

# Data-Driven Models with Expert Influence: A Hybrid Approach to Spatiotemporal Process Estimation

Jun Liu and Ryan K. Williams

**Abstract**—In this paper, our motivating application lies in precision agriculture where accurate modeling of forage is essential for informing rotational grazing strategies. Unfortunately, a major difficulty arises in modeling forage processes as they evolve on large scales according to complex ecological influences. As robots can collect data over large scales in a forage environment, they act as a promising resource for the forage modeling problem when combined with a data-driven Gaussian processes (GPs) technique. However, GPs are non-parametric in nature and may be blind to certain nuances of a process that a parameterized expert model may predict well. Indeed, for the forage modeling problem specifically, there exist several highly parameterized models from agricultural experts that exhibit powerful predictive capabilities. Expert models, however, often come with two shortcomings: (1) parameters may be difficult to determine in general; and (2) the model may not make complete spatiotemporal predictions. For example, a stochastic differential equation (SDE) that models the dynamics of the *average* output of an environment may be available from experts (a typical case). In such cases, we propose to take advantage of both data-driven (GPs) and expert (SDE) models, by fusing data collected by robots, which often yields spatial insight, with models from experienced professionals that often yield temporal insights. Specifically, we propose to leverage *Bayesian estimation* to combine these two methods, resulting in a posterior prediction that is a *hybrid* of data-driven and expert models. Finally, we provide simulations to demonstrate the effectiveness of the proposed method.

## I. INTRODUCTION

As a motivating example for this paper, consider the problem of deploying multi-robot teams for dynamic forage monitoring in a precision grazing application. Fig. 1 is an example of such an environment. To take advantage of the forage resource and avoid environmental degradation in precision grazing practices, we require predictive models of the dynamics of the forage environment [1]–[4]. A good starting point, as we have explored in previous work [5], [6], is to deploy multi-robot teams to collect data from the environment and then build a spatiotemporal model based on these data. Indeed, a significant amount of work has been proposed based on this basic monitoring concept in the robotics community. One of the pioneering works on this problem of process monitoring was proposed in [7]. The problem setup consists of one or more robots acting as mobile sensors whose measurements are used to estimate a spatial field. The underlying field is modeled and estimated using Gaussian process regression [8], [9],

This work was supported by the National Institute of Food and Agriculture under Grant 2018-67007-28380.

The authors are with the Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061 USA (e-mail: junliu@vt.edu; rywilli1@vt.edu).



Fig. 1. An example of a slowly evolving environment which needs to be modeled in precision grazing applications.

which in addition to estimating the value at every point, also quantifies the uncertainty in the estimate by using the submodularity of mutual information [10], [11]. Other environmental modeling techniques using information-based metrics can also be found in [12], [13]. Importantly, the Gaussian process regression method is a *non-parametric* method, which is data-driven. To address *spatiotemporal* modeling of an environment, the work in [5], [12], [14] utilized separable and non-separable covariance kernels and Gaussian process regression in different cases. The work in [15] exploits an annealing algorithm to predict the estimation error instead of information-based errors by using Gaussian processes. This non-parametric environmental modeling method was also exploited in building an information distribution map [16] for an unknown environment.

The above described studies, and many more, are based on a typical robotics perspective: how do I produce useful models with all of my robots' data? However, other *parametric* factors that affect the evolution of environmental processes may be overlooked from such a perspective. For instance, expert models for forage growth are highly parameterized and often yield powerful predictive capability when coupled with appropriate parameters [17], [18]. The reasons for this data-driven vs. expert model gap is multi-faceted. Roboticians often focus more on autonomy, data collection, and modeling processes with collected data, and often less on the expert knowledge in the application domain (such as forage estimation for precision grazing). As opposed to the works in the robotics community, agricultural professionals often put more effort into developing *parametric* models. These models are usually (stochastic) partial differential equations (PDEs) when

biosystems from different perspectives are considered [19]–[21]. Meanwhile, agricultural professionals lack powerful autonomy tools, as the robotics community possesses.

To take advantage of both sides of this picture, we argue that fusing data-driven approaches with expert models will result in a better environmental process model. There are some works, such as [22]–[24], that have demonstrated the mixture of Gaussian processes, and these works usually involve Bayesian inference [25]. The work in [26] utilized Gaussian processes to build an information distribution map and then used a PDE to model the diffusion of the information. However, the fusion of GPs and PDEs was not the primary focus. The work proposed in [27] takes advantage of functional Gaussian regression when a PDE is used to model the dynamics of a process while considering measurements. Still, this work does not focus on hybrid models that belong to the data-driven and expert model categories.

In this paper, we aim to combine a data-driven approach to spatiotemporal process estimation, represented as a GP fit with robot-collected data, with a parameterized expert model of the same process. We begin by demonstrating through *non-parametric* methods a manner in which spatiotemporal processes can be modeled using the Gaussian process. Next, we present a stochastic differential equation for the dynamic change of a quantity in an environment, allowing for a *parametric* modeling for the environment from an expert perspective. By analyzing these models of the same environment from different viewpoints, we argue and then demonstrate that Bayesian estimation is a reasonable path towards a hybrid model. Finally, Monte Carlo simulations are presented that indicate the effectiveness of the hybrid model for dynamic environmental modeling.

## II. PRELIMINARIES

As we will apply a non-parametric method (Gaussian processes) with a parametric approach (SDE) for different aspects of environmental modeling, we will briefly introduce these two modeling methods.

