

Article

# Fully Bayesian Prediction Algorithms for Mobile Robotic Sensors under Uncertain Localization Using Gaussian Markov Random Fields

Mahdi Jдалиha <sup>1</sup>, Jinho Jeong <sup>2</sup>, Yunfei Xu <sup>3</sup>, Jongeun Choi <sup>2,\*</sup>  and Junghoon Kim <sup>4</sup> <sup>1</sup> Monsanto, St. Louis, MO 63146, USA; jдалиha@gmail.com<sup>2</sup> School of Mechanical Engineering, Yonsei University, Seoul 03722, Korea; jinhojeong@yonsei.ac.kr<sup>3</sup> Denso International America, Inc., San Jose, CA 95110, USA; feiyun.xu@gmail.com<sup>4</sup> Department of Civil and Environmental Engineering, Yonsei University, Seoul 03722, Korea; junghoon@yonsei.ac.kr

\* Correspondence: jongsunchoi@yonsei.ac.kr; Tel.: +82-02-2123-2813

Received: 3 July 2018; Accepted: 26 August 2018; Published: 30 August 2018



**Abstract:** In this paper, we present algorithms for predicting a spatio-temporal random field measured by mobile robotic sensors under uncertainties in localization and measurements. The spatio-temporal field of interest is modeled by a sum of a time-varying mean function and a Gaussian Markov random field (GMRF) with unknown hyperparameters. We first derive the exact Bayesian solution to the problem of computing the predictive inference of the random field, taking into account observations, uncertain hyperparameters, measurement noise, and uncertain localization in a fully Bayesian point of view. We show that the exact solution for uncertain localization is not scalable as the number of observations increases. To cope with this exponentially increasing complexity and to be usable for mobile sensor networks with limited resources, we propose a scalable approximation with a controllable trade-off between approximation error and complexity to the exact solution. The effectiveness of the proposed algorithms is demonstrated by simulation and experimental results.

**Keywords:** Gaussian markov random field; fully Bayesian; mobile sensor network; localization uncertainty

## 1. Introduction

In recent years, there has been an increasing exploitation of mobile robotic sensors in environmental monitoring [1,2]. Gaussian processes defined by mean and covariance functions over a continuum space have been frequently used for mobile sensor networks to statistically model physical phenomena such as harmful algal blooms, pH, and temperature, e.g., [3–5]. The significant computational complexity in Gaussian process regression due to the growing number of observations has been tackled in different ways. Xu et al. [4] analyzed the conditions under which near-optimal prediction can be achieved, using truncated observations when the covariance function is known a priori. Xu and Choi [6] developed a new efficient and scalable inference algorithm for a class of static Gaussian processes that builds on a Gaussian Markov random field (GMRF) is developed for known hyperparameters. In terms of the computational cost reduction, Ref. [7,8] showed that Gaussian processes can be formulated as infinite-dimensional Kalman filtering and such approach can scale down computational complexity. Ref. [9] combined Gaussian process and Kalman filter for efficient computation. On the other hand, unknown hyperparameters in the covariance function can be estimated by a maximum likelihood (ML) estimator or a maximum a posteriori (MAP) estimator and then can be used for the prediction [10]. However, the point estimate itself needs to be identified using a certain amount of measurements and it does not fully incorporate the uncertainty in the estimated hyperparameters into the prediction in a Bayesian perspective.

The advantage of a fully Bayesian approach is the capability of incorporating various uncertainties in the model parameters and measurement processes in the prediction [2]. However, the solution often requires an approximation technique such as Markov Chain Monte Carlo (MCMC), Laplace approximation, or variational Bayes methods, which still requires a high level of computational complexity [2]. Xu et al. [11] designed a sequential Bayesian prediction algorithm and its distributed version for Gaussian processes to deal with uncertain bandwidths. Xu et al. [12] presented sequential fully Bayesian prediction algorithms for a static GMRF with unknown hyperparameters.

Inexpensive wireless/mobile sensor networks [13] are widespread at the cost of precision in localization. Due to their growing usage, there are many practical opportunities where continuously sampled measurements need to be fused for sensors with localization uncertainty [14]. Secure indoor localization has been proposed based on extracting trusted fingerprint [15]. Theoretically-correct yet efficient (or scalable) inference algorithms need to be developed to meet such demands.

Related works involving uncertain localization in our context are as follows. Gaussian process regression has been used in building maps and localization in many practical applications. Brooks et al. [16] utilized Gaussian process regression to model geo-referenced sensor measurements (obtained from a camera) in a supervised learning manner. Kemppainen et al. [17] used Gaussian process regression to implement simultaneous localization and mapping (SLAM) using a magnetic field. O’Callaghan et al. [18] investigated the problem of using laser range-finder data to probabilistically classify a robot’s environment. McHutchon and Rasmussen [19] presented a Gaussian process for training on input points corrupted by independent and identically distributed (i.i.d.) Gaussian noise. Do et al. [20] conducted visual feature selection via Gaussian process regression for position estimation using an omnidirectional camera. The works in [19,20] assume that all hyperparameters are trained offline a priori. Jadalaha et al. [21] and Choi et al. [22] formulated and solved the problem of Gaussian process regression with uncertain localization and known hyperparameters for both centralized and distributed fashions, respectively. A key limitation of such approaches with fixed hyperparameters arises from the fact that, after the initial training phase, learning is discontinued. If the environment changes, it is desirable that the localization algorithm adapts to the changes on the fly. A fully Bayesian approach that treats hyperparameters as random variables can address this issue with increased computational complexity.

The novelty of our work in contrast to ones in [11,12] is to fully consider the uncertainty on the sampling positions along with other uncertainties such as hyperparameters, observation noise, etc., in a *fully Bayesian* manner. The fully Bayesian field SLAM presented in [23] is quite similar to our current work in this paper. However, it is limited to a static random field while ours can deal with the time-varying random field. To the best of our knowledge, fully Bayesian prediction algorithms for spatio-temporal random fields that can take into account uncertain localization are scant to date. Hence, this paper aims to develop such inference algorithms for robotic sensor networks in practical situations by building on spatio-temporal models developed by Lynch et al. [1] and Xu et al. [2]. With continuous improvement in computation power in embedded systems, it is very important to prepare theoretically-correct, and flexible fully Bayesian approach to cope with such practical problems.

The contributions of the paper are as follows. Firstly, we model a physical spatio-temporal random field as a GMRF with uncertain hyperparameters and formulate the prediction problems with and without localization uncertainty. Next, we derive the exact Bayesian solution to the problem of computing the predictive inference of the random field, taking into account uncertain hyperparameters, measurement noise, and uncertain localization in a fully Bayesian point of view. We show that the exact solution for uncertain localization is not scalable as the number of observations increases. To cope with this increasing complexity, we propose a scalable approximation with a controllable trade-off between approximation error and complexity to the exact solution. The effectiveness of the proposed algorithms is demonstrated by experimental results in both static and dynamical environments.

The paper is organized as follows. In Section 2, we explain how a GMRF can be viewed as a sparse and discretized version of a Gaussian process. In Sections 3 and 4, we introduce a spatio-temporal

field model based on a GMRF and the mobile sensor network. In Section 5.1, we present a fully Bayesian inference approach to estimate the spatio-temporal field. The Bayesian prediction algorithm is extended for uncertain sampling positions in Section 5.2. Finally, we evaluate our approach on a real experimental setup in Section 6.

Standard notation is used. 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  $\mu$  and covariance matrix  $\Sigma$ , 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. From Gaussian Processes to Gaussian Markov Random Fields

There are efforts to fit a computationally efficient GMRF on a discrete lattice to a Gaussian random field on a continuum space [12,24,25]. It has been demonstrated that GMRFs with small neighborhoods can approximate Gaussian fields surprisingly well [24]. This approximated GMRF and its regression are efficient and very attractive [26] as compared to the standard Gaussian process and its regression. Fast kriging of large data sets by using a GMRF as an approximation of a Gaussian random field has been proposed in [25].

We now briefly review a GMRF as a discretized Gaussian process on a lattice. Consider a zero-mean Gaussian process:  $z(q) \sim \mathcal{GP}(0, \Sigma(q, q'))$ , 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  $\mathcal{S} := \{s^{[1]}, \dots, s^{[n]}\} \subset \mathbb{R}^2$ , where  $n = hx_{max} \times hy_{max}$ .  $h$  will be chosen such that  $n \in \mathbb{Z}_{>0}$ . Note that  $n \rightarrow \infty$  as  $h \rightarrow \infty$ . The collection of realized values of the random field in  $\mathcal{S}$  is denoted by  $z := (z^{[1]}, \dots, z^{[n]})^T \in \mathbb{R}^n$ , 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

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

where  $\Sigma_0 \in \mathbb{R}^{n \times n}$  is the covariance matrix. The  $i, j$ -th element of  $\Sigma_0$  is defined as  $\Sigma_0^{[ij]} = \text{Cov}(z^{[i]}, z^{[j]}) = \Sigma(z^{[i]}, z^{[j]})$ . 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 Gaussian process (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 a local neighborhood that can represent the original Gaussian process can be found, for example, making  $\hat{Q}_0$  close to  $Q_0$  in some norm [24,25]. 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 Gaussian process precisely.

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

$$y^{[i]} := y(q_c^{[i]}) = z(q_c^{[i]}) + \epsilon^{[i]}, \forall i = 1, \dots, N, \quad (2)$$

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

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

$$z|q_c, y \sim \mathcal{N}(\mu, \Sigma). \quad (3)$$

The predictive mean  $\mu \in \mathbb{R}^n$  and covariance matrix  $\Sigma \in \mathbb{R}^{n \times n}$  can be obtained by  $\mu = K^T C^{-1}y$ ,  $\Sigma = \Sigma_0 - K^T C^{-1}K$ , where the covariance matrices are defined as  $K := \text{Cov}(y, z) \in \mathbb{R}^{N \times n}$ ,  $C := \text{Cov}(y, y) \in \mathbb{R}^{N \times N}$ , and  $\Sigma_0 := \text{Cov}(z, z) \in \mathbb{R}^{n \times n}$ .

