2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Modeling and Decision Making in Spatio-Temporal Processes for Environmental Surveillance

Amarjeet Singh*, Fabio Ramos†, Hugh Durrant Whyte†, and William J Kaiser‡
*Indraprastha Institute of Information Technology, Delhi
†Australian Center for Field Robotics, University of Sydney
‡Department of Electrical Engineering, University of California, Los Angeles

Abstract— **The need for efficient monitoring of spatio-temporal dynamics in large environmental surveillance applications motivates the use of robotic sensors to achieve sufficient spatial and temporal coverage. A common approach in machine learning to model spatial dynamics is to use the non-parametric Bayesian framework known as *Gaussian Processes* (GPs) (*c.f.*, [1]) which are fully specified by a mean and a covariance function. However, defining suitable covariance functions that are able to appropriately model complex space-time dependencies in the environment is a challenging task. In this paper, we develop a generic approach for constructing several classes of covariance functions for spatio-temporal GP modeling. The GP models are then extended to perform efficient path planning in continuous space while maximizing the information gain. Extensive empirical evaluation for the different classes of covariance functions using real world sensing datasets is discussed, including experiments on a tethered robotic system - Networked Info Mechanical System (NIMS).**

I. Introduction

A broad class of environmental monitoring applications, including meteorology and climatology, epidemiology, ecology, demography, forestry, fishery and others, require distributed sensing capabilities [2] due to the dynamics exhibited in both space and time. Understanding and modeling such complex space time dynamics with only static sensors would require an impractically large number of sensors to be distributed across the complete spatial extent of the observed environment. Mobile robots equipped with sensors offer an alternative to a network of static sensing elements for high spatial coverage but at the cost of increased delay (sampling latency). Several path planning approaches have been proposed in the literature to adaptively sample the environment to reduce the latency while still providing high fidelity sampling [3]–[6].

At the core of each path planning approach is a model representing the space-time dependencies. Learning such complex and non-deterministic spatio-temporal dynamics, for accurate predictions, is challenging. A typical modeling procedure adopted in the field of environmental science is to manually specify partial differential equations governing the behavior of the observed phenomena [7]. This, however, relies on experts and usually requires time consuming validation experiments [8].

In this paper we consider a machine learning approach to the problem where a spatio-temporal Gaussian process model is learned through an optimization procedure. The model is then used for autonomous path planning for mobile robots. *Gaussian processes* (GPs) are a very popular non-parametric Bayesian technique for modeling spatially correlated data. Initially known as *kriging*, the technique has its roots in geostatistics where it was mainly used for resource estimation of mineral resources [9]. In a machine learning context, GPs are specified by a mean and a covariance function, and a set of (hyper) parameters which can be determined from a training set. The learning procedure maximises a Bayesian quantity (marginal likelihood) that naturally handles overfitting by possessing the Occam Razor's principle (simpler models are always preferable) [10].

However, a major difficulty for modeling spatio-temporal stochastic processes with GPs is the definition of a valid covariance function that can accurately account for space-time dependencies. In this paper, we study a generic approach for creating several classes of valid non-stationary, spatio-temporal GP models. After learning the parameters associated with the corresponding covariance model, we propose an efficient path planning strategy in continuous domains that maximizes the information gain. We present extensive empirical evaluation comparing several classes of GP models using different real world sensing datasets. Finally, we validate the proposed methodology of model learning and path planning using a Networked Info Mechanical System (NIMS, a tethered robotic system).

II. Related Work

Using GP models for environmental surveillance has been studied in a number of previous applications. In [3] a mutual information criterion was used to perform efficient path planning for multiple robots with strong approximation guarantees. Although, they used empirically learned covariance model, their path planning algorithm is independent of the covariance model. This work was extended in [11] to perform non-myopic path planning for a finite time horizon accounting for temporal correlations. However, for each of these algorithms, path planning is performed over a discrete set of observation locations distributed throughout the search space. The time complexity of these approaches will be intractable for performing path planning in a continuous space (with presumably infinite possible observation locations to select from). However, learning a suitable covariance function, as outlined in this work, is tangential to their work and can help improve the empirical performance of such path planning algorithms even in the discrete domain.