### A. Non-Parametric Modeling: Gaussian Processes

A Gaussian process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution [8]. Generally, a Gaussian process can be specified by a mean function  $\mu(\mathbf{x}) = \mathbb{E}[z(\mathbf{x})]$  and a covariance function  $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{E}[(z(\mathbf{x}_i) - \mu(\mathbf{x}_i))(z(\mathbf{x}_j) - \mu(\mathbf{x}_j))]$ , where  $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^p$  are inputs. This GP can be denoted by  $z(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), \kappa(\mathbf{x}_i, \mathbf{x}_j))$ . Before making an observation, we can use a Gaussian process to predict the output and the corresponding uncertainty for any input  $\mathbf{x}' \in \mathbb{R}^p$  given a dataset  $\mathcal{D} = \{(\mathbf{x}_i, z_i) \mid i = 1, \dots, n\}$ , where  $\mathbf{x}_i \in \mathbb{R}^p$  denotes an input and  $z_i \in \mathbb{R}$  is the corresponding output. The prediction for  $z \mid \mathbf{x}', \mathcal{D}$  is specified by mean and covariance as

$$\begin{aligned} \bar{z}(\mathbf{x}') &= \mu(\mathbf{x}') + K_{\mathbf{x}'X}(K_{XX} + \sigma_\epsilon^2 I)^{-1}(\mathbf{z} - \mu(X)), \\ \Sigma(\mathbf{x}') &= K_{\mathbf{x}'\mathbf{x}'} - K_{\mathbf{x}'X}(K_{XX} + \sigma_\epsilon^2 I)^{-1}K_{X\mathbf{x}'}, \end{aligned} \quad (1)$$

where  $\epsilon$  is a Gaussian white noise,  $X = [\mathbf{x}_1^\top, \dots, \mathbf{x}_p^\top]^\top \in \mathbb{R}^{n \times p}$  is the vector stacked with training inputs, and  $\mathbf{z} = [z_1, \dots, z_p]^\top \in \mathbb{R}^n$  is the vector stacked with training outputs.  $K_{\mathbf{x}'\mathbf{x}'}, K_{\mathbf{x}'X}, K_{X\mathbf{x}'}$  and  $K_{XX}$  are the covariance matrices constructed through covariance function  $\kappa(\cdot)$ . The key factors that influence the prediction performance are the mean and covariance function. A commonly used covariance function is the *squared exponential* (SE) function

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \sigma_k^2 \exp\left(-\frac{1}{2\ell_k^2}(\mathbf{x}_i - \mathbf{x}_j)^2\right),$$

where  $\sigma_k^2$  and  $\ell_k$  are the signal variance and length-scale, which are called *hyperparameters*. Gaussian processes are a *non-parametric* model as the structure of the model is not fixed but specified by these hyperparameters. As we collect more data, the prediction will be refined, and thus we refer to GPs as a *data-driven* method. To give an intuition of how a GP utilizes training data, Fig. 2 shows a 1D Gaussian process example. We also direct the reader to work [8] for more details.

### B. Parametric Modeling: Stochastic Differential Equations

A stochastic differential equation (SDE) is a differential equation that contains one or more stochastic processes, and the solution is also a stochastic process. An SDE can be viewed as an ordinary differential equation that is perturbed by noise. Generally, an SDE can be written as

$$\dot{\mathbf{m}}(t) = f(\mathbf{m}(t), t, \xi(t)),$$

where  $\mathbf{m}(t) \in \mathbb{R}^q$  denotes a state (of the modeled process),  $t \in \mathbb{R}$  is the time, and  $\xi(t) \in \mathbb{R}^q$  is a Gaussian white noise. This equation is usually used as a state transition equation in controls. We can write it in another form as

$$\frac{d\mathbf{m}(t)}{dt} = \mathbf{a}(\mathbf{m}(t), t) + \mathbf{b}(\mathbf{m}(t), t)\xi(t),$$

where  $\mathbf{a}, \mathbf{b} : \mathbb{R}^q \times \mathbb{R}_+ \mapsto \mathbb{R}^q$ . This is the general form of the Langevin equation [28]. To deal with the noise  $\xi(t)$ , we can multiply  $dt$  on both sides and rewrite it as

$$d\mathbf{m}(t) = \mathbf{a}(\mathbf{m}(t), t) dt + \mathbf{b}(\mathbf{m}(t), t) dW(t).$$

where  $dW(t) = \xi(t) dt$ . In this case,

$$W(t) = W(0) + \int_0^t \xi(\tau) d\tau,$$

with  $W(0)$  as the initial condition.  $W(t)$  is an  $n$ -dimensional Brownian motion or Wiener process, which is initially used for describing the random motion of particles.  $dW(t)$  is an increment of the Brownian motion  $W(t)$  and  $\xi(t)$  is the derivative of this Brownian motion  $W(t)$ . In Wiener processes, the distribution of  $W(t_2) - W(t_1)$  follows a Gaussian distribution, i.e.,  $(W(t_2) - W(t_1)) \sim \mathcal{N}(0, t_2 - t_1)$ , if  $0 \leq t_1 < t_2$ . Also,  $\mathbf{m}(t)$  is referred to as an *Itô process* [28]. If  $\mathbf{a}(\mathbf{m}(t), t)$  and  $\mathbf{b}(\mathbf{m}(t), t)$  satisfy the following Lipschitz continuity condition,

$$|\mathbf{a}(\mathbf{m}, t) - \mathbf{a}(\mathbf{y}, t)| + |\mathbf{b}(\mathbf{m}, t) - \mathbf{b}(\mathbf{y}, t)| \leq C|\mathbf{m} - \mathbf{y}|,$$

