by applying 3kPZS at two minutes intervals [41]. To describe

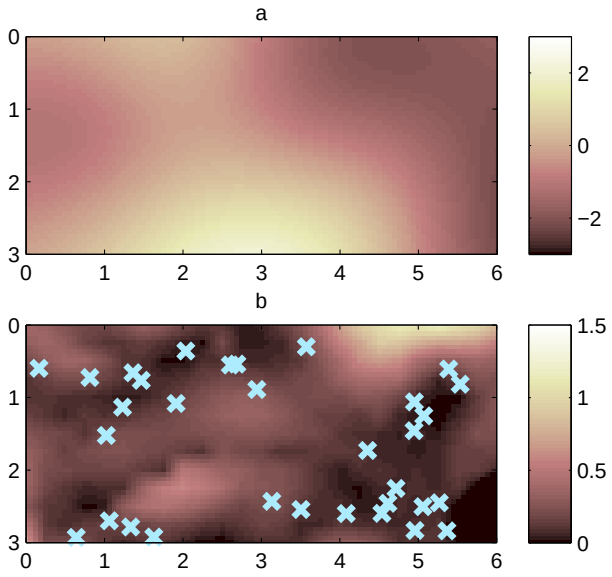


Fig. 7. The results of the Laplace approximations using $\bar{\mathbf{x}}$. a) Predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

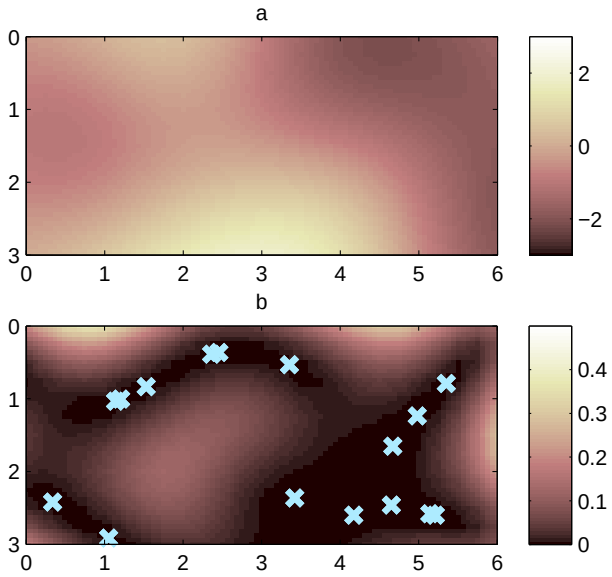
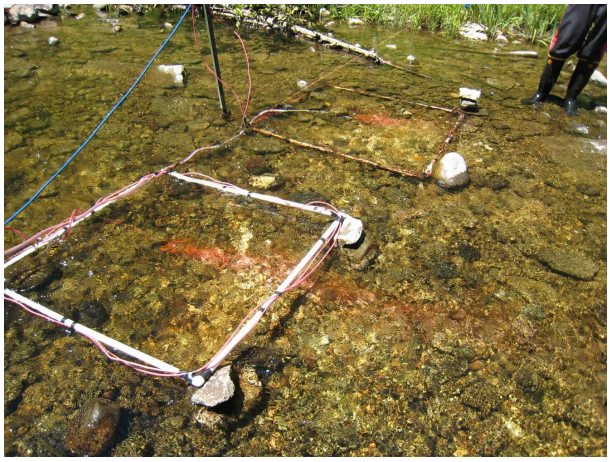


Fig. 8. The results of the simple Laplace approximations using $\bar{\mathbf{x}}$. a) The predictive mean estimation. b) The predictive variance estimation, and $\hat{\mathbf{x}}$ shown with aquamarine crosses. $\hat{\mathbf{x}}$ is the estimation of the true sampling positions, computed as a by-product of both fully exponential Laplace and simple Laplace approximations.

3kPZS concentration throughout the experimental stream, rhodamine dye (Turner Designs, Rhodamine WT, Sunnyale, CA, USA) was applied at the pheromone release point (or source location) to reach a final in-stream concentration of 1.0 mug/L (measure the concentration of the 3kPZS). The same pulsing strategy is used when 3kPZS was applied. The dye and 2kPZS pumping systems are shown in Fig. 9-(a). An example of the highly visible dye plume near the source location is shown in Fig. 9-(c).

To quantify dye concentrations in the stream, water samples

(a)



(b)

(c)

Fig. 9. (a) The dye pumping system. (b) An example of normal dye concentration applied to the stream for the pheromone distribution estimation. (c) An example of the highly visible dye plume near the source location.