The pose of a robot can be estimated by fusing different sensory information producing its estimate and estimation error statistics [27,28]. Throughout the paper, we assume that the positions of mobile robotic sensors and their uncertainties are estimated by a standard technique. Having had the aforementioned assumption, from a localization algorithm, the prior distribution for sampling location  $q_c^{[i]}$  is given as  $\pi(q_c^{[i]}|\tilde{q}_c^{[i]})$ , possibly with a compact support in  $\mathcal{S}_c$ . Then, the predictive distribution of  $z$  given the measured locations  $\tilde{q}_c = (\tilde{q}_c^{[1]T}, \dots, \tilde{q}_c^{[N]T})^T$  is thus given by

$$\pi(z|\tilde{q}_c, y) = \int_{q \in \mathcal{S}_c} \pi(z|q, y) \pi(q|\tilde{q}_c, y) dq, \quad (4)$$

where  $\pi(z|q, y)$  can be obtained by Equation (3). However, the predictive distribution in Equation (4) does not have a closed-form solution and needs to be computed either by MCMC methods or approximation techniques [29].

Now we consider a discretized version of the Gaussian process, i.e., (GMRF) with a precision matrix  $Q_0$  on  $\mathcal{S}$ . Since the sampling points of Gaussian process regression are not necessarily on a finite compact domain  $\mathcal{S}$ , we use the nearest grid point of a given sampling point  $q_c$  in  $\mathcal{S}$   $q^{[i]} = \arg \min_{q \in \mathcal{S}} \|q_c^{[i]} - q\|$ . 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  $q = (q^{[1]T}, \dots, q^{[N]T})^T \in \mathcal{S}^N$  is then obtained by

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

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

$$H^{[ij]} = \begin{cases} 1, & \text{if } s^{[i]} = q^{[j]}, \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

The proof of this posterior distribution of  $z$  in Equation (5) is very similar to that of Proposition 6.1 in Chapter 6 of [2], which was derived by using the Woodbury matrix identity.

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  $q_c^{[i]}$  of the Gaussian process. 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

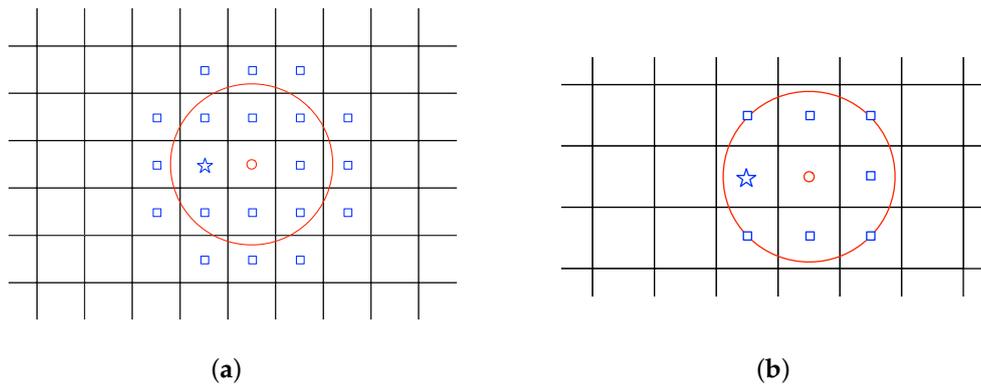
$$\pi(q^{[i]} = s^{[j]}|\tilde{q}^{[i]}) = \int_{s \in \mathcal{V}_j} \pi(s|\tilde{q}^{[i]}) ds, \quad (7)$$

where  $\pi(s|\tilde{q}^{[i]})$  is the continuous prior as in Gaussian process regression and  $\mathcal{V}_j$  is the Voronoi cell of the  $j$ -th grid point  $s^{[j]}$  given by  $\mathcal{V}_j := \{s \in \mathcal{S} \mid \|s - s^{[j]}\| \leq \|s - s^{[i]}\|, \forall i \neq j\}$ . The predictive distribution of  $z$  given  $y$  and  $\tilde{q}$  is thus given by

$$\pi(z|\tilde{q}, y) = \sum_{q \in \mathcal{S}} \pi(z|q, y) \pi(q|\tilde{q}, y), \quad (8)$$

where  $\pi(z|q, y)$  can be obtained by Equation (5) and the summation is over all possible locations in  $\mathcal{S}$ .

Figure 1 shows two examples of using this approximation approach with  $h_1 > h_2$  to convert a continuous space to a discrete one. When  $h \rightarrow \infty$ ,  $\tilde{q} \rightarrow \tilde{q}_c$  and the standard Gaussian process regression in a continuum space shall be recovered from the prediction using the GMRF in a discretized space.



**Figure 1.** Example of localization uncertainty for  $q^{[i]}$ . The measured sampling location  $\tilde{q}^{[i]}$  is indicated in a small red circle which is the closest point in the discrete support to the measured sampling position in the continuous space. 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  $q^{[i]}$  which is the closest point in the discrete support to the true sampling position in the continuous space. **(a)**  $h_1$ ; **(b)**  $h_2$ .

### 3. Mobile Sensor Networks

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

Consider  $N$  spatially distributed mobile sensing agents indexed by  $j \in \mathcal{J} := \{1, \dots, N\}$  sampling at time  $t \in \mathbb{Z}_{>0}$ . At time  $t$ , agent  $j$  takes a noise corrupted measurement at its current location  $q_t^{[j]} = s^{[j]} \in \mathcal{S}$ , i.e.,

$$y_t^{[j]} = z_t^{[i]} + \epsilon_t^{[j]}, \quad \epsilon_t^{[j]} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \sigma_\epsilon^2), \quad (9)$$

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

$$y_t = H_t^T z_t + \epsilon_t, \quad (10)$$

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

$$H_\tau^{[ij]} = \begin{cases} 1, & \text{if } s^{[i]} = q_\tau^{[j]}, \\ 0, & \text{otherwise.} \end{cases} \quad (11)$$

### 4. Spatio-Temporal Field Model

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

$$z_t^{[i]} = \lambda_t^{[i]} + \eta_t^{[i]}, \quad \forall i \in \{1, \dots, n\}, t \in \mathbb{Z}_{>0}. \quad (12)$$

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

$$\lambda_t^{[i]} = f(s^{[i]})^T \beta_t, \quad (13)$$

where  $f(s^{[i]}) = (f_1(s^{[i]}), \dots, f_p(s^{[i]}))^T \in \mathbb{R}^p$  is a known regression function and  $\beta_t = (\beta_t^{[1]}, \dots, \beta_t^{[p]})^T \in \mathbb{R}^p$  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:

$$\beta_{t+1} = A_t \beta_t + B_t \omega_t, \quad (14)$$

where  $\omega_t \sim \mathcal{N}(0, W)$ ,  $\beta_0 \sim \mathcal{N}(\mu_{\beta_0}, \Sigma_{\beta_0})$ , and  $A_t$  and  $B_t$  are known system parameters or can be found as discussed in [30].

In addition, we consider a zero-mean GMRF [31]  $\eta_t = (\eta_t^{[1]}, \dots, \eta_t^{[n]})^T \in \mathbb{R}^n$  whose covariance matrix is given by

$$E(\eta_t \eta_k^T | \theta) = Q_\theta^{-1} \delta(t - k), \quad (15)$$

where  $t$ , and  $k$  are time indices, and  $\delta(\cdot)$  is the Kronecker delta defined by

$$\delta(k) = \begin{cases} 1, & k = 0, \\ 0, & \text{otherwise,} \end{cases} \quad (16)$$

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

There are different parameterizations of the GMRF (i.e., the precision matrix  $Q_\theta$ ) [31]. 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  $\eta$  be a GMRF on a regular two-dimensional lattice. The associated Gaussian full conditional mean is

$$E(\eta_t^{[i]} | \eta_t^{[-i]}, \theta) = -\frac{1}{Q_\theta^{[ii]}} \sum_{j=1}^n Q_\theta^{[ij]} \eta_t^{[j]}, \quad (17)$$

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  $\eta_t$  values everywhere except  $s^{[i]}$ . The hyperparameter vector is defined as  $\theta = (\kappa, \alpha)^T \in \mathbb{R}_{>0}^2$ , where  $\alpha = a - 4$ . The value of  $Q_\theta^{[ii]}$  is  $4 + a^2$  as denoted at the center node of the graph. That of  $Q_\theta^{[ij]}$  is  $-2a$  if  $j$  is one of the four closest neighbors of  $i$  in the vector 1-norm sense. 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), as shown in [32]. The equation in Equation (17) states that the conditional expectation of  $\eta_t^{[i]}$  given the value of  $\eta_t$  everywhere else (i.e.,  $\eta_t^{[-i]}$ ) can be determined just by knowing the value of  $\eta_t$  on the twelve closest neighbors (see more details in [32]). The resulting GMRF accurately represents a Gaussian random field with the Matérn covariance function as shown in [32]

$$G(r) = \sigma_f^2 \frac{2^{1-\rho}}{\Gamma(\rho)} \left( \frac{\sqrt{2\rho}r}{\ell} \right)^\rho K_\rho \left( \frac{\sqrt{2\rho}r}{\ell} \right), \quad (18)$$

where  $K_\rho(\cdot)$  is a modified Bessel function [33], 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 hyperparameter  $\alpha > 0$  guarantees the positive definiteness of the precision matrix  $Q_\theta$ . In the case where  $\alpha = 0$ , the resulting GMRF is a second-order polynomial intrinsic GMRF [31,34].