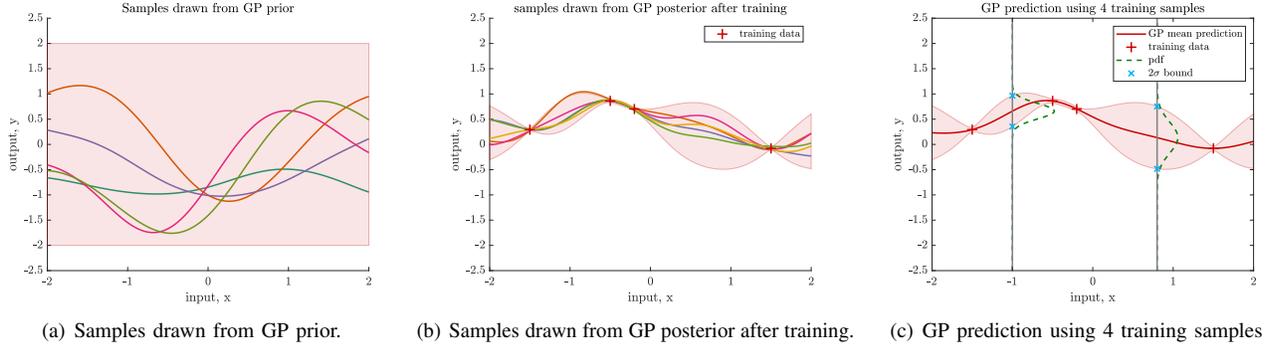


Fig. 2. 1D GP example of GP prior, posterior and prediction. (a). 5 samples are drawn randomly from prior. The prior mean is  $m(\mathbf{x}) = 0$ . The covariance kernel is SE kernel with  $(\ell, \sigma) = (1, 1)$ . The shaded area represents the  $2\sigma$  bound (corresponding to 95% confidence region). (b). 5 samples (curves) are drawn from posterior conditioning on 4 observations (training inputs). (c) GP prediction and the corresponding  $2\sigma$  bounds. Two dash lines represent the PDF's for  $x = -1$  and  $x = 0.8$ .

for a constant  $C$  and  $\mathbf{m}, \mathbf{y} \in \mathbb{R}^q$ , then there exists a unique solution for  $\mathbf{m}(t)$ . However, note that not all SDEs can be solved analytically. In general, the typical approach for solving a differential equation is to calculate the integral on both sides while considering initial conditions. This process can be expressed as

$$\mathbf{m}(t) = \mathbf{m}(0) + \int_{t_0}^t \mathbf{a}(\mathbf{m}(\tau), \tau) d\tau + \int_{t_0}^t \mathbf{b}(\mathbf{m}(\tau), \tau) dW(\tau),$$

where  $\mathbf{m}(0)$  is an initial condition. The basic requirement is that the function is integrable in Riemann's sense. However, since the Brownian motion  $W(t)$  is a random process with discontinuous derivatives, the second integral in Riemann's sense doesn't exist. To deal with this case, we resort to *Itô calculus* [28] that is derived by forming the Taylor expansion of  $f(\mathbf{m}(t), t, W(t))$  and ignoring higher-order terms.

In the next section, we will use an SDE to model the evolution of a quantity that describes a spatiotemporal process of interest, interpreted as an expert model, and then demonstrate how to utilize Itô calculus to solve it. We direct the reader to [28] for more details about SDEs.

### III. BAYESIAN INFERENCE FOR A HYBRID MODEL

As in our motivating example, we are tasked with monitoring a large-scale forage process with multi-robot deployments. Let us consider a spatiotemporal environmental process that evolves over a 2D space. From the robotics practitioners' perspective, we can use a spatiotemporal Gaussian process to model the dynamics and refine it as we collect more data using robots. On the other hand, agricultural experts often use an SDE to model the dynamics of the average output of the environment. This SDE is commonly used to predict the average output for the next time (e.g., day/month/year), and is an essential indicator in precision grazing applications.

#### A. Environmental Modeling from Different Perspectives

We model the forage environment that we must monitor as a spatiotemporal Gaussian process

$$z(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), \kappa(\mathbf{x}_i, \mathbf{x}_j)). \quad (2)$$

that evolves over a 2D space.  $\mathbf{x}_i = [x_i, y_i, t]^\top \in \mathbb{R}^3$  is the state that contains a 2D location  $(x_i, y_i)$  and a time  $t \in \mathbb{R}_+$ . As Gaussian processes generally operate in discrete space, we discretize the 2D environment into  $P \times Q$  cells. We then use  $\mathcal{P} = \{1, \dots, P \times Q\}$  to denote the set that contains the indexes of these  $|\mathcal{P}| = P \times Q$  locations. In this case,  $i \in \mathcal{P}$ . As the monitoring time horizon is not infinite, we use  $\mathcal{T} = \{1, \dots, T\}$  to represent the indexes of different times where  $T$  is the time horizon, i.e.,  $t \in \mathcal{T}$ . Given  $\mathcal{P}$  and  $\mathcal{T}$ , the *ground set* of this GP can be represented as  $\mathcal{V} = \mathcal{P} \times \mathcal{T}$ , which is the Cartesian product of  $\mathcal{P}$  and  $\mathcal{T}$ . The ground set contains all the available location and time combinations during monitoring and process prediction. Therefore, any pair of location  $(x, y)$  at any monitoring time  $t$  can be uniquely defined.