TABLE II
SIMULATION RMS ERROR

Estimator & Method	Figure	RMS error
Gaussian process regression in (3)	Fig. 3	0.1281
QDS 1 in (32)	Fig. 4	0.9126
QDS 2 in (33)	Fig. 5	0.7374
Monte Carlo method in (22)	Fig. 6	0.3403
Laplace's method in (27)	Fig. 7	0.3320
Simple Laplace method in (30)	Fig. 8	0.3503

were collected along transects occurring every 5 m from the sea lamprey release cage (0 m) to the application location (73 m upstream of release cage) as shown in Fig. 9-(b). Three sampling locations were fixed along each transect at 1/4, 1/2 and 3/4ths the channel width from the left bank. Water was collected from the three sampling sites along a transect simultaneously (three samplers) every 10 sec. for 2 minutes time interval. Further, a series of samples over 2 min were collected exactly 1 meter downstream of the dye application location. Water samples were collected in 5 ml glass vials and subsequently the fluorescence intensity of each sample measured at 556 nm was determined in a luminescence spectrometer (Perkin Elmer LSS55, Downers Grove, IL, USA) and rhodamine concentration was estimated using a standard curve ($R_2 = 0.9998$).

The objective here is to predict the spatio-temporal field of the dye concentration. In fact, the sampling positions are exactly known from the experiment. Therefore, we will intentionally inject some noises in the sampling positions to evaluate the proposed prediction algorithms in this paper. Before applying the proposed algorithms, the hyperparameters such as σ and σ_x were identified from the experimental data by a maximum a posteriori probability (MAP) estimator as described in [16]. On the other hand, the value for σ_w was set to $\sigma_w = 5 \times 10^{-3}$ mug/L according to the noise level from the data sheet of the sensor. We consider the anisotropic covariance function [16] to deal with the case that the correlations along different directions in $x = [x_1, x_2, x_3]^T \in \mathbb{R}^3$ are different. Next, we performed a change of variables such that in a transformed space and time. The covariance function, given (1), could be used for the proposed approaches.

For the case of this experiment, the spatial correlation along the river flow direction is different from the correlation perpendicular to the river flow direction. We consider the following covariance function in the units for the experimental data.

$$\mathcal{K}(x, x') = \sigma^2 \exp\left(-\sum_{j=1}^{3} \frac{(x_j - x'_j)^2}{2\sigma_{xj}^2}\right). \quad (34)$$

Using this covariance function in (34) and computing likelihood function with true value of $\mathbf{x}$, the hyperparameter vector $\theta := [\sigma, \sigma_{x1}, \sigma_{x2}]$ can be computed by the MAP estimator as follow:

$$\theta_{MAP} = \arg\max_{\theta \in \Theta} \pi(\bar{\mathbf{y}}|\theta)\pi(\theta). \quad (35)$$

Using the optimization in (35), we obtained $\sigma = \sqrt{0.3}$ mug/L, $\sigma_{x1} = 1.6$ m and $\sigma_{x2} = 10.7$ m. x_1 and x_2 are the coordinates of the vertical and horizontal (flow direction) axes of Fig. 10.

TABLE III
EXPERIMENT PARAMETERS

Description	Parameter	Value
Number of sensors	q	15
Number of measurements	n	15
The variability at a fixed point	σ	$\sqrt{0.3}$
Bandwidth	σ_{x1} (vertical)	1.6
	σ_{x2} (horizontal)	10.7
Noise covariance matrix	Σ_v	$\begin{bmatrix} 4 & 0 \\ 0 & 0.089 \end{bmatrix}$
Measurement noise level	σ_w	0.005

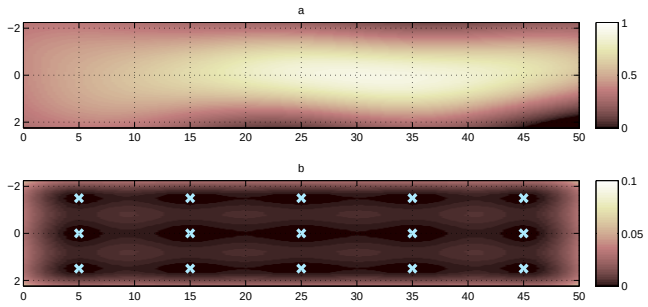


Fig. 10. The results of Gaussian process regression using true $\mathbf{x}$. a) The posterior mean estimation. b) The posterior variance estimation, and $\mathbf{x}$ shown with aquamarine crosses.

As we expect, we have $\sigma_{x1} < \sigma_{x2}$, i.e., the field is more correlated along the river flow direction as compared to the perpendicular to the river flow direction. Next, after finding MAP estimates of hyperparameters σ_{x1} and σ_{x2}, we change the scale of each component of x, such that we have the same values for hyperparameters in the transformed coordinates. The new coordinates are given by $\{x_i^{new} = \frac{\sigma_x}{\sigma_{xi}} x_i\}$ and we recover

$$\mathcal{K}(x, x') = \sigma^2 \exp\left(-\sum_{i=1}^{d} \frac{\left((\frac{\sigma_x}{\sigma_{xi}})x_i - (\frac{\sigma_x}{\sigma_{xi}})x'_i\right)^2}{2\sigma_x^2}\right),$$

as in (1). Note that scaling x_i with $\frac{\sigma_x}{\sigma_{xi}}$, also transforms the covariance matrix Σ_v in a normalized space. Table III shows parameter values which are estimated and used for the experimental data.