From the presented model in Equations (12), (14), and (15), the distribution of  $z_t$  given  $\beta_t$  and  $\theta$  is

$$z_t | \beta_t, \theta \sim \mathcal{N}(F_s \beta_t, Q_\theta^{-1}), \quad (19)$$

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

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  $\Sigma_\theta$  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  $\Sigma_\theta^{[ij]} = \text{Cov}(z_t^{[i]}, z_t^{[j]})$ , 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

$$\begin{aligned} \mu_{z_t|\mathcal{B}_t} &= F_s\beta_t + \Sigma_\theta H_t \left( H_t^T \Sigma_\theta H_t + \sigma_c^2 I \right)^{-1} \left( y_t - H_t^T F_s \beta_t \right), \\ \Sigma_{z_t|\mathcal{B}_t} &= \Sigma_\theta - \Sigma_\theta H_t \left( H_t^T \Sigma_\theta H_t + \sigma_c^2 I \right)^{-1} H_t^T \Sigma_\theta. \end{aligned} \quad (20)$$

The basic idea behind the model introduced in Equations (12), (14), and (15) stems from the space-time Kalman filter model proposed in [35]. The advantage of this spatio-temporal model with known hyperparameters is to make inferences in a recursive manner as the number of observations increases.

In this paper, however, uncertainties in the precision matrix and sampling positions are considered in a fully Bayesian manner. In contrast to [35,36], the GMRF with a sparse precision matrix is used to increase the computational efficiency.

## 5. Fully Bayesian Predictive Inference

### 5.1. Uncertain Hyperparameters and Exact Localization

In this section, we consider the problem of predicting a spatio-temporal random field, using successive noisy measurements sampled by a mobile sensor network. For a known covariance function, the prediction can be shown to be recursive [37] based on Gaussian process regression. The uncertainty in  $\theta$  in a GMRF has been considered and its sequential prediction algorithms are derived in [12]. However, only the static field has been considered, i.e.,  $\mu_t = \mu_0$ . In this section, we use a Bayesian approach 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 1–5.

**Assumption 1.** *The spatio-temporal random field is generated by Equations (12), (14), and (15);*

**Assumption 2.** *The precision matrix  $Q_\theta$  is a given function of an uncertain hyperparameter vector  $\theta$ ;*

**Assumption 3.** *The noisy measurements  $\{y_t\}$ , as in Equation (10), are continuously collected by mobile robotic sensors in time  $t$ ;*

**Assumption 4.** *The sample positions  $\{q_t\}$  are measured precisely by mobile robotic sensors in time  $t$ ;*

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

A.1 and A.2 stem from the discretization of a Gaussian process as we described in Section 2. From the model in A.1, the zero-mean GMRF 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. A.3 is a standard assumption over the measured observations [36]. A.4 will be relaxed to A.6 in Section 5.2 to deal with localization uncertainty. A.5 is from the discretization of the hyperparameter vector to replace an integration with a summation over possible hyperparameters.

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 1.** Consider the Assumptions 1–5. Our problem is to find the predictive distribution, mean, and variance of  $x_t$  conditional on  $\mathcal{D}_{1:t}$ .

We then summarize the intermediate steps to obtain the solution to Problem 1. In what follows, our results are presented in terms of lemmas and theorems. We provide proofs when they are not straightforward.

**Lemma 1.** Under Assumptions 1 and 2, the predictive distribution of  $x_t$  conditional on the hyperparameter vector  $\theta$  and the measurements  $\mathcal{D}_{1:t-1}$  is Gaussian with the following mean and precision matrix:

$$\begin{aligned} \mu_{x_t|\theta, \mathcal{D}_{1:t-1}} &= \begin{pmatrix} F_s \mu_{\beta_t|\theta, \mathcal{D}_{1:t-1}} \\ \mu_{\beta_t|\theta, \mathcal{D}_{1:t-1}} \end{pmatrix}, \\ Q_{x_t|\theta, \mathcal{D}_{1:t-1}} &= \begin{pmatrix} Q_\theta & -Q_\theta F_s \\ -F_s^T Q_\theta & F_s^T Q_\theta F_s + \Sigma_{\beta_t|\theta, \mathcal{D}_{1:t-1}}^{-1} \end{pmatrix}, \end{aligned} \tag{21}$$

where  $\mu_{\beta_t|\theta, \mathcal{D}_{1:t-1}} = A_t \mu_{\beta_{t-1}|\theta, \mathcal{D}_{1:t-1}}$  denotes the expectation of  $\beta_t$  conditional on  $\theta$  and  $\mathcal{D}_{1:t-1}$  and  $\Sigma_{\beta_t|\theta, \mathcal{D}_{1:t-1}} = A_t \Sigma_{\beta_{t-1}|\theta, \mathcal{D}_{1:t-1}} A_t^T + B_t W B_t^T$  denotes the associated estimation error covariance matrix.

**Proof of Lemma 1.** It can be shown by the update using the prior precision matrix and the previous iteration as in [12] (see more details in Section 3 of [12]). □

For a given hyperparameter vector  $\theta$ , Equation (21) provides the optimal prediction of the spatio-temporal field in time  $t$  using data up to time  $t - 1$ .

The following lemma is used to compute the posterior distribution of  $\theta$ , recursively.

**Lemma 2.** Under Assumptions 3 and 4, the posterior distribution of the hyperparameter vector  $\theta$  can be obtained recursively via

$$\pi(\theta|\mathcal{D}_{1:t}) \propto \pi(y_t|\theta, \mathcal{D}_{1:t-1}, q_t) \pi(\theta|\mathcal{D}_{1:t-1}), \tag{22}$$

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

$$\begin{aligned} \mu_{y_t|\theta, \mathcal{D}_{1:t-1}, q_t} &= \Gamma_{q_t}^T \mu_{x_t|\theta, \mathcal{D}_{1:t-1}}, \\ \Sigma_{y_t|\theta, \mathcal{D}_{1:t-1}, q_t} &= \Gamma_{q_t}^T \Sigma_{x_t|\theta, \mathcal{D}_{1:t-1}} \Gamma_{q_t} + \sigma_\epsilon^2 I, \end{aligned} \tag{23}$$

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

**Proof of Lemma 2.** The posterior distribution of  $\theta$  given in Equation (22) is computed by applying Bayes' rule on  $\pi(\theta|y_t, \mathcal{D}_{1:t-1})$ . The predictive statistics of  $y_t|\theta, \mathcal{D}_{1:t-1}$  are straightforward results of using Equation (10). Note that  $\pi(\theta|\mathcal{D}_{1:t-1})$  is equal to  $\pi(\theta|\mathcal{D}_{1:t-1}, q_t)$ . □

**Lemma 3.** Under Assumptions 1–4, the full conditional distribution of  $x_t$  for a given hyperparameter vector and data up to time  $t$  is

$$x_t|\theta, \mathcal{D}_{1:t} \sim \mathcal{N}(\mu_{x_t|\theta, \mathcal{D}_{1:t}}, Q_{x_t|\theta, \mathcal{D}_{1:t}}^{-1}),$$

where

$$\begin{aligned} Q_{x_t|\theta, \mathcal{D}_{1:t}} &= Q_{x_t|\theta, \mathcal{D}_{1:t-1}} + \sigma_\epsilon^{-2} \Gamma_{q_t} \Gamma_{q_t}^T, \\ \mu_{x_t|\theta, \mathcal{D}_{1:t}} &= \mu_{x_t|\theta, \mathcal{D}_{1:t-1}} + \sigma_\epsilon^{-2} Q_{x_t|\theta, \mathcal{D}_{1:t}}^{-1} \Gamma_{q_t} (y_t - \Gamma_{q_t}^T \mu_{x_t|\theta, \mathcal{D}_{1:t-1}}). \end{aligned} \tag{24}$$

In order to keep computing the prediction error covariance matrix  $Q_{x_t|\theta, \mathcal{D}_{1:t}}^{-1}$  alone, the Woodbury lemma could be used to reduce the computational load as follows:

$$Q_{x_t|\theta, \mathcal{D}_{1:t}}^{-1} = Q_{x_t|\theta, \mathcal{D}_{1:t-1}}^{-1} - Q_{x_t|\theta, \mathcal{D}_{1:t-1}}^{-1} \Gamma_{q_t} \left( \sigma_\epsilon^2 I + \Gamma_{q_t}^T Q_{x_t|\theta, \mathcal{D}_{1:t-1}}^{-1} \Gamma_{q_t} \right)^{-1} \Gamma_{q_t}^T Q_{x_t|\theta, \mathcal{D}_{1:t-1}}^{-1}, \quad (25)$$

where  $Q_{x_t|\theta, \mathcal{D}_{1:t-1}}^{-1}$  can be computed with blockwise inversion using Equation (21),

$$Q_{x_t|\theta, \mathcal{D}_{1:t-1}}^{-1} = \begin{pmatrix} Q_\theta^{-1} + F_s \Sigma_{\beta_t|\theta, \mathcal{D}_{1:t-1}} F_s^T & F_s \Sigma_{\beta_t|\theta, \mathcal{D}_{1:t-1}} \\ \Sigma_{\beta_t|\theta, \mathcal{D}_{1:t-1}}^T F_s^T & \Sigma_{\beta_t|\theta, \mathcal{D}_{1:t-1}} \end{pmatrix}.$$

The blockwise inversion needs to be updated only with  $\Sigma_{\beta_t|\theta, \mathcal{D}_{1:t-1}}$ .

The following theorem explicitly illustrates how the results of Lemmas 2 and 3 lead to the predictive statistics of  $x_t$  under Assumptions 1–5, which will be the solution to Problem 1.