In this GP, we also denote by  $\mathcal{D} = \{(\mathbf{x}_i, z_i) \mid i = 1, \dots, n\}$  the training dataset used to calculate the *joint probability* of different known inputs through (1).  $\mathbf{x}_i \in \mathbb{R}^3$  is an input, and  $z_i \in \mathbb{R}$  is the corresponding output. Through our motivating example, we consider the output  $z_i$  to be the height of forage in location  $(x_i, y_i)$ . Stacking these inputs and outputs into different vectors, we get  $X = [\mathbf{x}_1^\top, \dots, \mathbf{x}_n^\top]^\top \in \mathbb{R}^{n \times 3}$  and  $\mathbf{z} = [z_1, \dots, z_n]^\top \in \mathbb{R}^n$ . Before calculating the joint probability over all known inputs, we need to define the mean function and covariance function properly. Given our problem setting, we know that agricultural experts model the dynamics of the *mean output* more rigorously through their expertise. Thus, we can rely on this expert model (SDE) and use a zero-mean function in our GP modeling. That is

$$\mu(\mathbf{x}) = h(t). \quad (3)$$

Here, we model the mean as a general function of  $t$ . We will demonstrate how to get  $h(t)$  in the next section when fusing our two models. At the same time, as there is no covariance information from agricultural professionals, we use a separable covariance function [5] to model the spatial and temporal correlation of this environment. That is,

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = \kappa_p(\mathbf{p}_i, \mathbf{p}_j) \kappa_t(t_i, t_j) + \epsilon_c^2, \quad (4)$$

where  $\mathbf{p}_i = [x_i, y_i]$  is the location of  $i$ th point from the training input  $\mathbf{x}_i$ . Similarly,  $\mathbf{p}_j$  is the location of  $j$ th point from the training input  $\mathbf{x}_j$ . Also,  $\kappa_p$  is the spatial kernel which characterizes the spatial relationship between different locations. A commonly used spatial kernel is the SE kernel that we also adopt in this paper.  $k_t$  is the temporal kernel that models the relationship between different times. Many temporal kernels are suitable for capturing temporal properties [29]. For example, SE kernel, Spectral Mixture Kernel, and so on. In this work, we will use the SE kernel to model the temporal property as our main focus is how to fuse models from different fields.

Now, let us look at the problem from another point of view. Agricultural professionals tend to use an SDE to model the dynamics of average output rather than the output for every location since they can perform regression using historical data to yield an empirical SDE. A general SDE that is often used for modeling this dynamic is as follows

$$dm(t) = a(t) dt + b(t) dW(t), \quad m(0) = m_0 \quad (5)$$

where  $m(t) \in \mathbb{R}$  is the average output in this discretized 2D space,  $W(t)$  is a Wiener process,  $a, b \in \mathbb{R}$  are constants,  $m(0) \in \mathbb{R}$  is the initial condition. In general, this model is a linear SDE. The solution to this equation is

$$m(t) = m_0 + \int_0^t a(\tau) d\tau + \int_0^t b(\tau) dW(\tau).$$

The first integral is a general integral, and the second integral is an Itô integral as  $W(t)$  is a Wiener process. From the structure of this equation, we can see that  $m(t)$  is a Gaussian distribution. This observation results from the fact that the second integral is a weighted Gaussian distribution, and the first integral is a deviation. Therefore, we can write  $m(t)$  as  $m(t) \sim \mathcal{N}(\mu_1(t), \sigma_1^2(t))$ , where

$$\mu_1(t) = m_0 + \int_0^t a(\tau) d\tau,$$

and

$$\sigma_1^2(t) = \mathbb{E} \left[ \int_0^t b(\tau) dW(\tau) \int_0^t b(\tau) dW(\tau) \right].$$

This integral is an Itô integral and thus we need to resort to *Itô calculus* [28]. That is,  $(dW(t))^2 = dt$ . After substitution for  $dW(t)$ , we have

$$\begin{aligned} \sigma_1^2 &= \mathbb{E} \left[ \left( \int_0^t b(\tau) dW(\tau) \right)^2 \right] \\ &= \int_0^t b(\tau)^2 d\tau. \end{aligned}$$

Therefore, when we define  $a(t) = \alpha$  and  $b(t) = \beta$ , we obtain the distribution of  $m(t)$  as  $m(t) \sim \mathcal{N}(\mu_1(t), \sigma_1^2(t))$  and

$$\mu_1(t) = \alpha t + m_0, \quad \sigma_1^2(t) = \beta^2 t. \quad (6)$$

Thus, we observe that the dynamics of the mean of this 2D field follows a Gaussian distribution, where the mean is an affine function of time  $t$  with slope  $\alpha$  and intercept  $m_0$  and

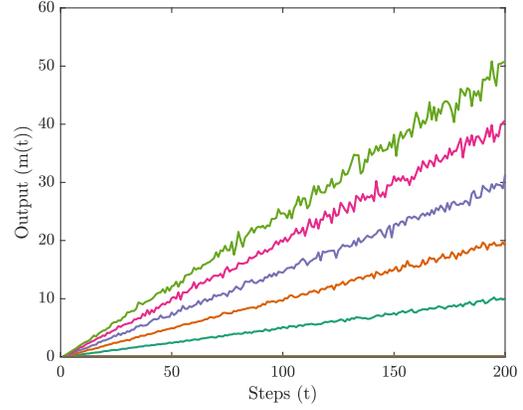


Fig. 3. Realization of 5 samples of the dynamics of  $m(t)$  with different parameters.

is disturbed by noise  $\beta^2 t$ .