Fig. 10 shows the predicted field by applying Gaussian process regression on the true positions $\mathbf{x}$. To illustrate the

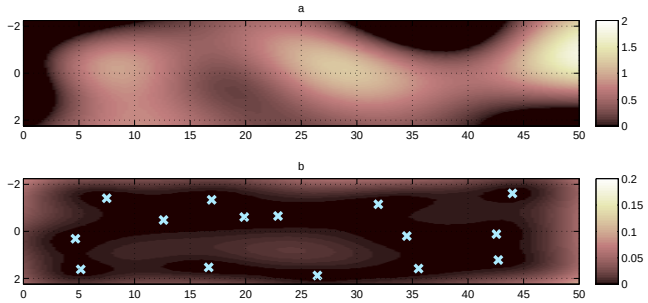


Fig. 11. The results of the QDS 1 approach using noisy positions $\bar{\mathbf{x}}$. a) The predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

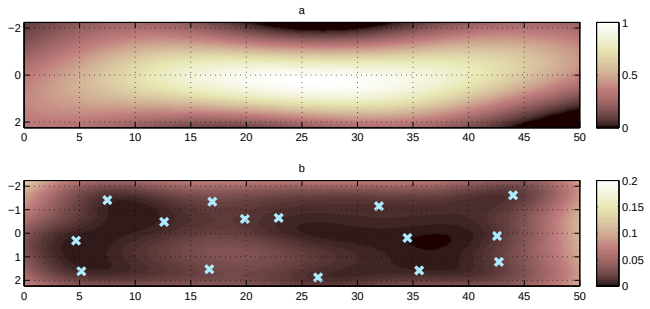


Fig. 12. The results of the Monte Carlo method with 1000 samples using noisy positions $\bar{\mathbf{x}}$. a) The predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

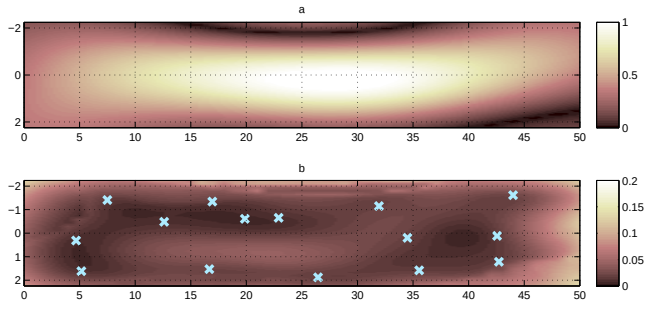


Fig. 13. The results of the Laplace method using noise augmented positions $\bar{\mathbf{x}}$ shown with aquamarine crosses. a) The predictive mean estimation. b) The predictive variance estimation and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

advantage of proposed methods over the QDS approach in dealing with the noisy localizations, we show the results of applying QDS 1, Monte Carlo, Laplace, and simple Laplace approximations in Figs. 11, 12, 13, and 14, respectively. Since there is only one measurement per position, QDS 2 is same as QDS 1. As can be seen in these figures, the peaks of the predicted fields by Monte Carlo and Laplace's method with noisy $\bar{\mathbf{x}}$ are closer to the peak predicted by the true $\mathbf{x}$ than that of the QDS approach. Indeed, QDS 1 has failed in Fig. 11 to produce a good estimation of the true field with neglecting the uncertainty in the positions, while Monte Carlo, Laplace and simple Laplace methods, proposed in this paper, produce good estimations in Figs. 12, 13, and 14, respectively.

Gaussian regression using true $\mathbf{x}$ is our best estimation of the true field (Fig. 10). Table IV shows the RMS errors of the different estimation approaches using the noisy $\mathbf{x}$ (i.e. $\bar{\mathbf{x}}$) with respect to Gaussian regression using the true $\mathbf{x}$. As can be seen in Table IV, the proposed approaches outperform QDS 1 in terms of the RMS error.

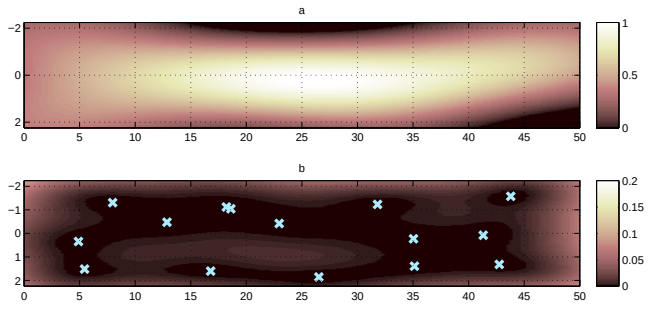


Fig. 14. The results of the simple Laplace method using noise augmented positions $\bar{\mathbf{x}}$ shown with aquamarine crosses. a) The predictive mean estimation. b) The predictive variance estimation and $\hat{\mathbf{x}}$ shown with aquamarine crosses.