[12] proposed a decentralized approach for simultaneously learning the GP model and performing path planning. Their algorithm also constrains path planning to discrete observation locations while studying only a

specific class of covariance function (Matérn class). Our approach for efficient computation of predicted distribution (discussed in Section III-C) can be easily applied to efficient decentralized model learning. The generic approach for constructing diverse classes of covariance functions discussed here will also help in selecting a suitable covariance function for the observed environment.

GPs were also used to model the gas distribution in confined and open environments with a mobile robot in [13], and WiFi signal strength in [14]. The methodology presented here extends these approaches by explicitly modeling the time domain (dynamics), comparing different covariance functions (stationary, non-stationary, separable, non-separable), and presenting a path planning algorithm maximizing the information gain.

Kriged Kalman filters have been used in geostatistics to model the spatio-temporal distribution of dynamic environments [15], [16]. While being similar to modeling stochastic processes with spatio-temporal covariance functions, kriged kalman filter involves learning many more parameters which might lead to overfitting.

III. Gaussian Process Modeling

In this section we introduce notation and review Gaussian Processes models for regression. We consider the supervised learning problem where a set of observations $\mathcal{S} = \{\mathbf{x}_i, y_i\}_{i=1}^N$ consisting of N input locations $\mathbf{x}_i \in \mathbb{R}^D$ and the corresponding outputs $y_i \in \mathbb{R}$ is given. A Gaussian process model places a multivariate Gaussian distribution over the space of function variables $f(\mathbf{x})$ mapping input to output spaces i.e. $f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right)$, where $m(\mathbf{x})$ specifies a mean function and $k(\mathbf{x}, \mathbf{x}')$ specifies a covariance function (also called kernel). Using the learned GP model from the observations, the objective is to compute the predictive distribution $f\left(\mathbf{x}_*\right)$ at a new unobserved location $\mathbf{x}_*$.

Typically, the observations in the real environment are noisy observations of the underlying model. Observations are assumed made to be made as per $y = f(\mathbf{x}) + \epsilon$, where ϵ is the zero mean Gaussian noise with variance σ_n^2. The set of locations, model output and observations at the observed locations is represented as $(X, \mathbf{f}, \mathbf{y}) = (\{\mathbf{x}_i\}, \{f_i\}, \{y_i\})_{i=1}^N$, and at the unobserved locations as $(X_*, \mathbf{f}_*, \mathbf{y}_*) = (\{\mathbf{x}_{*,i}\}, \{f_{*,i}\}, \{y_{*,i}\})_{i=1}^N$. The zero mean ($m(\mathbf{x}) = 0$) joint Gaussian distribution can then be written as

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix}\right), \quad (1)$$

where $\mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ represents a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and K represents the covariance matrix computed between all input locations. Conditioning on the observed locations, the predictive distribution at the unobserved locations can be obtained as:

$$p\left(f_* \mid X_*, X, \mathbf{y}\right) = \mathcal{N}\left(\boldsymbol{\mu}_*, \boldsymbol{\Sigma}_*\right), \quad (2)$$

$$\boldsymbol{\mu}_* = K\left(X_*, X\right)\left[K\left(X, X\right) + \sigma_n^2 I\right]^{-1}\mathbf{y},$$
$$\boldsymbol{\Sigma}_* = K(X_*, X_*) - K(X_*, X)\left[K(X,X) + \sigma_n^2 I\right]^{-1} K(X, X_*)$$

From Equation 2, it can be observed that the predictive mean is a linear combination of N kernel functions, each centered on an observed location, $\mathbf{x}_i$, i.e. $\boldsymbol{\mu}_* = \sum_{i=1}^N \alpha_i k\left(\mathbf{x_i}, \mathbf{x}_*\right)$, with $\boldsymbol{\alpha} = \left(K\left(X, X\right) + \sigma_n^2 I\right)^{-1}\mathbf{y}$. Therefore, a GP can also be viewed as a *best unbiased linear estimator* [17], [18] in the mean squared error sense.