*Remark 3.1.* Here, we assume  $\alpha(t)$  and  $\beta(t)$  are constants, which result in a linear SDE. The mean of this solution is an affine function of time, which perfectly reflects the fact that the growth of forage is also a function of time in our motivating example.

To give an intuition of the dynamics of  $m(t)$ , Fig. 3 shows an example of 5 different paths/samples of  $m(t)$ . In this example, the parameters are set as follows:  $m_0 = 0$ ,  $\alpha$ 's are integers from 1 to 5, and  $\beta$ 's are from 0.15 to 0.4 with equal interval. Here, we use this model to characterize the temporal property for the mean. This model is a *parametric* model, which doesn't need to be refined when we collect more data.

## B. Bayesian Estimation

After deriving models from different perspectives, we are ready to fuse these two models to produce the posterior. On the one hand, the result (6) from the agricultural professional model (5) gives the distribution of the average/mean of the monitored process at different times  $t$ . On the other hand, the model (2) from the robotics practitioner provides the distribution of the output in the same process. Since the process evolution of every location and the mean should follow the same pattern, we have the following assumption.

*Remark 3.2.* We assume that the professional model (5) also represents the dynamics of the mean for every location in the same environment.

Therefore, from the Bayesian point of view, it is reasonable to use the information from agricultural professionals as a prior distribution for every location and the information from the robotics community as likelihood information. According to the above procedures, we can compute the posterior distribution for every location in this field.

First, let us deal with the GP model from the robotics community. From (2), we know that the output from the GP follows a normal distribution  $\mathcal{N}(\mu(\mathbf{x}), \kappa(\mathbf{x}_i, \mathbf{x}_j))$ . If we have  $n$  predicted outputs  $z_i, \forall i = 1, \dots, n$  from the GP prediction,

we then can write the distribution of  $z_i$  as

$$z_i \sim \mathcal{N}(\mu, \sigma^2),$$

Now we can write the likelihood function, which is the joint probability distribution of  $z_i$ 's, as<sup>1</sup>

$$\begin{aligned} & \mathcal{L}(\mu, \sigma^2 | z_1, \dots, z_n) \\ &= \mathbb{P}(z_1, \dots, z_n | \mu, \sigma^2) \\ &= \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (z_i - \mu)^2\right) \\ &= \left(\prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}}\right) \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (z_i - \mu)^2\right). \end{aligned} \quad (7)$$

From the assumption (Remark 3.2), we know that the professional model (5) also represents the dynamics of the mean for every location in the same environment. Therefore, we have  $\mu \sim \mathcal{N}(\mu_1(t), \sigma_1^2(t))$ . To simplify the notation, we write it as

$$\mu \sim \mathcal{N}(\mu_1, \sigma_1^2),$$

where  $\mu_1 = \alpha t + m_0$  and  $\sigma_1^2 = \beta^2 t$  according to (6). This is a prior distribution of  $\mu$ .

Given the prior information of the mean  $\mu$  and the likelihood function  $\mathbb{P}(z_1, \dots, z_n | \mu, \sigma^2)$ , we are ready to get the posterior distribution of  $\mu$  through Bayesian estimation

$$\begin{aligned} & \mathbb{P}(\mu | z_1, \dots, z_n) \\ &= \frac{\mathbb{P}(z_1, \dots, z_n | \mu, \sigma^2) \mathbb{P}(\mu | \mu_1, \sigma_1^2)}{\int \mathbb{P}(z_1, \dots, z_n | \mu, \sigma^2) \mathbb{P}(\mu | \mu_1, \sigma_1^2) d\mu}. \end{aligned}$$

Since the denominator is a constant, we only need to focus on the numerator. Thus, the above equation becomes,

$$\begin{aligned} & \mathbb{P}(\mu | z_1, \dots, z_n) \\ & \propto \mathbb{P}(z_1, \dots, z_n | \mu, \sigma^2) \mathbb{P}(\mu | \mu_1, \sigma_1^2) \\ & \propto \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (z_i - \mu)^2\right) \exp\left(-\frac{1}{2\sigma_1^2} (\mu - \mu_1)^2\right). \end{aligned}$$

To simplify the calculation, let us define  $\rho = n^{-1} \sum_{i=1}^n z_i$ , which is the average output of all  $z_i$ 's. We then have

$$\begin{aligned} & \mathbb{P}(\mu | z_1, \dots, z_n) \\ & \propto \exp\left(-\frac{1}{2\sigma^2} n(\rho - \mu)^2\right) \exp\left(-\frac{1}{2\sigma_1^2} (\mu - \mu_1)^2\right) \\ & \propto \exp\left(-\frac{1}{2}(\mu^2 A - 2\mu B + C)\right), \end{aligned} \quad (8)$$

where

$$A = \frac{n}{\sigma^2} + \frac{1}{\sigma_1^2}, \quad B = \frac{n\rho}{\sigma^2} + \frac{\mu_1}{\sigma_1^2}, \quad C = \frac{n\rho^2}{\sigma^2} + \frac{\mu_1^2}{\sigma_1^2}.$$