TABLE IV
EXPERIMENTAL RMS ERROR W.R.T. GAUSSIAN ESTIMATION

Estimator & Method	Figure	RMS error
QDS 1 in (32)	Fig. 11	0.3955
Monte Carlo method in (22)	Fig. 12	0.1176
Laplace's method in (27)	Fig. 13	0.1305
Simple Laplace method in (30)	Fig. 14	0.1252

B. Experiment with noisy sampling positions

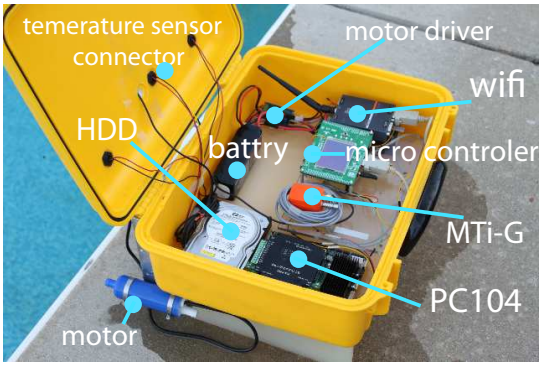
In Section VIII-A we used the true and simulated noisy sampling positions to compare the performance of the proposed approaches. In this section, we present another set of experimental results under real localization errors. The experimental results were obtained using a single aquatic surface robot in an outdoor swimming pool as shown in Fig. 15. The robot is capable of monitoring the water temperature in an autonomous manner while could be remotely supervised by a central station as well. To measure the location of the robot, we used a Xsense MTi-G sensory unit [42] (as shown in the center of Fig. 15-(a)) which consists of an accelerometer, a gyroscope, a magnetometer, and a Global Positioning System (GPS) unit. A Kalman filter is implemented inside of this sensor by the company, which produces the localization of the robot with accuracy of one meter. More information about this experiment can be found in [43]. In this experiment, we have selected and identified the system parameters based on a priori knowledge about the process and sensor noise characteristics. Therefore, model hyperparameters such as $\sigma_x = 2$, $\Sigma_v = I$, $\sigma_w = 0.005$, and $\sigma^2 = 0.3$ are known. The number of sampling positions and sampled measurements are $q = n = 10$. To validate our approaches, we have controlled the temperature field by turning the hot water pump on and off. The hot water outlet locations have been shown in Fig. 15-(b). We turned on the hot water pump for a while. After that, the hot water pump was turned off, and after 6 minutes, the robot collected 10 measurements with 10 sec. time intervals.

The estimated temperature and its error variance fields, by applying QDS 1, Monte Carlo, fully exponential Laplace, and simple Laplace methods are shown in Figs. 16, 17, 18, and 19, respectively. For each method, the estimated temperature and its error variance fields are shown in subfigures of (a) and (b), respectively.

As can be seen in Figs. 17, 18, and 19, results from Monte Carlo, fully exponential Laplace, and simple Laplace methods are well matched.

IX. CONCLUSION

We have formulated Gaussian process regression with observations under the localization uncertainty due to (possibly mobile) sensor networks with limited resources. Effects of the measurements noise, localization uncertainty, and prior distributions have been all correctly incorporated in the posterior



(a)

(b)

Fig. 15. a) The developed robotic sensor. b) The experimental environment- a 12×6 meters outdoor swimming pool.

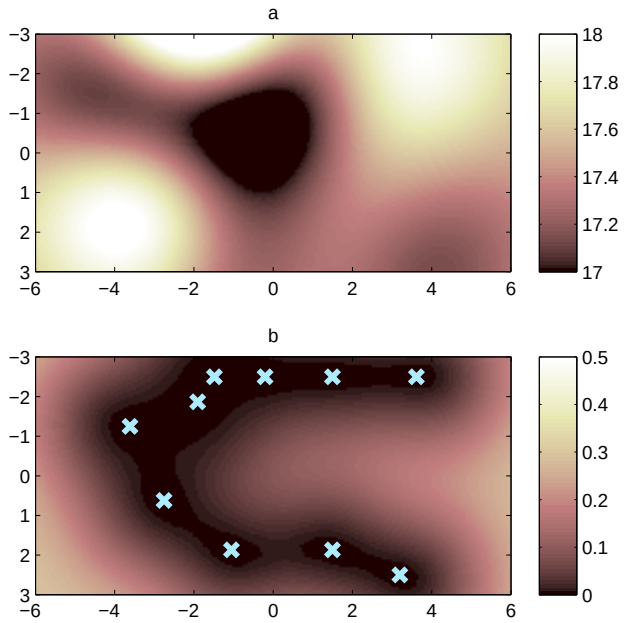


Fig. 16. The results of the QDS 1 approach using noisy positions $\bar{\mathbf{x}}$. a) The predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