A. Covariance Function

For k to be a valid covariance function, it must satisfy the condition of positive-definitiveness i.e., for any $\mathbf{x}_1, \ldots, \mathbf{x}_m$, any real $a_1, \ldots, a_m$, and any positive integer m, k must satisfy the condition:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j k(\mathbf{x}_i, \mathbf{x}_j) \geq 0. \quad (3)$$

Satisfying the necessary condition of positive definiteness makes the specification of covariance functions non-trivial. Often, for simplicity, the covariance functions are assumed to be stationary i.e. only a function of $\mathbf{x} - \mathbf{x}'$. A common choice for a stationary covariance function is the squared exponential:

$$k(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T M (\mathbf{x} - \mathbf{x}')\right) \quad (4)$$

with $M = \text{diag}(\mathbf{l})^{-2}$ wherein $\mathbf{l}$ being a vector of positive numbers representing the length-scales in each dimension. In this paper, we also study a non-stationary *neural network* covariance function [19] described as:

$$k(\mathbf{x},\mathbf{x}') = \sigma_f^2 \sin^{-1}\left(\frac{\beta + 2\mathbf{x}^T M \mathbf{x}'}{\sqrt{(1+\beta+2\mathbf{x}^T M \mathbf{x})(1+\beta+2\mathbf{x}'^T M \mathbf{x}')}}\right) \quad (5)$$

Additionally, for each covariance model, we assumed a zero mean observation noise ϵ with variance σ_n^2, thereby writing the final covariance matrix as $K(X,X) + \sigma_n^2 I$.

B. Learning Hyper-Parameters

Commonly, the covariance function $k(\mathbf{x}, \mathbf{x}')$ is defined by a set of hyper-parameters $\boldsymbol{\theta}$, and written as $k\left(\mathbf{x}, \mathbf{x}' \mid \boldsymbol{\theta}\right)$, once these hyper-parameters are known. For example, in the case of the *squared-exponential* covariance function (as given in Eq. (4)), $\boldsymbol{\theta} \in \{\sigma_f, M, \sigma_n\}$ and in the case of neural network covariance function (as given in Eq. (5)), $\boldsymbol{\theta} \in \{\sigma_f, M, \beta, \sigma_n\}$. Thus, learning a GP model is equivalent to determining the hyper-parameters of the covariance function from some training dataset consisting of observations available at a set of given locations. In a Bayesian framework this can be performed by maximising the log of the marginal likelihood

$$\log p(\mathbf{y}|X, \theta) = -\frac{1}{2}\mathbf{y}^T K_y^{-1}\mathbf{y} - \frac{1}{2}\log|K_y| - \frac{N}{2}\log 2\pi, \quad (6)$$

where $K_y = K(X,X) + \sigma^2 I$ is the covariance matrix for the observations $\mathbf{y}$ made at locations X and where $\boldsymbol{\theta}(= \theta)$ represents a set of hyper-parameters specified according to a given covariance function. The marginal likelihood has three terms (from left to right), the first accounts for the data fit; the second is a complexity penalty term

(encoding the Occam's Razor principle) and the last is a normalisation constant. Maximization of the marginal likelihood requires the computation of partial derivatives w.r.t the hyper-parameters $\boldsymbol{\theta}_i$ (see [1] for further details).

C. Online Covariance Matrix Update

Assume that X_i represents the set of observation locations $\mathbf{x}_1, \ldots, \mathbf{x}_i$. After making a new observation at $\mathbf{x}_{i+1}$ the new set can then be represented as X_{i+1}. If we were to update the predicted distribution at unobserved locations X_*, as in Eq. (2), after making each new observation, we would have to compute the matrix inverse $K(X_i, X_i)^{-1} \forall i \in N$. This requires computation effort of $\mathcal{O}\left(N^3\right)$ if implemented naïvely. Instead, following a combination of the matrix inversion lemma and submatrix inversion, the inverse can be computed efficiently.

After making a new observation at $\mathbf{x}_{i+1}$, the covariance matrix for the observation locations (X_{i+1}) can be written as

$$K(X_{i+1}, X_{i+1}) = \begin{bmatrix} K(X_i, X_i) & K(X_i, \mathbf{x}_{i+1}) \\ K(\mathbf{x}_{i+1}, X_i) & K(\mathbf{x}_{i+1}, \mathbf{x}_{i+1}) \end{bmatrix}.$$