From this result, we observe that the posterior of  $\mu$  follows

<sup>1</sup>The expression of the PDF function  $\mathbb{P}(\cdot)$  and the likelihood function  $\mathcal{L}(\cdot)$  are the same for the same distribution. In PDF functions  $\mathbb{P}(\cdot)$ , the model parameters are known, and data is generated through the model. In likelihood functions  $\mathcal{L}(\cdot)$ , the model parameters are estimated by the known data.

a normal distribution  $\mathcal{N}(\mu', \sigma'^2)$ . Now, we can replace  $\sigma^2$ ,  $\mu_1$ , and  $\sigma_1^2$  with the result from (4) and (6), and get the posterior mean  $\mu'$  and the posterior covariance  $\sigma'^2$  of  $\mu$  as

$$\begin{aligned} & \mu' \\ &= \left(\frac{\sum_{i=1}^n z_i}{\sigma^2} + \frac{\mu_1}{\sigma_1^2}\right) \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_1^2}\right)^{-1}, \\ &= \left(\frac{\sum_{i=1}^n z_i}{\sigma^2} + \frac{\mu_1}{\beta^2 t}\right) \left(\frac{n}{\sigma^2} + \frac{1}{\beta^2 t}\right)^{-1}, \end{aligned} \quad (9)$$

and

$$\sigma'^2 = \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_1^2}\right)^{-1} = \left(\frac{n}{\sigma^2} + \frac{1}{\beta^2 t}\right)^{-1}. \quad (10)$$

Therefore, this posterior distribution  $\mathbb{P}(\mu | z_1, \dots, z_n)$  combines the information from the model of the robotics community and the agricultural experts' model.

#### IV. SIMULATIONS

In this section, we demonstrate the performance of our proposed hybrid model. Specifically, we will first compare the performance of using a single non-parametric model (GP) with a ground truth spatiotemporal process. Then, we will examine the hybrid model (GP + SDE) with ground truth. Finally, we will check and compare these two performances.

The basic settings are as follows: we have a  $200 \times 200$  space that we need to monitor. We then downsample this space into  $20 \times 20$  to speed up the simulation. First, we need to build the ground truth for comparisons. Regarding this, we first use a GP to generate the ground truth. The mean and the covariance parameters are set by using (9) and (10). Note that we also need to generate  $n$  initial data to get  $\mu'$ . In the simulation, we set the length-scale  $\ell_p = 30$  and signal variance  $\sigma_p^2 = 1$  for the temporal kernel  $\kappa_p$ . Also, for the spatial kernel, we set the length-scale and the signal variance as  $\ell_t = 1$  and  $\sigma_t = 1$ . Then, using this setting, we generate 10 different data for each of the first 5 times. Thus, we have  $n = 50$  data points for making the  $\mu'$  in (9). This is the model from the robotics community. Next, we need to set the SDE model (6) from the agricultural expert's perspective. The parameters for SDE model are:  $\alpha = 1$ ,  $\beta = 0.15$ . Then, using these parameters, we generate the ground truth for the comparisons.

We first demonstrate the procedure to make predictions only through the GP model. First, at each time from  $t = 1$  to  $t = 5$ , we collect 20 samples randomly. Note that these data are not the same as the data we mentioned above. The collected data in this stage is used to train GP as a GP model needs training data. The data we mentioned before is used to build the ground truth. We then use the same parameters for the length-scale and signal variances as we used in generating the ground truth. Though the initial settings are the same, these parameters will be trained and changed based on the data. Then, according to these 100 data, we train the GP model and then predict for every location from time  $t = 6$  to  $t = 20$ . Next, we demonstrate how to make predictions using

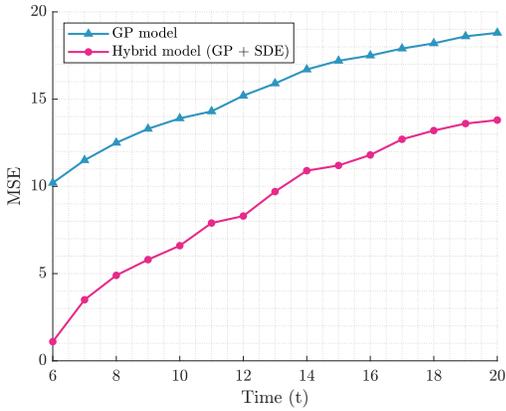


Fig. 4. The comparison of mean squared error (MSE) of both methods.

the hybrid model. First, we collect 20 samples for each time from  $t = 1$  to  $t = 5$ . As we know that the mean of these data follows an SDE model (5), we can use these data to fit this model and get the parameters  $\alpha$ ,  $\beta^2$ , and  $m_0$  through regression. These settings are for the model that comes from agricultural professionals. Next, we set the parameters for the GP model. Here, we use the same parameters as we used in the simple GP method. Then, we make predictions for the environment from time  $t = 6$  to  $t = 20$ .