**Theorem 1.** Under Assumption 5, the predictive distribution of  $x_t|\mathcal{D}_{1:t}$  is given by

$$\pi(x_t|\mathcal{D}_{1:t}) = \sum_{\theta \in \Theta} \pi(x_t|\theta, \mathcal{D}_{1:t}) \pi(\theta|\mathcal{D}_{1:t}), \quad (26)$$

where  $\pi(\theta|\mathcal{D}_{1:t})$  and  $\pi(x_t|\theta, \mathcal{D}_{1:t})$  are given by Lemmas 2 and 3, respectively. The predictive mean and variance follow as

$$\begin{aligned} \mu_{x_t|\mathcal{D}_{1:t}} &= \sum_{\theta \in \Theta} \mu_{x_t|\theta, \mathcal{D}_{1:t}} \pi(\theta|\mathcal{D}_{1:t}), \\ \Sigma_{x_t|\mathcal{D}_{1:t}} &= \sum_{\theta \in \Theta} \left[ \Sigma_{x_t|\theta, \mathcal{D}_{1:t}} + (\mu_{x_t|\theta, \mathcal{D}_{1:t}} - \mu_{x_t|\mathcal{D}_{1:t}}) (\mu_{x_t|\theta, \mathcal{D}_{1:t}} - \mu_{x_t|\mathcal{D}_{1:t}})^T \right] \pi(\theta|\mathcal{D}_{1:t}). \end{aligned} \quad (27)$$

**Proof of Theorem 1:** The predictive mean and variance is obtained by marginalizing over the conditional distribution of  $\theta$  given  $\mathcal{D}_{1:t}$ . The marginal mean and variance are  $E_Y(Y) = E_X(E_X(Y|X))$  and  $\text{Var}_Y(Y) = E_X(\text{Var}(Y|X)) + \text{Var}_X(E(Y|X))$ , where  $E_X$  and  $\text{Var}_X$  denote the expectation and the variance with respect to the random variable  $X$ . Having  $Y := x_t|\mathcal{D}_{1:t}$  and  $X := \theta|\mathcal{D}_{1:t}$  completes the proof of Theorem 1.  $\square$

The optimal prediction of the spatio-temporal field  $x_t|\theta, \mathcal{D}_{1:t-1}$  using predictive statistics of  $x_{t-1}|\theta, \mathcal{D}_{1:t-1}$  is provided by Lemma 1. Lemma 3 provides the optimal estimator of  $x_t|\theta, \mathcal{D}_{1:t}$ , using predictive statistics of  $x_t|\theta, \mathcal{D}_{1:t-1}$  which is given by Lemma 1. Using Lemmas 1 and 3 sequentially, we can update predictive statistics of  $x_t|\theta, \mathcal{D}_{1:t}$  for known hyperparameters. Lemma 2 gives us the posterior distribution of  $\theta$  based on the measured data. Finally, Theorem 1 provides the optimal Bayesian prediction of the spatio-temporal random field with a time varying mean function and uncertain hyperparameters by marginalizing  $\theta$  over the conditional distribution of  $\theta|\mathcal{D}_{1:t}$ .

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

---

**Algorithm 1** Sequential Bayesian Predictive Inference

---

Initialization:  
 1: initialize  $F_s$   
 2: for  $\theta \in \Theta$ , initialize  $Q_\theta$ , and compute  $Q_\theta^{-1}$   
 At time  $t \in \mathbb{Z}_{>0}$ , do:  
 1: obtain new observations  $y_t$  at current locations  $q_t$   
 2: find the map  $\Gamma_{q_t}$  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  $\mu_{x_t|\theta, \mathcal{D}_{1:t-1}}$  and  $Q_{x_t|\theta, \mathcal{D}_{1:t-1}}$  using measurements up to time  $t - 1$ , given by Equation (21).  
 5:   compute  $\mu_{x_t|\theta, \mathcal{D}_{1:t}}$  and  $Q_{x_t|\theta, \mathcal{D}_{1:t}}$  given by Equation (24).  
 6:   compute  $\mu_{y_t|\theta, \mathcal{D}_{1:t-1}, q_t}$  and  $\Sigma_{y_t|\theta, \mathcal{D}_{1:t-1}, q_t}$  by Equation (23).  
 7:   calculate  $\pi(\theta|\mathcal{D}_{1:t})$  given by Equation (22).  
 8: **end for**  
 9: compute the predictive mean and variance using Equation (27).

---

5.2. 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 this section, the proposed method in the previous section will be extended for the uncertain localization data.

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

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

A straightforward consequence of Assumption 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,  $\mathcal{I}(k : r) := \prod_{g=k}^r \mathcal{I}(g)$  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 2.** Consider Assumptions 1–3, 5, and 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  $\mathcal{R}_{r:k} := \{\mathcal{P}_{r-1}, y_{r:k}\}$ . We then have  $\mathcal{R}_{r:k} \subset \mathcal{D}_{r:k}$ , where we recall that  $\mathcal{D}_{r:k} := \{\mathcal{P}_{r-1}, q_{r:k}, y_{r:k}\}$ .

To solve Problem 2, we first look for a way to compute the posterior distribution of  $q_t$  as summarized in the following lemma.

**Lemma 4.** Consider  $\pi(y_t|\theta, \mathcal{D}_{1:t-1}^{(n)}, q_t^{(k)})$  given by Equation (23) and  $\pi(\theta|\mathcal{D}_{1:t-1}^{(n)})$  given by Equation (22), where  $n \in \mathcal{I}(1 : t - 1)$ ,  $k \in \mathcal{I}(t)$ , and  $\mathcal{D}_{1:t-1}^{(n)} := \{\mathcal{P}_0, q_{1:t-1}^{(n)}, y_{1:t-1}\}$ . Under Assumption 5, we have

$$\pi(y_t|\mathcal{D}_{1:t-1}^{(n)}, q_t^{(k)}) = \sum_{\theta \in \Theta} \pi(y_t|\theta, \mathcal{D}_{1:t-1}^{(n)}, q_t^{(k)}) \pi(\theta|\mathcal{D}_{1:t-1}^{(n)}).$$

Under Assumption 6, the posterior distribution of  $q_t$  can be obtained, recursively, via

$$\pi(q_{1:t-1}^{(n)}, q_t^{(k)} | \mathcal{R}_{1:t}) \propto \pi(q_{1:t-1}^{(n)} | \mathcal{R}_{1:t-1}) \pi(y_t | \mathcal{D}_{1:t-1}^{(n)}, q_t^{(k)}) \pi(q_t^{(k)}). \tag{28}$$

We now give the exact solution to Problem 2 as follows.

**Theorem 2.** Consider the predictive distribution  $\pi(x_t|\mathcal{D}_{1:t}^{(i)})$  given by Theorem 1 and the posterior  $\pi(q_{1:t}^{(i)}|\mathcal{R}_{1:t})$  given by Lemma 4, where  $q_{1:t}^{(i)} \in \Omega(1:t)$  and  $\mathcal{D}_{1:t}^{(i)} = \{\mathcal{P}_0, q_{1:t}^{(i)}, y_{1:t}\}$ . Under Assumption 6, the predictive distribution of  $x_t|\mathcal{R}_{1:t}$  can be obtained as follows:

$$\pi(x_t|\mathcal{R}_{1:t}) = \sum_{i \in \mathcal{I}(1:t)} \pi(x_t|\mathcal{D}_{1:t}^{(i)}) \pi(q_{1:t}^{(i)}|\mathcal{R}_{1:t}). \quad (29)$$

Consequently, the predictive mean and variance are given by the formulas:

$$\begin{aligned} \mu_{x_t|\mathcal{R}_{1:t}} &= \sum_{i \in \mathcal{I}(1:t)} \mu_{x_t|\mathcal{D}_{1:t}^{(i)}} \pi(q_{1:t}^{(i)}|\mathcal{R}_{1:t}), \\ \Sigma_{x_t|\mathcal{R}_{1:t}} &= \sum_{i \in \mathcal{I}(1:t)} \left[ \Sigma_{x_t|\mathcal{D}_{1:t}^{(i)}} + \left( \mu_{x_t|\mathcal{D}_{1:t}^{(i)}} - \mu_{x_t|\mathcal{R}_{1:t}} \right) \left( \mu_{x_t|\mathcal{D}_{1:t}^{(i)}} - \mu_{x_t|\mathcal{R}_{1:t}} \right)^T \right] \pi(q_{1:t}^{(i)}|\mathcal{R}_{1:t}). \end{aligned} \quad (30)$$

**Proof of Theorem 2:** The proof is similar to that of Theorem 1. Hence, the predictive mean and variance are obtained by marginalizing over the conditional distribution of  $q_{1:t}$ . The marginal mean and variance are  $E_Y(Y) = E_Y(E_X(Y|X))$  and  $\text{Var}_Y(Y) = E_X(\text{Var}(Y|X)) + \text{Var}_X(E(Y|X))$ , where  $E_X$  and  $\text{Var}_X$  denote the expectation and the variance with respect to the random variable  $X$ . Having  $Y := x_t|\mathcal{R}_{1:t}$  and  $X := q_{1:t}|\mathcal{R}_{1:t}$  proves Theorem 2.  $\square$

The complexity of the proposed algorithm in Theorem 2 is proportional to the number of possibilities for  $q_{1:t}$ . The result of Lemma 4 enables us to compute these probable possibilities recursively. However, the number of the non-zero probable combinations grows exponentially by a power of time index  $t$ .

In what follows, we propose an approximation, with a controllable trade-off between approximation error and complexity, to the exact solution given in Theorem 2 by including an option of ignoring uncertainties on past position data. This approximation will include a case of the exact solution with the maximal and original 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. To formulate our idea clearly, we present first the following results.

**Lemma 5.** 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

$$\pi(q_{t-m+1:t-1}^{(j)}, q_t^{(k)}|\mathcal{R}_{t-m+1:t}) \propto \pi(q_{t-m+1:t-1}^{(j)}|\mathcal{R}_{t-m+1:t-1}) \pi(y_t|\mathcal{D}_{t-m+1:t-1}^{(j)}, q_t^{(k)}) \pi(q_t^{(k)}), \quad (31)$$

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

**Theorem 3.** Consider  $\mu_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}}$  and  $\Sigma_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}}$  computed by Theorem 1. Under Assumption 6, the predictive statistics of  $x_t|\mathcal{R}_{t-m+1:t}$  are as follows:

$$\begin{aligned} \mu_{x_t|\mathcal{R}_{t-m+1:t}} &= \sum_{h \in \mathcal{I}(t-m+1,t)} \mu_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}} \pi(q_{t-m+1:t}^{(h)}|\mathcal{R}_{t-m+1:t}), \\ \Sigma_{x_t|\mathcal{R}_{t-m+1:t}} &= \sum_{h \in \mathcal{I}(t-m+1,t)} \left[ \Sigma_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}} + \left( \mu_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}} - \mu_{x_t|\mathcal{R}_{t-m+1:t}} \right) \left( \mu_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}} - \mu_{x_t|\mathcal{R}_{t-m+1:t}} \right)^T \right] \pi(q_{t-m+1:t}^{(h)}|\mathcal{R}_{t-m+1:t}). \end{aligned} \quad (32)$$

To implement approximations to the predictive statistics of  $x_t|\mathcal{R}_{1:t}$  which are given by Theorem 2, we consider the following conditions.