predictive statistics in a Bayesian approach. We have reviewed the Monte Carlo sampling and Laplace's method, which have been applied to compute the analytically intractable posterior predictive statistics of the Gaussian processes with localization uncertainty. The approximation error and complexity of all the proposed approaches have been analyzed. In particular, we have provided tradeoffs between the error and complexity of Laplace approximations and their different degrees such that one can choose a tradeoff taking into account the performance requirements and computation complexity due to the resource-constrained sensor network. Simulation study demonstrated that the proposed approaches perform much better than approaches without considering the localization uncertainty properly. Finally, we applied the proposed approaches on the experimentally collected real data to provide proof of concept tests and evaluation of the proposed schemes in practice. From both simulation and experimental results, the proposed methods outperformed the quick-and-dirty solutions often used in practice.

ACKNOWLEDGMENT

This work has been supported, in part, by the National Science Foundation through CAREER Award CMMI-0846547. This article is Contribution 1713 of the U.S. Geological Survey Great Lakes Science Center. In-stream field experiments were supported by the Great Lakes Fishery Commission. W. Li was supported in part by the National Science Foundations Awards IOB0517491 and IOB0450916. These support are gratefully

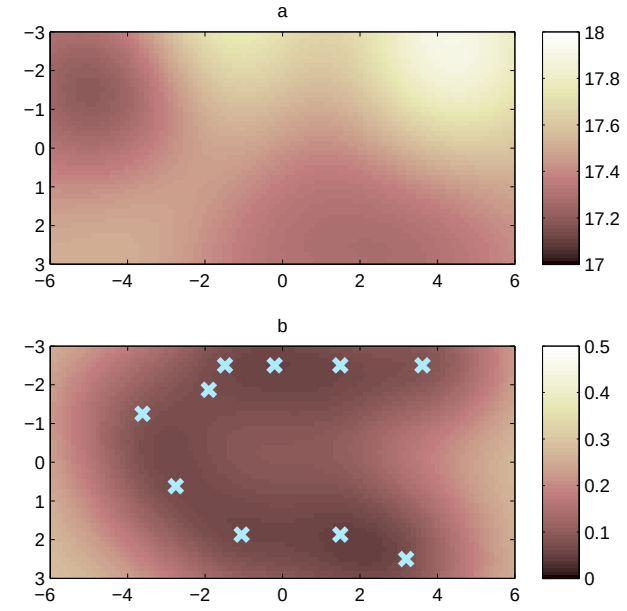


Fig. 17. The results of the Monte Carlo method with 1000 samples using noisy positions $\bar{\mathbf{x}}$. a) The predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

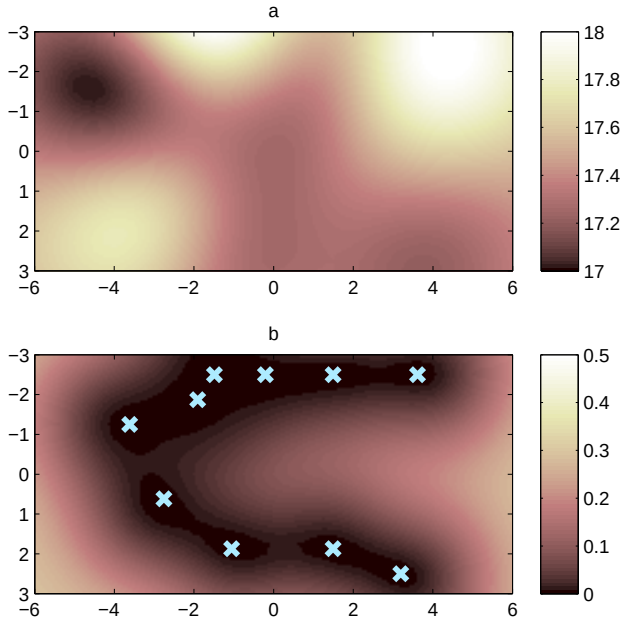


Fig. 18. The results of the Laplace approximations using $\bar{\mathbf{x}}$. a) Predictive mean estimation. b) The predictive variance estimation, and $\bar{\mathbf{x}}$ shown with aquamarine crosses.