We then calculate the mean squared error (MSE) of both methods for each time from  $t = 6$  to  $t = 20$ . The result is shown in Fig. 4. We observe that the hybrid model performs better than the data-driven only method. The ground truth is generated through the hybrid model with known parameters. Though we don't know the parameters when we use the hybrid method, we can make an initial estimation using the collected data. Also, from the result, we observe that the increments of MSE of both methods decrease as we perform more testing steps. Though we have a temporal kernel to model the temporal relationships, the predicted variance will increase because there are no observations in these steps. Also, Fig. 5 shows the result of ground truth ((a)- (c)), the result of the proposed hybrid model method ((d) - (f)), and the result using GP prediction only ((g)-(i)) at time  $t = 6$ ,  $t = 10$ , and  $t = 20$ .

## V. CONCLUSIONS AND FUTURE WORK

In this paper, we demonstrated how to fuse a non-parametric model with a parametric model for a dynamic environmental modeling problem. Specifically, we use a GP model as a non-parametric method and an SDE as a parametric method. We then fuse these two models using the SDE model as a prior. Finally, we demonstrated how to compute the posterior of the hybrid model. As we noted that not all SDEs are solvable, we need to resort to Monte Carlo simulations. We suggest that handling non-solvable SDEs is an avenue for future research.

## REFERENCES

- [1] J. P. Fentanes, I. Gould, T. Duckett, S. Pearson, and G. Cielniak, "3-d soil compaction mapping through kriging-based exploration with a mobile robot," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3066–3072, 2018.
- [2] J. Liu and R. K. Williams, "Optimal intermittent deployment and sensor selection for environmental sensing with multi-robot teams," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2018, pp. 1078–1083.
- [3] D. L. Zimmerman, "Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 17, no. 6, pp. 635–652, 2006.
- [4] Z. Zhu and M. L. Stein, "Spatial sampling design for prediction with estimated parameters," *Journal of agricultural, biological, and environmental statistics*, vol. 11, no. 1, p. 24, 2006.
- [5] J. Liu and R. K. Williams, "Monitoring over the long term: Intermittent deployment and sensing strategies for multi-robot teams," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2020, pp. 7733–7739.
- [6] J. Liu and R. K. Williams, "Coupled temporal and spatial environment monitoring for multi-agent teams in precision farming," in *Proceedings of IEEE Conference on Control Technology and Applications*, 2020, pp. 273–278.
- [7] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *Journal of Machine Learning Research*, vol. 9, no. Feb, pp. 235–284, 2008.
- [8] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006.
- [9] S. T. O'Callaghan and F. T. Ramos, "Gaussian process occupancy maps," *The International Journal of Robotics Research*, vol. 31, no. 1, pp. 42–62, 2012.
- [10] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Mathematical programming*, vol. 14, no. 1, pp. 265–294, 1978.
- [11] J. Liu and R. K. Williams, "Submodular optimization for coupled task allocation and intermittent deployment problems," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3169–3176, 2019.
- [12] J. Choi and D. Milutinović, "Tips on stochastic optimal feedback control and bayesian spatiotemporal models: Applications to robotics," *J. Dyn. Syst. Meas. Contr.*, vol. 137, no. 3, 2015.
- [13] G. A. Hollinger, B. Englot, F. Hover, U. Mitra, and G. S. Sukhatme, "Uncertainty-driven view planning for underwater inspection," in *2012 IEEE International Conference on Robotics and Automation*, 2012, pp. 4884–4891.
- [14] A. Singh, F. Ramos, H. D. Whyte, and W. J. Kaiser, "Modeling and decision making in spatio-temporal processes for environmental surveillance," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2010, pp. 5490–5497.
- [15] L. Van Nguyen, S. Kodagoda, R. Ranasinghe, and G. Dissanayake, "Simulated annealing based approach for near-optimal sensor selection in gaussian processes," in *International Conference on Control, Automation and Information Sciences (ICCAIS)*, 2012, pp. 142–147.
- [16] M. G. Jadidi, J. V. Miró, R. Valencia, and J. Andrade-Cetto, "Exploration on continuous gaussian process frontier maps," in *Proceedings of IEEE International Conference on Robotics and Automation*, 2014, pp. 6077–6082.
- [17] M. S. Corson, C. A. Rotz, and R. H. Skinner, "Evaluating warm-season grass production in temperate-region pastures: a simulation approach," *Agricultural Systems*, vol. 93, no. 1-3, pp. 252–268, 2007.
- [18] T. Zhai, R. H. Mohtar, A. R. Gillespie, G. R. von Kiparski, K. D. Johnson, and M. Neary, "Modeling forage growth in a midwest usa silvopastoral system," *Agroforestry systems*, vol. 67, no. 3, pp. 243–257, 2006.
- [19] A. J. Franzluebbers, L. K. Paine, J. R. Winsten, M. Krome, M. A. Sanderson, K. Ogles, and D. Thompson, "Well-managed grazing systems: A forgotten hero of conservation," *Journal of Soil and Water Conservation*, vol. 67, no. 4, pp. 100A–104A, 2012.
- [20] M. A. Sanderson, L. W. Jolley, and J. P. Dobrowolski, "Pastureland and hayland in the usa: Land resources, conservation practices, and ecosystem services," *Conservation outcomes from pastureland and hayland practices: Assessment, recommendations, and knowledge gaps*. Allen Press, Lawrence, KS, pp. 25–40, 2012.
- [21] H. C. J. Godfray, "Food and biodiversity," *Science*, vol. 333, no. 6047, pp. 1231–1232, 2011.
- [22] C. E. Rasmussen and Z. Ghahramani, "Infinite mixtures of gaussian