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

$$\pi(x_t|\mathcal{R}_{1:t}) \approx \pi(x_t|\mathcal{R}_{t-m+1:t}). \quad (33)$$

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

$$\mathcal{P}_t \approx \{\mu_{x_t|\mathcal{R}_{t-m+1:t}}, \Sigma_{x_t|\mathcal{R}_{t-m+1:t}}\} \cup \{\pi(\theta|\mathcal{R}_{t-m+1:t})|\theta \in \Theta\}. \tag{34}$$

Under conditions C.1 and C.2, it is natural for us to propose the following approximations:

$$\mu_{x_t|\mathcal{R}_{1:t}} \approx \mu_{x_t|\mathcal{R}_{t-m+1:t}}, \quad \Sigma_{x_t|\mathcal{R}_{1:t}} \approx \Sigma_{x_t|\mathcal{R}_{t-m+1:t}}. \tag{35}$$

In Theorem 3, the predictive statistics of  $x_t|\mathcal{D}_{t-m+1:t}^{(h)}$  are obtained from Algorithm 1, which is given in Section 5.1. The only difference is that we start from time  $t - m + 1$  instead of time 1 with  $\mathcal{P}_{t-m}$  instead of  $\mathcal{P}_0$ . Note that, without condition C.2, we cannot use Algorithm 1 to calculate the statistics of  $x_t|\mathcal{D}_{t-m+1:t}^{(h)}$ . The proposed approximation for the case of uncertain localization in Equation (35) is quite different from a mere truncation of old data in the sense that past measurements still affect the current estimation through the approximately updated prior information using Equation (34). Note that we update the prior information from  $\mathcal{P}_{t-m}$  to  $\mathcal{P}_t$  with the cumulative data collected from time  $t - m + 1$  up to time  $t$ , which is different from only using truncated observations. The proposed approximation for the formulated problem is summarized by Algorithm 2.

To further understand the nature of the proposed approximation, consider the following two extreme special cases.

**Corollary 1.** As a special case of Theorem 3 for  $m = 1$ , the posterior distribution of  $q_t$  can be obtained via

$$\pi(q_t^{(k)}|\mathcal{P}_{t-1}, y_t) \propto \pi(y_t|\mathcal{P}_{t-1}, q_t^{(k)})\pi(q_t^{(k)}), \tag{36}$$

where  $k \in \mathcal{I}(t)$ . The predictive distribution of  $x_t|\mathcal{R}_{1:t}$  can be approximated in a constant time as time  $t$  increases in a sequential way.

$$\pi(x_t|\mathcal{R}_{1:t}) \approx \pi(x_t|\mathcal{P}_{t-1}, y_t),$$

where

$$\pi(x_t|\mathcal{P}_{t-1}, y_t) = \sum_{k \in \mathcal{I}(t)} \pi(x_t|\mathcal{P}_{t-1}, q_t^{(k)}, y_t) \pi(q_t^{(k)}|\mathcal{P}_{t-1}, y_t) \tag{37}$$

and the posterior  $\pi(q_t^{(k)}|\mathcal{P}_{t-1}, y_t)$  is given by Equation (36). Consequently, the predictive mean  $\mu_{x_t|\mathcal{R}_{1:t}}$  and variance  $\Sigma_{x_t|\mathcal{R}_{1:t}}$  can be approximated by  $\mu_{x_t|\mathcal{P}_{t-1}, y_t}$  and  $\Sigma_{x_t|\mathcal{P}_{t-1}, y_t}$ , respectively, i.e.,

$$\begin{aligned} \mu_{x_t|\mathcal{P}_{t-1}, y_t} &= \sum_{k \in \mathcal{I}(t)} \mu_{x_t|\mathcal{P}_{t-1}, q_t^{(k)}, y_t} \pi(q_t^{(k)}|\mathcal{P}_{t-1}, y_t), \\ \Sigma_{x_t|\mathcal{P}_{t-1}, y_t} &= \sum_{k \in \mathcal{I}(t)} \left[ \Sigma_{x_t|\mathcal{P}_{t-1}, q_t^{(k)}, y_t} + \left( \mu_{x_t|\mathcal{P}_{t-1}, q_t^{(k)}, y_t} - \mu_{x_t|\mathcal{P}_{t-1}, y_t} \right) \right. \\ &\quad \left. \left( \mu_{x_t|\mathcal{P}_{t-1}, q_t^{(k)}, y_t} - \mu_{x_t|\mathcal{P}_{t-1}, y_t} \right)^T \right] \pi(q_t^{(k)}|\mathcal{P}_{t-1}, y_t). \end{aligned} \tag{38}$$

**Corollary 2.** For another special case with  $m = t$ , Theorem 3 becomes Theorem 2.

For a fixed  $m \in \mathbb{Z}_{>0}$ , Algorithm 2 is scalable as time  $t$  increases. In our approach, the level of the approximation can be controlled by users by selecting a trade-off (or by choosing  $m$ ) between the approximation error and complexity. Most simplistic and practical approximation can be obtained by choosing  $m = 1$  as in Corollary 1. The original exact solution with maximal complexity is recovered by selecting  $m = t$  as shown in Corollary 2.

**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:   predict  $\mu_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}}$ ,  $\Sigma_{x_t|\mathcal{D}_{t-m+1:t}^{(h)}}$  and  $\pi(y_t|\mathcal{D}_{t-m+1:t-1}^{(j)}, q_t^{(k)})$  using Algorithm 1
- 4:   compute  $\pi(q_t^{(h)}|\mathcal{R}_{t-m+1:t})$  by Lemma 5.
- 5: **end for**
- 6: compute  $\mu_{x_t|\mathcal{R}_{t-m+1:t}}$  and  $\Sigma_{x_t|\mathcal{R}_{t-m+1:t}}$  using Theorem 3.
- 7: use following approximation to update estimations

$$\begin{aligned} \mu_{x_t|\mathcal{R}_{1:t}} &\approx \mu_{x_t|\mathcal{R}_{t-m+1:t}}, & \Sigma_{x_t|\mathcal{R}_{1:t}} &\approx \Sigma_{x_t|\mathcal{R}_{t-m+1:t}}, \\ \mathcal{P}_t &\approx \{\mu_{x_t|\mathcal{R}_{t-m+1:t}}, \Sigma_{x_t|\mathcal{R}_{t-m+1:t}}, \pi(\Theta|\mathcal{R}_{t-m+1:t})\}. \end{aligned}$$

### 5.3. Complexity of Algorithms

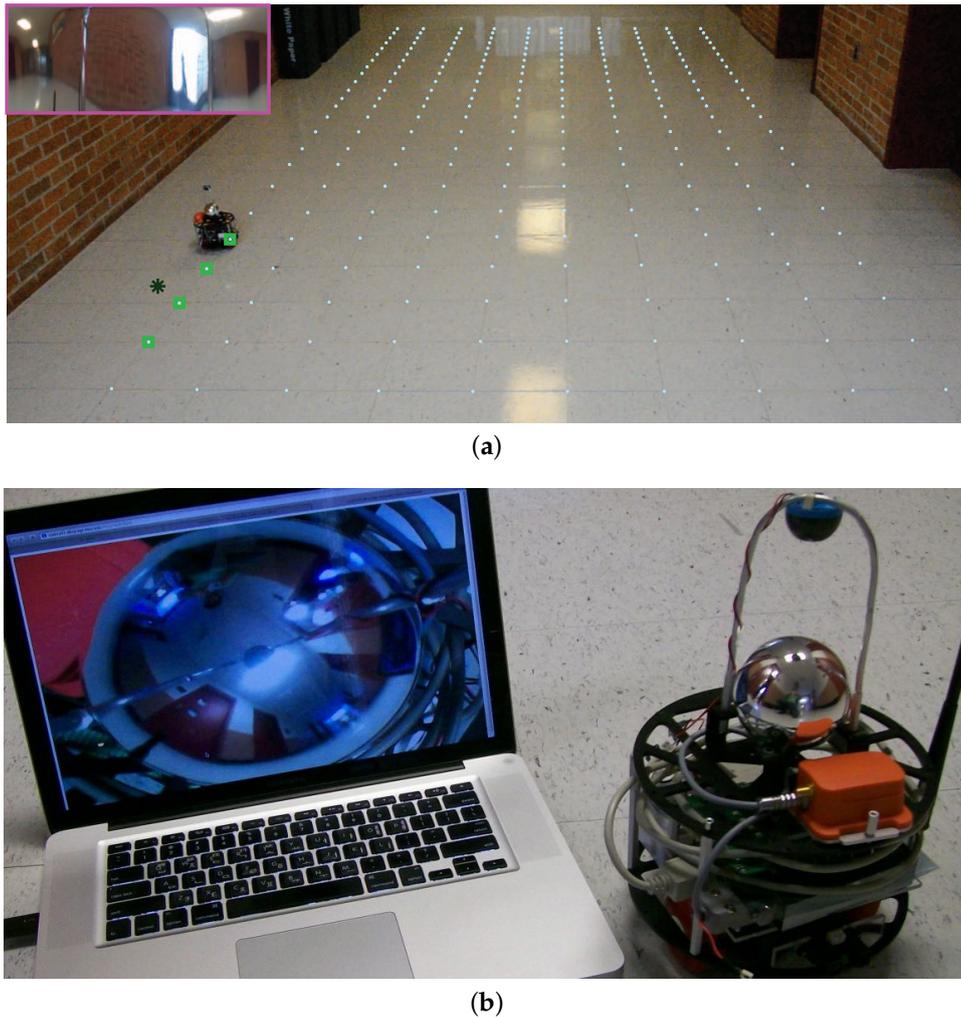
In this section, we discuss complexity aspects of the proposed algorithms. For a fixed number of the radial basis functions (i.e.,  $p$ ) and a fixed number of the special sites (i.e.,  $n$ ), the computational complexity of Algorithm 1 is dominated by Equation (24). The complexity of Algorithm 1 in each time step is  $O(LN^2)$ , where  $L$  is the number of possible hyperparameter vectors and  $N$  is the number of agents. The complexity of Algorithm 2 in time  $t$  is  $O(\gamma(t - m + 1 : t))$  times the complexity of Algorithm 1 for  $m$  time steps. Hence, the complexity of Algorithm 2 in time  $t$  is  $O(\gamma(t - m + 1 : t)LN^2M)$ . The numbers of special sites and radial basis functions affect the complexity of Algorithm 1 as well. The complexity of the three cases is  $O(n^3)$  with respect to the number of special sites due to the matrix inversion. Thus, if we use finer grids, we increase the number of special sites, i.e.,  $n$ , and the complexity increase cubical. For a fixed set of  $L$ ,  $n$  and  $N$ , the complexity of Algorithm 1 with respect to  $p$  is  $O(p^3)$ .