To compute the predicted covariance at unobserved locations X_*, we now need to calculate the inverse covariance matrix $K(X_{i+1}, X_{i+1})^{-1}$. Instead of calculating the inverse covariance matrix from scratch, we can calculate this inverse using submatrix inversion and the matrix inversion lemma as:

$$K(X_{i+1}, X_{i+1})^{-1} = \begin{bmatrix} K(X_i, X_i) & K(X_i, \mathbf{x}_{i+1}) \\ K(\mathbf{x}_{i+1}, X_i) & K(\mathbf{x}_{i+1}, \mathbf{x}_{i+1}) \end{bmatrix}^{-1}$$
$$= \begin{bmatrix} F_{11} & -F_{12} \\ -F_{12}^T & F_{22}^{-1} \end{bmatrix}$$

$$F_{22} = K(\mathbf{x}_{i+1}, \mathbf{x}_{i+1}) - K(\mathbf{x}_{i+1}, X_i) K(X_i, X_i)^{-1} K(X_i, \mathbf{x}_{i+1})$$
$$F_{11} = K(X_i, X_i)^{-1} + K(X_i, X_i)^{-1} K(X_i, \mathbf{x}_{i+1}) F_{22}^{-1}$$
$$K(\mathbf{x}_{i+1}, X_i) K(X_i, X_i)^{-1}$$
$$F_{12} = K(X_i, X_i)^{-1} K(X_i, \mathbf{x}_{i+1}) F_{22}^{-1}.$$

This shows that the new inverse covariance matrix after making a new observation, $K(X_{i+1},X_{i+1})^{-1}$, can be easily calculated from the previous inverse, $K(X_i,X_i)^{-1}$, in $\mathcal{O}\left(N^2\right)$.

IV. Constructing spatio-temporal covariance functions

We now consider spatio-temporal stochastic processes, $y = f(\mathbf{x}) + \epsilon$, where the inputs consist of both space and time coordinates, such that $\mathbf{x} = (\mathbf{s}; t)$ where $\mathbf{s} \in \mathbb{R}^D$; $t \in \mathbb{R}$ and ϵ is again assumed to be zero mean Gaussian noise with variance σ_n^2. Similar to the definition of GPs for spatial process we define a mean for the spatio-temporal stochastic process as $m(\mathbf{s}; t) \equiv E\left[y(\mathbf{s}; t)\right]$ and the covariance function as $k((\mathbf{s}; t), (\mathbf{s}'; t')) \equiv \operatorname{cov}\left(y(\mathbf{s}; t), y(\mathbf{s}'; t')\right); \mathbf{s}, \mathbf{s}' \in D, t > 0, t' > 0$.

A covariance function is stationary in space if it is a function of $\mathbf{h} = \mathbf{s} - \mathbf{s}'$ and stationary in time if it is a function of $u = t - t'$. In these cases we can simplify notation by writing $k((\mathbf{s}; t), (\mathbf{s}'; t')) = k(\mathbf{h}; u)$. Again, for k to be a valid covariance function, it must satisfy the condition of positive-definitiveness (*c.f.,* Eq. (3)). Additionally, if k has a set of parameters, we can write a covariance function stationary in space and time as $\operatorname{cov}(y(\mathbf{s}; t), y(\mathbf{s}+\mathbf{h}; t+u)) = k^0(\mathbf{h}; u \mid \boldsymbol{\theta})$, where k^0 satisfies Eq. (3) for all $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$, where Θ represents the superset of all possible parameter assignments in the p dimension domain.

To simplify the computational effort, another assumption regularly taken in modeling the spatio-temporal covariance functions is to assume separability i.e. the dynamics across space and time are assumed to be independent as:

$$k^0(\mathbf{h}; u \mid \theta) = k^1(\mathbf{h} \mid \theta_1) k^2(u \mid \theta_2), \tag{7}$$

where k^1 is a positive definite function in $\mathbb{R}^D$, k^2 is a positive definite function in $\mathbb{R}^1$, and $\theta =< \theta_1, \theta_2 >$. Separable functions are often chosen for their convenience but they fail to model interaction across space-time.