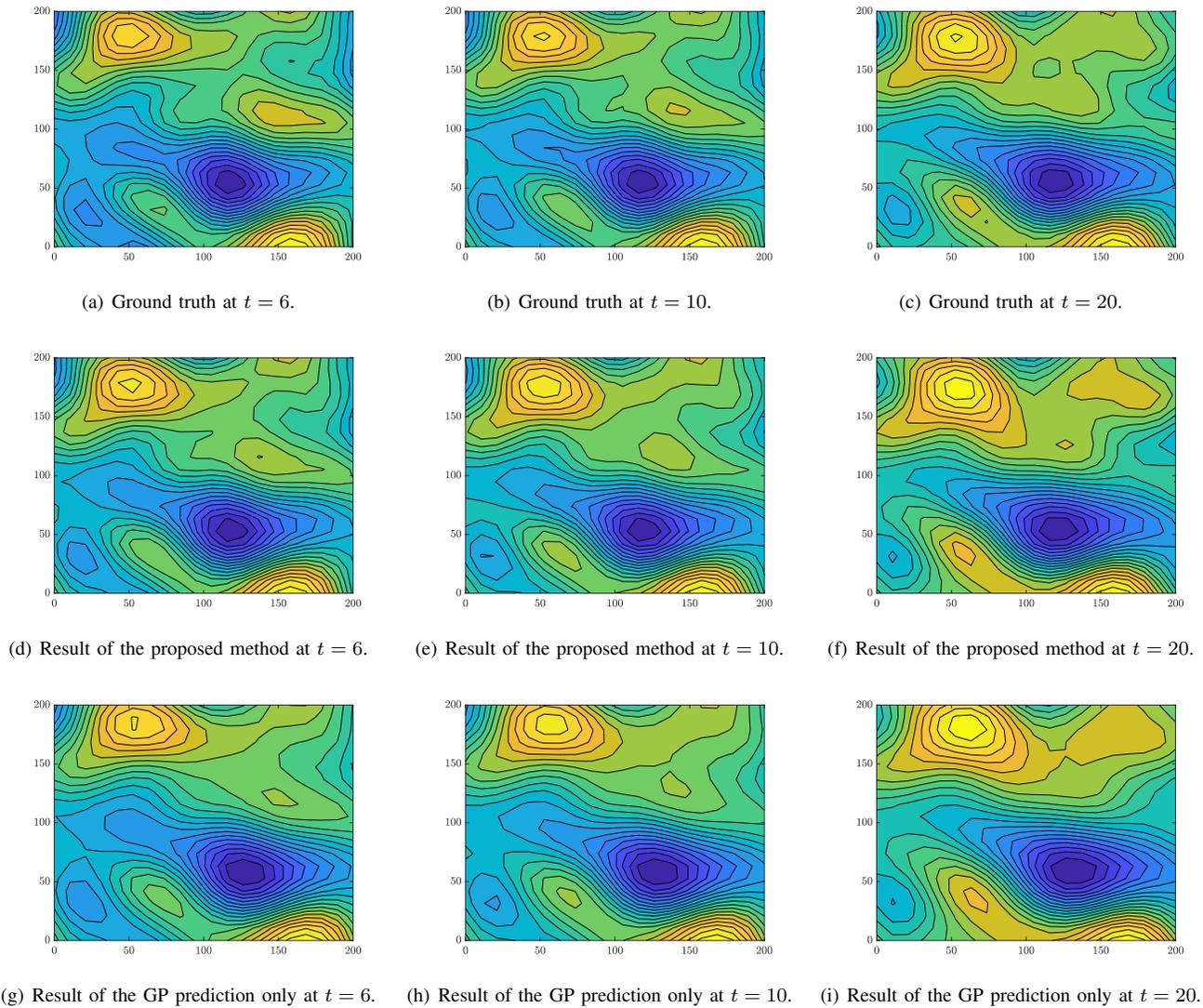


Fig. 5. A comparison of the ground truth (a)-(c), the result from the proposed hybrid model (d)-(f), and the result using GP model only (g)-(i) at time  $t = 6$ ,  $t = 10$ , and  $t = 20$ .

process experts,” in *Proceedings of Conference on Neural Information Processing Systems*, 2002, pp. 881–888.

[23] C. Stachniss, C. Plagemann, A. J. Lilienthal, and W. Burgard, “Gas distribution modeling using sparse gaussian process mixture models,” in *Proceedings of Robotics: Science and Systems*, vol. 4. MIT Press, 2008, pp. 310–317.

[24] C. Stachniss, C. Plagemann, and A. J. Lilienthal, “Learning gas distribution models using sparse gaussian process mixtures,” *Autonomous Robots*, vol. 26, no. 2-3, pp. 187–202, 2009.

[25] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[26] M. G. Jadidi, J. V. Miró, R. Valencia, J. Andrade-Cetto, and G. Disanayake, “Exploration using an information-based reaction-diffusion process,” in *Proc. Australas. Conf. Robot. Autom.*, 2013, pp. 1–10.

[27] N. C. Nguyen, H. Men, R. M. Freund, and J. Peraire, “Functional regression for state prediction using linear pde models and observations,” *SIAM Journal on Scientific Computing*, vol. 38, no. 2, pp. B247–B271, 2016.

[28] B. Oksendal, *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[29] A. Wilson and R. Adams, “Gaussian process kernels for pattern discovery and extrapolation,” in *Proceedings of International Conference on Machine Learning*, 2013, pp. 1067–1075.