## 6. Experimental Results

Similarly to [38], a vision-based robotic sensor was built to validate the proof of concept. The wireless mobile robot is equipped with two motorized wheels, a micro-controller, and a 360-degree omnidirectional camera, a motion sensor, a wireless receiver, and a transmitter. The omnidirectional camera is homemade from a cheap Wi-Fi remote CCD camera (Ai-Ball<sup>®</sup>, Trek 2000 International, Singapore) and a globe mirror. The vision images of the 360-degree environment around the robot produced by the omnidirectional camera are streamed via 802.11 b/g Wi-Fi interface to a laptop for image processing (see Figure 2).

In this section, we apply the proposed prediction algorithms to real experimental data. Figure 2 shows our experimental setup in which a redness intensity field is sampled by the captured images from the CCD camera on top of the mobile robot. The CCD camera captures 360-degree images. The redness intensity is computed by simply averaging the red component of the RGB picture. Noisy visual measurements are sampled at random sampling positions by our robot. The position of the robot has been measured by an image processing software which is built by the authors and the true sampling positions are obtained manually by inspection.

The experimental objective is to predict the redness intensity field over a spatial space or a spatio-temporal space using the scalar field model proposed in Section 4. Note that each image has a lot of information. However, we only used the redness out of each image as a scalar value of interest. In the future, we plan to extend this experiment for multivariate random fields for multiple features from each image. Here we are considering two different scenarios. First, we consider a static field and then a time-varying field with a moving person in the surveillance region.

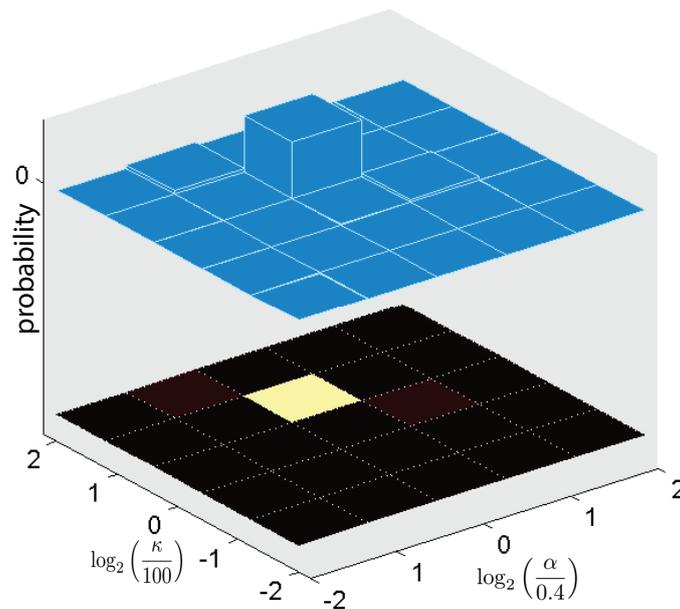


**Figure 2.** The experimental setup. (a) The measured position of the robot and four possible sampling positions are shown in dark green star (\*) and light green squares (□), respectively. The spatial sites are marked with aqua blue dots on the ground. The panoramic view of the robot is pictured on the upper left-hand side of the figure; (b) a vision-based robot is built with a 360-degree omnidirectional camera.

### 6.1. Spatial Field in the Static Experiment

In this study, the spatial sites in  $\mathcal{S}$  are considered to be  $10 \times 26$  grid points, i.e.,  $n = 260$ . The grid points are shown in Figure 2 with aqua-blue dots. The mean function  $\mu_t$  consists of only one radial basis function that keeps moving average of the field. Note that this basis function can model the changes in the brightness of the images caused by the slow changes of the environment lights. The center of the radial basis function is  $(5, 13)$ , and its bandwidth  $\sigma_1$  is  $\infty$ . The prior distribution of the hyperparameter vector  $\theta$  is chosen to be discrete with a support  $\Theta = \{25, 50, 100, 200, 400\} \times \{0.1, 0.2, 0.4, 0.8, 1.6\}$  and the associated uniform probabilities. The measurement noise variance  $\sigma_\epsilon = 0.01$  is estimated.

To demonstrate the usefulness of our model in Equations (12), (14), and (15), and our prediction algorithm, we simulate eighty mobile robots with a single mobile robot that measures spatially distributed eighty samples, i.e.,  $n = 80$ , where each of nine sampling positions is uncertain with four possibilities. Thus, there is  $4^9$  possible combinations for this set of sampling positions. After two sets of observations, the resulting posterior probabilities of the hyperparameters for Case 1 are shown in Figure 3. Figure 3 shows that the estimated hyperparameters converge to  $\kappa = 100$  and  $\alpha = 0.8$ , which is equivalent to  $\ell = 1.58$  and  $\sigma_f = 0.0315$ .



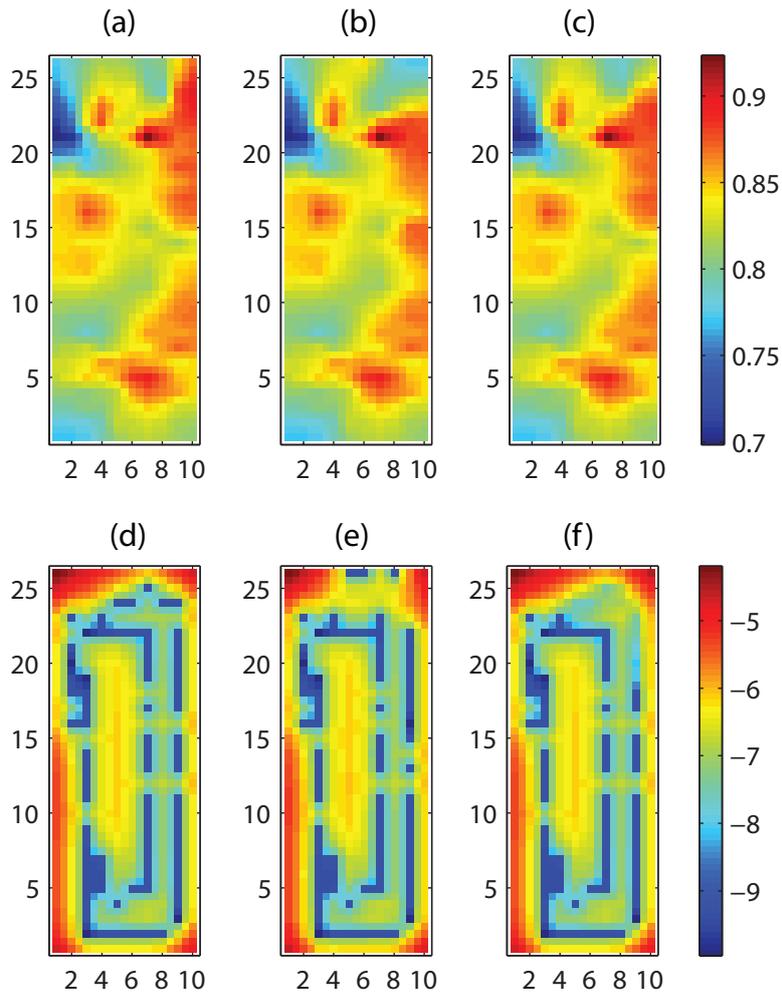
**Figure 3.** The posterior probability of the hyperparameter vector at  $t = 2$ .

We consider three cases: Case 1 using Algorithm 1 with exact sampling positions, Case 2 using Algorithm 1 naively to measured sample positions including noisy locations, and Case 3 using by applying Algorithm 2 with  $m = 1$ . The prediction and prediction error variance are computed for Cases 1, 2, and 3. The prediction and the prediction error variance using true sampling positions (Case 1), are shown in Figure 4a,d, respectively. Red and blue colors represent the highest and the lowest values, respectively. Figure 4b,e show the resulting fields, by applying Algorithm 1 naively to measured sample positions including noisy locations (Case 2).