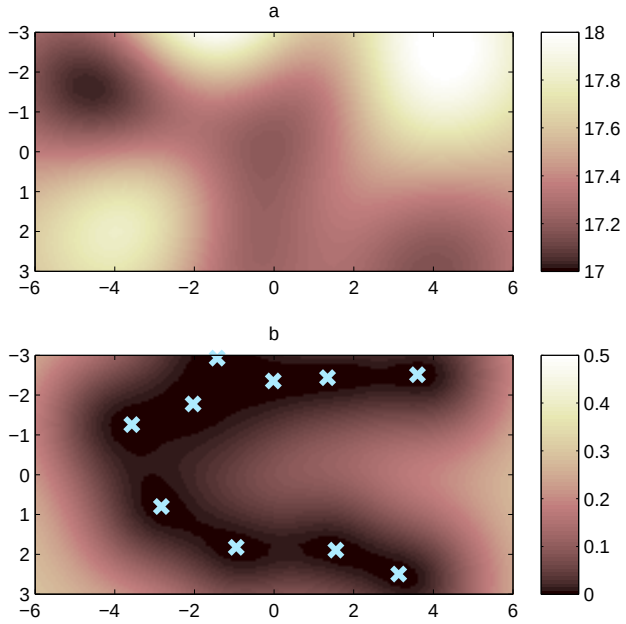


Fig. 19. The results of the simple Laplace method using noise augmented positions $\bar{\mathbf{x}}$ shown with aquamarine crosses. a) The predictive mean estimation. b) The predictive variance estimation and $\hat{\mathbf{x}}$ shown with aquamarine crosses.

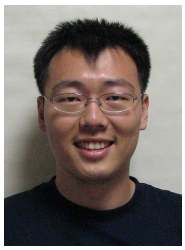
acknowledged. Use of trademark names does not represent endorsement by the US Government.

REFERENCES

[1] D. Culler, D. Estrin, M. Srivastava, Guest editors' introduction: overview of sensor networks, Computer 37 (8) (2004) 41–49.
[2] D. Estrin, D. Culler, K. Pister, G. Sukhatme, Connecting the physical world with pervasive networks, Pervasive computing 1 (1) (2002) 59–69.
[3] K. M. Lynch, I. B. Schwartz, P. Yang, R. A. Freeman, Decentralized environmental modeling by mobile sensor networks, IEEE Transactions on Robotics 24 (3) (2008) 710–724.
[4] J. Choi, S. Oh, R. Horowitz, Distributed learning and cooperative control for multi-agent systems, Automatica 45 (2009) 2802–2814.
[5] J. Cortés, Distributed Kriged Kalman filter for spatial estimation, IEEE Transactions on Automatic Control 54 (12) (2010) 2816–2827.
[6] N. Cressie, Kriging nonstationary data, Journal of the American Statistical Association 81 (395) (1986) 625–634.
[7] C. E. Rasmussen, C. K. I. Williams, Gaussian processes for machine learning, The MIT Press, Cambridge, Massachusetts, London, England, 2006.
[8] A. Krause, C. Guestrin, A. Gupta, J. Kleinberg, Near-optimal sensor placements: maximizing information while minimizing communication cost, in: Proceedings of the 5th international conference on Information processing in sensor networks, ACM, 2006, pp. 2–10.
[9] A. Krause, A. Singh, C. Guestrin, Near-optimal sensor placements in Gaussian processes: theory, efficient algorithms and empirical studies, The Journal of Machine Learning Research 9 (2008) 235–284.
[10] Y. Xu, J. Choi, S. Oh, Mobile sensor network coordination using Gaussian processes with truncated observations, IEEE Transactions on Robotics 27 (6) (2011) 1118–1131.
[11] K. Ho, Alleviating sensor position error in source localization using calibration emitters at inaccurate locations, IEEE Transactions on Signal Processing 58 (1) (2010) 67–83.
[12] L. Hu, D. Evans, Localization for mobile sensor networks, in: Proceedings of the 10th annual international conference on mobile computing and networking, ACM, 2004, pp. 45–57.
[13] P. Oguz-Ekim, J. Gomes, J. Xavier, P. Oliveira, Robust localization of nodes and time-recursive tracking in sensor networks using noisy range measurements, IEEE Transactions on Signal Processing 59 (99) (2011) 3930–3942.
[14] R. Karlsson, F. Gustafsson, Bayesian surface and underwater navigation, IEEE Transactions on Signal Processing 54 (11) (2006) 4204–4213.
[15] D. Titterton, J. Weston, Strapdown inertial navigation technology, Peter Peregrinus Ltd, 2004.
[16] Y. Xu, J. Choi, Adaptive sampling for learning Gaussian processes using mobile sensor networks, Sensors 11 (3) (2011) 3051–3066.
[17] M. Mysorewala, D. Popa, F. Lewis, Multi-scale adaptive sampling with mobile agents for mapping of forest fires, Journal of Intelligent and Robotic Systems 54 (4) (2009) 535–565.
[18] A. Girard, C. Rasmussen, J. Candela, R. Murray-Smith, Gaussian process priors with uncertain inputs-application to multiple-step ahead time series forecasting, Advances in Neural Information Processing Systems (2003) 545–552.
[19] J. Kocijan, R. Murray-Smith, C. Rasmussen, B. Likar, Predictive control with Gaussian process models, in: Proceedings of EUROCON 2003, Vol. 1, IEEE, 2003.
[20] R. Murray-Smith, D. Sbarbaro, C. Rasmussen, A. Girard, Adaptive, cautious, predictive control with Gaussian process priors, in: Proceeding of the 13th IFAC Symposium on System Identification, Rotterdam, Netherlands, 2003, pp. 1195–1200.
[21] M. Deisenroth, C. Rasmussen, J. Peters, Gaussian process dynamic programming, Neurocomputing 72 (7-9) (2009) 1508–1524.
[22] D. Nash, M. Hannah, Using Monte-Carlo simulations and Bayesian Networks to quantify and demonstrate the impact of fertiliser best management practices, Environmental Modelling & Software 26 (9) (2011) 1079–1088.
[23] M. Wainwright, M. Jordan, Graphical models, exponential families, and variational inference, Foundations and Trends® in Machine Learning 1 (1-2) (2008) 1–305.
[24] L. Tierney, J. Kadane, Accurate approximations for posterior moments and marginal densities, Journal of the American Statistical Association 81 (393) (1986) 82–86.
[25] L. Tierney, R. Kass, J. Kadane, Fully exponential Laplace approximations to expectations and variances of nonpositive functions, Journal of the American Statistical Association 84 (407) (1989) 710–716.