A. Non-Separable Spatio-Temporal Covariance Functions

To construct non-separable covariance functions we follow the procedure proposed in [20]. Assume that the stationary spatio-temporal covariance function $k(\mathbf{h}; u)$ is continuous and integrable. Then, by Bochner's Theorem [21], its spectral density is a positive and finite measure, obtained through the Fourier transform

$$\begin{aligned} g(\boldsymbol{\omega}; \tau) &= (2\pi)^{-D-1} \int\int e^{-i\mathbf{h}'\boldsymbol{\omega} - iu\tau} k(\mathbf{h}; u) d\mathbf{h} du \\ &= (2\pi)^{-1} \int e^{-iu\tau} h(\boldsymbol{\omega}; u) du, \end{aligned} \tag{8}$$

where $$h(\boldsymbol{\omega}; u) = (2\pi)^{-D} \int e^{-i\mathbf{h}'\boldsymbol{\omega}} k(\mathbf{h}; u) d\mathbf{h}.$$

We can then construct k, or equivalently g by specifying appropriate models for $h(\boldsymbol{\omega}; u)$. If $h(\boldsymbol{\omega}; u)$ is defined as

$$h(\boldsymbol{\omega}; u) = \rho(\boldsymbol{\omega}; u) \varphi(\boldsymbol{\omega}),$$

then for $k(\mathbf{h}; u)$ to be a valid (i.e. positive definite) continuous spatiotemporal function on $\mathbb{R}^D \times \mathbb{R}$, the following two conditions need to be satisfied:

- C1 For each $\boldsymbol{\omega} \in \mathbb{R}^D, \rho(\boldsymbol{\omega}; .)$ is a continuous autocorrelation function, $\int \rho(\boldsymbol{\omega}; u) du < \infty$ and $\varphi(\boldsymbol{\omega}) > 0$.
- C2 $\int \varphi(\boldsymbol{\omega}) < \infty$.

Using this generic approach, the following stationary spatio-temporal covariance functions can be constructed:

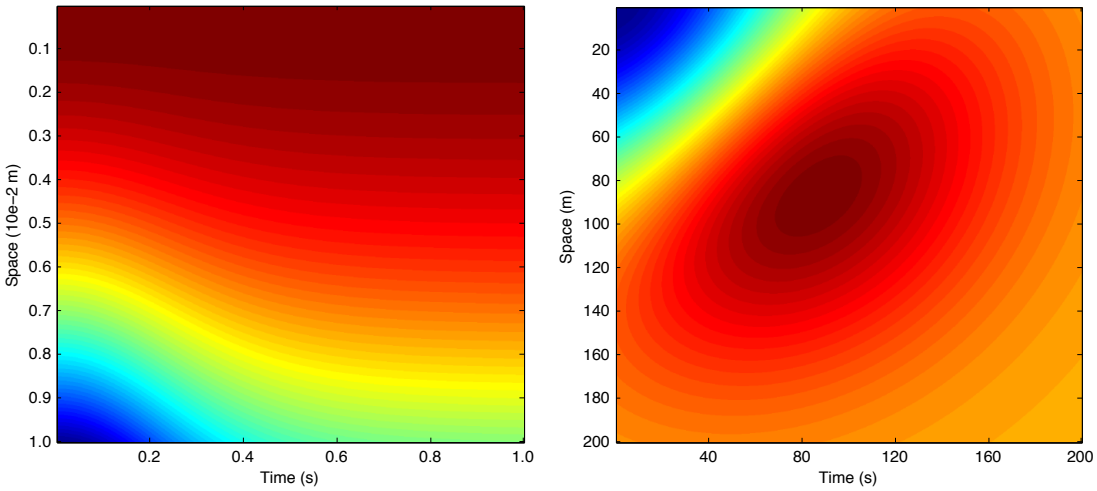
Example 1: Let $\rho(\boldsymbol{\omega}; u) = \exp\{-\|\boldsymbol{\omega}\| u^2\}$ and $\varphi = \exp\{-c_0 \|\boldsymbol{\omega}\|\}; \quad c_0 > 0$. Then the stationary, non-separable spatio-temporal covariance function (substituting c_0 = 1, without loss of generality) is given by

$$k(\mathbf{h}; u \mid \boldsymbol{\theta}) = \frac{\sigma^2 (a^2 u^2 + 1)}{[(a^2 u^2 + 1)^2 + b^2 \|\mathbf{h}\|^2]^{(D+1)/2}}, \tag{9}$$

where $\boldsymbol{\theta} = (a, b, \sigma^2)$ are the hyper-parameters with $a \geq 0$ and $b \geq \mathbf{0}$. Removing the time correlation (substituting $a = c = 0$), the covariance function reduces to a stationary spatial covariance function with

$$k(\mathbf{x}, \mathbf{x}') = \frac{\sigma^2}{[1 + b^2 \|\mathbf{h}\|^2]^{(D+1)/2}}. \tag{10}$$