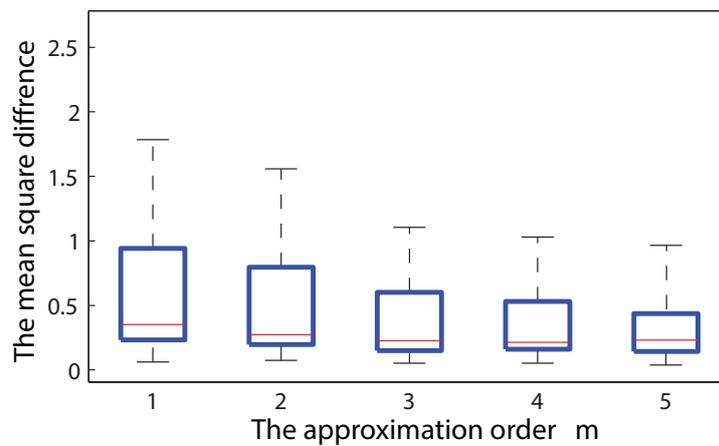
Figure 4c,f show the results by applying Algorithm 2 with  $m = 1$  (Case 3). Blue trails shown in Figure 4d–f represent the low predicted error variances due to sampling. The results confirm that the quality of the prediction in Case 3 are not very compromised as compared to Case 1 and demonstrate the capability of our proposed algorithm to deal with uncertain sampling positions.

Figure 5 shows the controllable trade-off between approximation error and complexity for another simulated example. The complexity of Algorithm 2 increases exponentially with respect to  $m$ . The predicted field is compared between Case 1, which is the best prediction quality expected, and Case 3. Clearly, by increasing  $m$ , the mean square difference between Case 1 and Case 3 decreases. However, this mild improvement results in the exponentially increasing computational load, which guides us to use  $m = 1$  for practical cases.

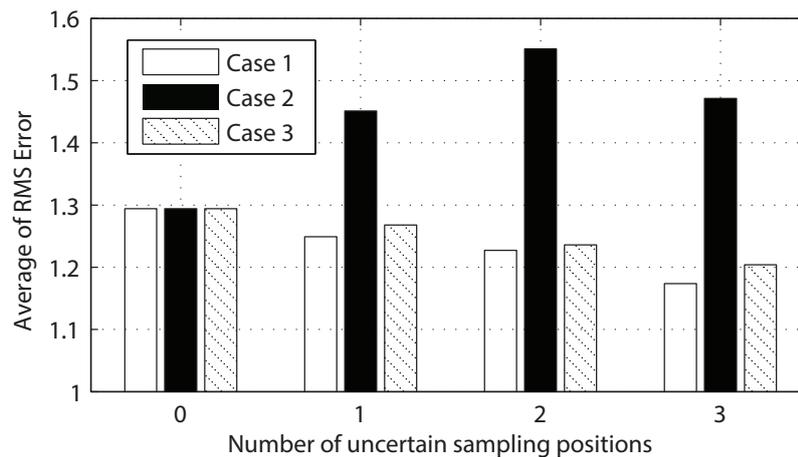
Figure 6 shows the effect of an increasing number of observations with uncertain sampling positions on the three cases from the same simulated example. Here we assume that we have seven observations with known true sampling positions plus a few observations with uncertain sampling positions. Clearly, the root mean square (RMS) error decreases in Case 1 and Case 3 by adding new observations with true and uncertain sampling positions, respectively. On the other hand, as shown in Figure 6, adding new observations with noisy sampling positions could increase RMS error for Case 2. Figure 6 also shows the efficacy of our proposed algorithm, which can be compared to the previous work [23]. The previous work in [23] was limited to a static random field and achieved 3.63 RMS error with a bandwidth of 4.47. Note that our proposed method (Case 3) considers the temporal random field and achieved averaged RMS error less than 1.2 with a bandwidth of 1.58. For a fair comparison against [23], we normalize the input space of the GMRF by the bandwidth and compare resulting values in terms of RMS over bandwidth. Our current work achieved 0.759 RMS/bandwidth outperforming the previous work in [23] with 0.812 RMS/bandwidth.



**Figure 4.** The prediction results of Cases 1, 2, and 3 at time  $t = 2$  are shown in the first, second, and third columns, respectively. The first and second rows correspond to the predictions and the natural logarithm of the prediction error variance.



**Figure 5.** The mean square difference between Case 1 and Case 3 is shown for the different approximation orders  $m = 1, \dots, 5$ . On each box, the central mark is the median, the upper and lower edges of the box are the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points.



**Figure 6.** The number of observations with uncertain sampling position on the three different cases is shown on the horizontal axis. The field prediction root mean square (RMS) error for Case 1, Case 2 and Case 3 are shown with white, black, and hatched bars, respectively.

## 6.2. Spatio-Temporal Field in the Dynamic Experiment

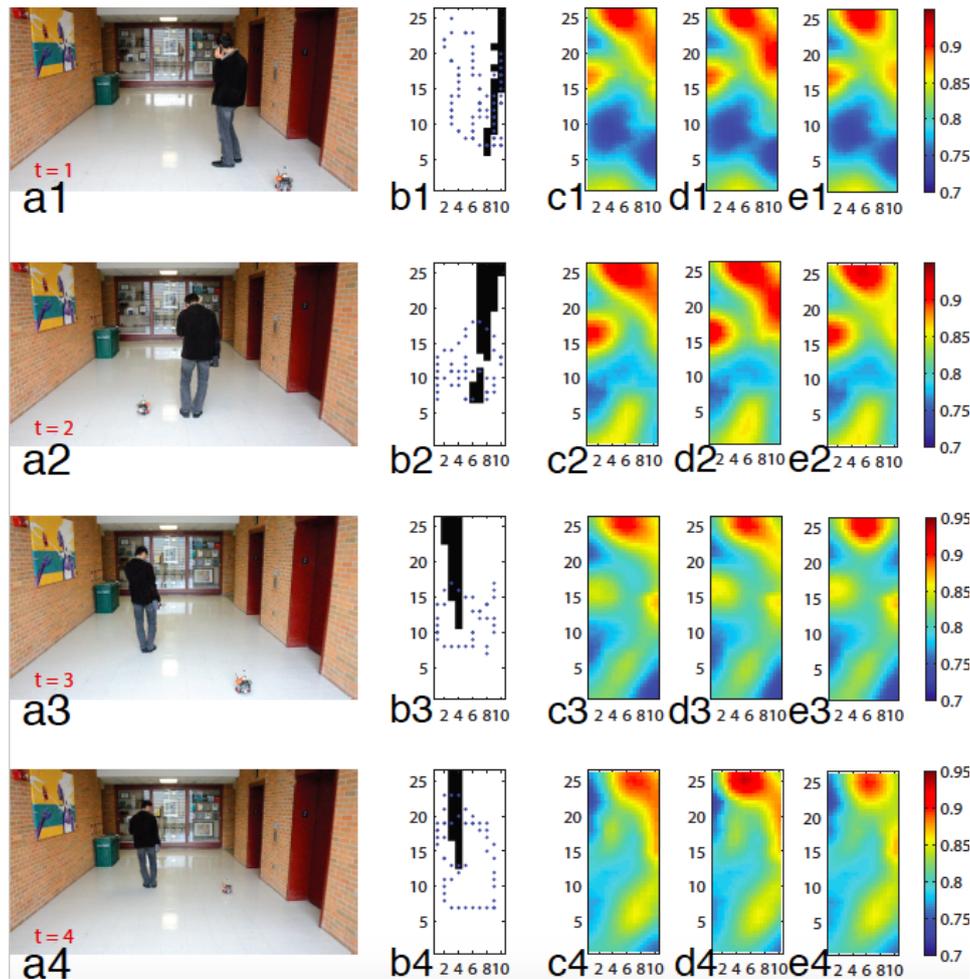
In this scenario, the spatial sites are the same as in the previous experiment. The mean function  $\mu_t$  consists of twenty nine radial basis functions. The centers of radial basis functions are  $\{(5, 13)\} \cup \{1, 4, 7, 10\} \times \{1, 5, 9, 13, 17, 21, 25\}$ . The time evolution of  $\beta_t$  is modeled by Equation (14), where the state matrix  $A_t$  and the input matrix  $B_t$  are given by  $I$  and  $0.05I$ , respectively. 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 that is equal to  $\sigma_j = 4$ . A person moves to a spot and stays still while a robot is collecting observations, the process of which is then repeated in order to efficiently simulate multiple robots. The robot collects 57, 40, 31, and 45 observations corresponding to times  $t = 1, 2, 3$ , and 4, respectively. The first column (a1–4) in Figure 7 shows the positions of moving robot and person in the domain. As mentioned, the position of the robot has been measured with a fixed camera and an image processing software. Sometimes, the robot is not visible by the positioning system since the moving person blocks the robot. The black areas in the second column (b1–4) of Figure 7 show the blind spots of the positioning system. Note that other positioning systems like GPS also have blind spots such as GPS denied areas. Therefore, when the robot moves to a blind spot the position cannot be determined precisely. In this case, we assign different probabilities on multiple sampling positions.

True sampling positions in each time step are shown with blue dots in the second column (b1–4) of Figure 7 while just blue dots in the white area are measured through the positioning system and the positions in the blind spots are recorded manually for a comparison purpose.

The prediction results of Case 1 and Case 3 are compared with a trivial method of prediction defined as follows.

Case 4: The fifth column (e1–4) of Figure 7 shows the resulting fields, by applying Algorithm 1 on just observed sampling positions. Here, all the observations whose sampling positions are uncertain are discarded. In particular, 12, 3, 2, and 4 observations have been discarded for time  $t = 1, 2, 3$  and 4, respectively.

The predicted field using true sampling positions (Case 1) and uncertain sampling positions (Case 3) are shown in the third (c1–4) and fourth (d1–4) columns of Figure 7, respectively. Since in the fully automated experiment true sampling positions in the blind spot are not available, we used the proposed algorithm in this paper to deal with uncertain sampling positions in the blind spot areas. The predicted field simply by discarding uncertain sampling positions is shown in the fifth column (e1–4) of Figure 7. The results obtained for Case 3 with  $m = 1$ , is not compromised considering the result for Case 1. Thus, the experimental result demonstrates the effectiveness of the proposed algorithm.



**Figure 7.** The time varying field is shown for four time iterations. The first column (a1–4) shows the moving object in the field. The second column (b1–4) shows the sampling positions with blue dots and the positioning blind spot with black areas. The third (c1–4), fourth (d1–4) and fifth (e1–4) columns show predicted field for Case 1, Case 3 and Case 4, respectively. The rows are correspond to the time intervals  $t = 1, 2, 3$  and 4.