[26] Y. Miyata, Fully exponential Laplace approximations using asymptotic modes, Journal of the American Statistical Association 99 (468) (2004) 1037–1049.
[27] Y. Miyata, Laplace approximations to means and variances with asymptotic modes, Journal of Statistical Planning and Inference 140 (2) (2010) 382–392.
[28] M. Jadaliha, Y. Xu, J. Choi, Gaussian process regression using Laplace approximations under localization uncertainty, in: Proceedings of the American Control Conference, 2012.
[29] A. J. Smola, P. L. Bartlett, Sparse greedy Gaussian process regression, in: Proceeding of the Advances in Neural Information Processing Systems 13, 2001, pp. 619–625.
[30] C. Williams, M. Seeger, Using the Nystrom method to speed up kernel machines, in: Proceeding of the Advances in Neural Information Processing Systems 13, MIT Press, 2001, pp. 682–688.
[31] N. Lawrence, M. Seeger, R. Herbrich, Fast sparse Gaussian process methods: The informative vector machine, Advances in Neural Information Processing Systems (2003) 625–632.
[32] M. Seeger, Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations, Ph.D. thesis, School of Informatics, University of Edinburgh (2003).
[33] V. Tresp, A Bayesian committee machine, Neural Computation 12 (11) (2000) 2719–2741.
[34] A. Brix, P. Diggle, Spatiotemporal prediction for log-Gaussian Cox processes, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (4) (2001) 823–841.
[35] C. Andrieu, N. De Freitas, A. Doucet, M. Jordan, An introduction to MCMC for machine learning, Machine learning 50 (1) (2003) 5–43.
[36] J. Geweke, Bayesian inference in econometric models using Monte Carlo integration, Econometrica: Journal of the Econometric Society 57 (6) (1989) 1317–1339.
[37] H. Durrant-Whyte, T. Bailey, Simultaneous localisation and mapping (SLAM): Part I the essential algorithms, Robotics and Automation Magazine 13 (2) (2006) 99–110.
[38] N. Johnson, S. Yun, H. Thompson, C. Brant, W. Li, A synthesized pheromone induces upstream movement in female sea lamprey and summons them into traps, Proceedings of the National Academy of Sciences 106 (4) (2009) 1021.
[39] B. Smith, J. Tibbles, Sea lamprey (petromyzon marinus) in lakes huron, michigan, and superior: history of invasion and control, 1936-78., Canadian Journal of Fisheries and Aquatic Sciences 37 (11) (1980) 1780–1801.
[40] G. Christie, C. Goddard, Sea lamprey international symposium (slis ii): advances in the integrated management of sea lamprey in the great lakes, Journal of Great Lakes Research 29 (2003) 1–14.
[41] N. Johnson, M. Luehring, M. Siefkes, W. Li, Mating pheromone reception and induced behavior in ovulating female sea lampreys, North American journal of fisheries management 26 (1) (2006) 88–96.
[42] Xsens company, MTi-G, miniature AHRS with integreted GPS (2009). URL http://www.xsens.com/images/stories/products/PDF_Brochures/mti-gleaflet.pdf
[43] M. Jadaliha, J. Choi, Environmental monitoring using autonomous aquatic robots: Sampling algorithms and experiments, IEEE Transactions on Control Systems TechnologyIn press, 2012.

Mahdi Jadaliha received his B.S. degree in Electrical Engineering from Iran University of Science and Technology, Tehran, Iran, in 2005. He also received a M.S. degree in Electrical Engineering from Sharif University of Technology, Tehran, Iran, in 2007. Currently, he is a Ph.D. candidate in Mechanical Engineering at the Michigan State University under the guidance of Prof. Jongeun Choi. His research interests include statistical learning and system identification, with applications to mobile robotic sensors and SLAM. His paper was a finalist for the Best Student Paper Award at the Dynamic System and Control Conference (DSCC) 2012.