Example 2: Let



(a) Example 1 output with, $\sigma = 1$, $a = 2$ and $b = 2$

(b) Example 2 output with, $\sigma = 1$, $a = 0.01$ and $b = 0.01$

Fig. 1: Covariance function output for variations in time and space. Note that in Example 1 (non-separable, non-stationary), the plot is not symmetric, indicating space-time dependencies not possible to exist in Example 2 (separable, non-stationary).

$$\rho(\boldsymbol{\omega}; u) = \frac{c_0^{D/2}}{(u^2 + c_0)^{D/2}} \exp\left\{-\frac{\|\boldsymbol{\omega}\|^2}{4(u^2 + c_0)} + \frac{\|\boldsymbol{\omega}\|^2}{4c_0}\right\}$$

and $$\varphi(\boldsymbol{\omega}) = \exp\left\{-\frac{\|\boldsymbol{\omega}\|^2}{4c_0}\right\}; \quad c_0 > 0.$$

Then the four parameter, non-separable spatio-temporal covariance function is given as:

$$k(\mathbf{h}; u \mid \boldsymbol{\theta}) = \sigma^2 \exp\{-a^2u^2 - b^2\|\mathbf{h}\|^2 - cu^2\|\mathbf{h}\|^2\} \quad (11)$$

where $\boldsymbol{\theta} = (a, b, c, \sigma^2)$ are the hyper-parameters with $a \geq 0$, $b \geq \mathbf{0}$ and $c \geq 0$. Substituting $c = 0$, we obtain a stationary, separable spatio-temporal covariance function. It is important to note that if we also remove the time correlation (substituting $a = 0$), this expression reduces to the *squared-exponential* covariance function (Eq. (4)).

B. Non-Stationary Spatio-Temporal Covariance Functions

The following corollary (Corollary 3.1 in [22]) provides a mechanism for constructing non-stationary spatio-temporal covariance functions from a spatio-temporal covariance function stationary in both space and time:

Corollary 1: If $k(\mathbf{s}-\mathbf{s}'; t-t')$ is a spatio-temporal covariance function in $\mathbb{R}^D \times \mathbb{R}$ which is stationary in both space and time, then the function $k(\mathbf{s}+\mathbf{s}'; t+t') + k(\mathbf{s}-\mathbf{s}'; t-t') - 2[k(\mathbf{s}; t) + k(\mathbf{s}'; t') - k(\mathbf{0}; 0)]$ represents a valid non-negative, non-stationary covariance function in $\mathbb{R}^D \times \mathbb{R}$.

Combining the above corollary with the proposed approach for creating non-separable, stationary spatio-temporal covariance functions (discussed in Section IV-A), we prove:

Theorem 2: Combining a continuous autocorrelation function $\rho(\omega; u)$ satisfying C1 and a function $\varphi(\boldsymbol{\omega})$ satisfying C2, one can construct a spectral density $h(\boldsymbol{\omega}; u) = \rho(\boldsymbol{\omega}; u)\varphi(\boldsymbol{\omega})$, representing a continuous and integrable function $k(\mathbf{h}; u)$ which can then be used to form a valid non-stationary spatio-temporal covariance function in $\mathbb{R}^D \times \mathbb{R}$ as $k(\mathbf{s}+\mathbf{s}'; t+t') + k(\mathbf{s}-\mathbf{s}'; t-t') - 2[k(\mathbf{s}; t) + k(\mathbf{s}'; t') - k(\mathbf{0}; \mathbf{0})]$, where $\mathbf{s}, \mathbf{s}' \in D \subset \mathbb{R}^D$ and $t, t' \in [0, \infty)$.