## 7. Conclusions

We have tackled a problem of predicting a spatio-temporal field using successive noisy scalar measurements obtained by mobile robotic sensors, some of which have uncertain localization. We developed the spatio-temporal field of interest using a GMRF and designed sequential prediction algorithms for computing the exact and approximated predictive inference from a Bayesian point of view. The most important contribution is that the computation times for Algorithms 1 and 2 with a finite  $m$  at each time step do not grow as the number of measurements increases. Two different static and time-varying experimental results along with a comparison study using simulation results provide a solid proof of concept on the proposed scheme. The proposed algorithm will be useful for robotics applications such as environmental monitoring by autonomous aquatic robots and drones. Future work is to apply our algorithms to spatio-temporal sensory information fusion for autonomous driving. In particular, we plan to predict a spatio-temporal scalar field of a risk measure. The self-driving vehicle will be designed to perform path-planning taking into account such predicted risk measures over space and time for better safety.

**Author Contributions:** Conceptualization, M.J. and J.C.; Methodology, M.J., Y.X. and J.C.; Software, M.J. and Y.X.; Validation, M.J. and Y.X.; Formal Analysis, M.J. and J.J.; Investigation, M.J., J.C. and J.K.; Writing—Original Draft Preparation, M.J. and J.C.; Writing—Review & Editing, M.J., J.J., J.C. and J.K.; Supervision, J.C.; Funding Acquisition, J.C.

**Funding:** This work of the first and third authors was supported in part by the National Science Foundation through CAREER Award CMMI-0846547. The work of other authors was funded by the Mid-career Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2018R1A2B6008063). These supports are gratefully acknowledged.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lynch, K.M.; Schwartz, I.B.; Yang, P.; Freeman, R.A. Decentralized environmental modeling by mobile sensor networks. *IEEE Trans. Robot.* **2008**, *24*, 710–724. [[CrossRef](#)]
2. Xu, Y.; Choi, J.; Dass, S.; Maiti, T. *Bayesian Prediction and Adaptive Sampling Algorithms for Mobile Sensor Networks: Online Environmental Field Reconstruction in Space and Time*; Springer: New York, NY, USA, 2016.
3. Krause, A.; Singh, A.; Guestrin, C. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *J. Mach. Learn. Res.* **2008**, *9*, 235–284.
4. Xu, Y.; Choi, J.; Oh, S. Mobile sensor network navigation using Gaussian processes with truncated observations. *IEEE Trans. Robot.* **2011**, *27*, 1118–1131. [[CrossRef](#)]
5. Todescato, M.; Carron, A.; Carli, R.; Pillonetto, G.; Schenato, L. Multi-robots Gaussian estimation and coverage control: From client–server to peer-to-peer architectures. *Automatica* **2017**, *80*, 284–294. [[CrossRef](#)]
6. Xu, Y.; Choi, J. Spatial prediction with mobile sensor networks using Gaussian processes with built-in Gaussian Markov random fields. *Automatica* **2012**, *48*, 1735–1740. [[CrossRef](#)]
7. Hartikainen, J.; Särkkä, S. Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In Proceedings of the 2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP), Kittila, Finland, 29 August–1 September 2010; pp. 379–384.
8. Särkkä, S.; Hartikainen, J. Infinite-dimensional Kalman filtering approach to spatio-temporal Gaussian process regression. In Proceedings of the International Conference on Artificial Intelligence and Statistics, Naha, Japan, 16–18 April 2012; pp. 993–1001.
9. Carron, A.; Todescato, M.; Carli, R.; Schenato, L.; Pillonetto, G. Machine learning meets Kalman filtering. In Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), Las Vegas, NV, USA, 12–14 December 2016; pp. 4594–4599.
10. Xu, Y.; Choi, J. Adaptive sampling for learning Gaussian processes using mobile sensor networks. *Sensors* **2011**, *11*, 3051–3066. [[CrossRef](#)] [[PubMed](#)]
11. Xu, Y.; Choi, J.; Dass, S.; Maiti, T. Sequential Bayesian prediction and adaptive sampling algorithms for mobile sensor networks. *IEEE Trans. Autom. Control* **2012**, *57*, 2078–2084.
12. Xu, Y.; Choi, J.; Dass, S.; Maiti, T. Efficient Bayesian spatial prediction with mobile sensor networks using Gaussian Markov random fields. *Automatica* **2013**, *49*, 3520–3530. [[CrossRef](#)]
13. Oh, S.; Schenato, L.; Chen, P.; Sastry, S. Tracking and coordination of multiple agents using sensor networks: System design, algorithms and experiments. *Proc. IEEE* **2007**, *95*, 234–254. [[CrossRef](#)]
14. Huang, S.; Wu, Z.; Misra, A. A Practical, Robust and Fast Method for Location Localization in Range-Based Systems. *Sensors* **2017**, *17*, 2869. [[CrossRef](#)] [[PubMed](#)]
15. Luo, J.; Yin, X.; Zheng, Y.; Wang, C. Secure Indoor Localization Based on Extracting Trusted Fingerprint. *Sensors* **2018**, *18*, 469. [[CrossRef](#)] [[PubMed](#)]
16. Brooks, A.; Makarenko, A.; Upcroft, B. Gaussian process models for indoor and outdoor sensor-centric robot localization. *IEEE Trans. Robot.* **2008**, *24*, 1341–1351. [[CrossRef](#)]
17. Kemppainen, A.; Haverinen, J.; Vallivaara, I.; Röning, J. Near-optimal SLAM exploration in Gaussian processes. In Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Salt Lake City, UT, USA, 5–7 September 2010; pp. 7–13.
18. O’Callaghan, S.; Ramos, F.; Durrant-Whyte, H. Contextual occupancy maps using Gaussian processes. In Proceedings of the IEEE International Conference on Robotics and Automation, Kobe, Japan, 12–17 May 2009; pp. 1054–1060.

19. McHutchon, A.; Rasmussen, C. Gaussian Process training with input noise. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Grandata, Spain, 12–15 December 2011; pp. 1–9.
20. Do, H.N.; Jadaliha, M.; Choi, J.; Lim, C.Y. Feature selection for position estimation using an omnidirectional camera. *Image Vis. Comput.* **2015**, *39*, 1–9. [[CrossRef](#)]
21. Jadaliha, M.; Xu, Y.; Choi, J.; Johnson, N.; Li, W. Gaussian process regression for sensor networks under localization uncertainty. *IEEE Trans. Signal Process.* **2013**, *61*, 223–237. [[CrossRef](#)]
22. Choi, S.; Jadaliha, M.; Choi, J.; Oh, S. Distributed Gaussian process regression under localization uncertainty. *J. Dyn. Syst. Meas. Control* **2015**, *137*, 031007. [[CrossRef](#)]
23. Do, H.N.; Jadaliha, M.; Temel, M.; Choi, J. Fully Bayesian Field SLAM Using Gaussian Markov Random Fields. *Asian J. Control* **2016**, *18*, 1175–1188. [[CrossRef](#)]
24. Rue, H.; Tjelmeland, H. Fitting Gaussian Markov random fields to Gaussian fields. *Scand. J. Stat.* **2002**, *29*, 31–49. [[CrossRef](#)]
25. Hartman, L.; Hössjer, O. Fast kriging of large data sets with Gaussian Markov random fields. *Comput. Stat. Data Anal.* **2008**, *52*, 2331–2349. [[CrossRef](#)]
26. Le Ny, J.; Pappas, G.J. On trajectory optimization for active sensing in Gaussian process models. In Proceedings of the 48th IEEE Conference on Decision and Control, Shanghai, China, 16–18 December 2009; pp. 6286–6292.
27. Jetto, L.; Longhi, S.; Venturini, G. Development and experimental validation of an adaptive extended Kalman filter for the localization of mobile robots. *IEEE Trans. Robot. Autom.* **1999**, *15*, 219–229. [[CrossRef](#)]
28. St-Pierre, M.; Gingras, D. Comparison between the unscented Kalman filter and the extended Kalman filter for the position estimation module of an integrated navigation information system. In Proceedings of the IEEE Intelligent Vehicles Symposium, Parma, Italy, 14–17 June 2004; pp. 831–835.
29. Jadaliha, M.; Xu, Y.; Choi, J. Gaussian process regression using Laplace approximations under localization uncertainty. In Proceedings of the American Control Conference, Montreal, QC, Canada, 27–29 June 2012; pp. 1394–1399.
30. Jadaliha, M.; Choi, J. Environmental monitoring using autonomous aquatic robots: Sampling algorithms and experiments. *IEEE Trans. Control Syst. Technol.* **2013**, *21*, 899–905. [[CrossRef](#)]
31. Rue, H.; Held, L. *Gaussian Markov Random Fields: Theory and Applications*; Chapman & Hall: London, UK, 2005.
32. Lindgren, F.; Rue, H.; Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B* **2011**, *73*, 423–498. [[CrossRef](#)]
33. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA; London, UK, 2006.
34. Rue, H.; Martino, S.; Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B* **2009**, *71*, 319–392. [[CrossRef](#)]
35. Cressie, N.; Wikle, C. Space–time Kalman filter. *Encycl. Environ.* **2002**, *4*, 2045–2049.
36. Ko, J.; Fox, D. GP-BayesFilters: Bayesian filtering using Gaussian process prediction and observation models. *Auton. Robots* **2009**, *27*, 75–90. [[CrossRef](#)]
37. Cortés, J. Distributed Kriged Kalman filter for spatial estimation. *IEEE Trans. Autom. Control* **2009**, *54*, 2816–2827. [[CrossRef](#)]
38. Plagemann, C.; Stachniss, C.; Hess, J.; Endres, F.; Franklin, N. A nonparametric learning approach to range sensing from omnidirectional vision. *Robot. Auton. Syst.* **2010**, *58*, 762–772. [[CrossRef](#)]