Yunfei Xu received his Ph.D. degree in Mechanical Engineering from Michigan State University in 2011. He also received his M.S. and B.S. degrees in Automotive Engineering from Tsinghua University, Beijing, China, in 2004 and 2007, respectively. Currently, he is a postdoctoral researcher in the Department of Mechanical Engineering at Michigan State University. His current research interests include environmental adaptive sampling algorithms, Gaussian processes and statistical learning algorithms with applications to robotics and mobile sensor networks.

Jongeun Choi received his Ph.D. and M.S. degrees in Mechanical Engineering from the University of California at Berkeley in 2006 and 2002 respectively. He also received a B.S. degree in Mechanical Design and Production Engineering from Yonsei University at Seoul, Republic of Korea in 1998. He is currently an Associate Professor with the Department of Mechanical Engineering and the Department of Electrical and Computer Engineering at the Michigan State University. His research interests include systems and control, system identification, and Bayesian inference, with applications to mobile robots, environmental adaptive sampling, engine control, and biomedical problems. He was a recipient of an NSF CAREER Award in 2009. His papers were finalists for the Best Student Paper Award at the 24th American Control Conference (ACC) 2005 and the Dynamic System and Control Conference (DSCC) 2011 and 2012. Dr. Choi is a member of ASME.

Nicholas Johnson received the B.S degree from University of Wisconsin Stevens Point in Fisheries/Limnology and Biology in 2004. He also received M.S. and Ph.D. degrees in Fisheries and Wildlife from Michigan State University in 2005 and 2008, respectively. He is currently a Research Ecologist with the United States Geological Survey, Great Lakes Science Center, Hammond Bay Biological Station and Adjunct Assistant Professor in the Department of Fisheries and Wildlife at Michigan State University. His research interests include behavioral and chemical ecology of fishes and specifically how to use chemical and electrical stimuli to aid restoration of valued fishes and control of invasive fishes.

Weiming Li received a B.S. degree in Aquaculture and a M.S. degree in Fish Physiology from Shanghai Ocean University, Shanghai, People's Republic of China in 1982 and 1987, respectively. He also received a Ph.D. degree in Fisheries from the University of Minnesota at Twin Cities, MN in 1994. He is a Professor and FEJ Fry Chair in Environmental Physiology in the Departments of Fisheries and Wildlife and of Physiology at Michigan State University, East Lansing, where he is also a faculty member of the Neuroscience Program and the Ecology, Evolutionary Biology and Behavior Program. His research interests include chemical ecology, physiology of olfaction and endocrine systems, and genomics.

APPENDIX

A. Regularity Conditions

In this section, we review a set of regularity conditions [26]. Let $B_\delta(\hat{\mathbf{x}}_a)$ denote the open ball of radius δ centered at $\hat{\mathbf{x}}_a$, i.e., $B_\delta(\hat{\mathbf{x}}_a) = \{\mathbf{x} \in \mathcal{X} : \|\mathbf{x} - \hat{\mathbf{x}}_a\| < \delta\}$. Let $\hat{\mathbf{x}}_a$ be the asymptotic mode of order n^{-1}. The followings are regularity conditions.

A.1 $h(\mathbf{x}) \in C^6$.

A.2 $\int_{\mathcal{X}} e^{-nh(\mathbf{x})} d\mathbf{x} < \infty$.

A.3 $\|\partial^c h(\mathbf{x})/\partial \mathbf{x}_{j_1} \cdots \partial \mathbf{x}_{j_c}\| < M \in \mathbb{R}_{>0}$, for all $\mathbf{x} \in B_\epsilon(\hat{\mathbf{x}}_a)$ and all $1 \leq j_1, \cdots, j_c \leq m$ with $c = 1, \cdots, 6$, where $\mathbf{x} \in \mathbb{R}^m$.

A.4 $\nabla^2 h(\hat{\mathbf{x}}_a)$ is positive definite and $|\nabla^2 h(\hat{\mathbf{x}}_a)| > \xi \in \mathbb{R}_{>0}$.

A.5 $B_\delta(\hat{\mathbf{x}}_a) \subseteq \mathcal{X}$ and

$$\frac{|n\nabla^2 h(\hat{\mathbf{x}}_a)|^{\frac{1}{2}}}{C(\hat{\mathbf{x}}_a)} \int_{\mathcal{X} - B_\delta(\hat{\mathbf{x}}_a)} e^{-n[h(\mathbf{x}) - h(\hat{\mathbf{x}}_a)]} d\mathbf{x} = O(n^{-2}),$$

for all δ for which $0 < \delta < \epsilon$,

where $C(\hat{\mathbf{x}}_a) = e^{(\frac{n}{2} \nabla h(\hat{\mathbf{x}}_a)(\nabla^2 h(\hat{\mathbf{x}}_a))^{-1} \nabla h(\hat{\mathbf{x}}_a)^T)}$.