Proof of this theorem follows from the proof for creating stationary, spatio-temporal covariance functions [20] and the proof of Corollary 1 as given in [22]. The proof is omitted here due to space constraints. This theorem allows several classes of spatio-temporal covariance functions to be constructed. Specifically, the following are examples of separable, non-separable and non-stationary spatio-temporal covariance functions extending the stationary spatio-temporal covariance functions presented in Section IV-A:

Example 1: Non-separable, non-stationary model

$$k(\mathbf{s}, \mathbf{s}'; t, t' \mid \boldsymbol{\theta}) = \frac{\sigma^2(a^2(t+t')^2 + 1)}{[(a^2(t+t')^2 + 1)^2 + b^2\|\mathbf{s} + \mathbf{s}'\|^2]^{\frac{D+1}{2}}} + \frac{\sigma^2(a^2(t-t')^2 + 1)}{[(a^2(t-t')^2 + 1)^2 + b^2\|\mathbf{s} - \mathbf{s}'\|^2]^{\frac{D+1}{2}}} \quad (12)$$

$$-2\left[\frac{\sigma^2(a^2t^2 + 1)}{[(a^2t^2+1)^2+b^2\|\mathbf{s}\|^2]^{\frac{D+1}{2}}} + \frac{\sigma^2(a^2t'^2 + 1)}{[(a^2t'^2+1)^2+b^2\|\mathbf{s}'\|^2]^{\frac{D+1}{2}}} - \sigma^2\right].$$

Example 2: Separable, non-stationary model

$$k(\mathbf{s}, \mathbf{s}'; t, t' \mid \boldsymbol{\theta}) = \sigma^2 \exp\{-b^2\|\mathbf{s} + \mathbf{s}'\|^2 - a^2(t+t')^2\} + \sigma^2 \exp\{-b^2\|\mathbf{s} - \mathbf{s}'\|^2 - a^2(t-t')^2\} \quad (13)$$

$$-2[\sigma^2\exp\{-b^2\|\mathbf{s}\|^2 - a^2(t)^2\} + \sigma^2\exp\{-b^2\|\mathbf{s}'\|^2 - a^2(t')^2\} - \sigma^2].$$

One of the reasons for using non-separable covariance functions is to be able to model complex space and time dependencies. An example of these dependencies occurs when monitoring temperatures in indoor environments over an extensive period of time. Regions closer to windows will exhibit larger temperature variations during the day as the sunlight will heat up the surfaces nearby. At night, a more homogenous distribution is expected as the amount of energy received in the environment is more equally distributed. By modeling the space-time dependencies, it is possible to predict when and where the temperature will be higher over the period of the day.

Non-stationarity can add another level of complexity to the model. There is a vast number of non-stationary kernels ranging from the simple linear kernel ($k(\mathbf{x}, \mathbf{x}') = \mathbf{x} \cdot \mathbf{x}'$) to complex kernels obtained through process convolutions as in [**?**]. In the later case, the kernel can have parameters that vary locally according to the expected values of the outputs, providing extreme flexibility to model complex phenomenons. For computational reasons the examples above employed a different (simpler) non-stationarity formulation. The non-stationarity property is obtained by summing and subtracting kernels, as motivated by the Brownian motion.

Fig.(1a) and Fig.(1b) show the output values for the two covariance functions above for variations in time (x-axis) and space (y-axis). Fig.(1a) depicts a non-uniform dependence between space and time as a result of the non-separability in Example 1. In Fig. (1b) variations in position are not dependent on variations in time and the resulting plot is symmetric. This demonstrates the separability behavior in Example 2. Note that both cases exhibit non-stationary properties as their behavior change depending on the particular position or time coordinate.

V. Path Planning

In this section we propose an example path planning approach in continuous domain that can exploit the Gaussian Process based modeling approach to guide informative sensing. Our proposed greedy path planning algorithm uses *information gain* as the objective function. We assume that $X_t = (\mathbf{s}_1; t_1), \ldots, (\mathbf{s}_N; t_N)$ be the training set of N observation locations, available to learn the hyper-parameters of the covariance function. With the learned hyper-parameters, the trained model is used for testing on the spatio-temporal dataset. We define a timestep to be an instance of the environment during which the observed phenomena is assumed to be static.

During the testing phase, we assume that a maximum of m observations can be obtained in a single timestep; constrained by the rate at which the observed phenomena is varying and constraints of the robotic system (such as speed). For simplicity, we assume $m = 1$ in the following. A similar analysis will hold true for $m \geq 1$. At each timestep i, we have a set of locations that are already observed in the previous timesteps, given by $X_i = \mathbf{x}_1, \ldots, \mathbf{x}_i$. The predicted covariance at some set of unobserved locations X_* after making these observations will be given by $K(X_*, X_*|X_i) = K(X_*, X_*) - K(X_*, X_i)K(X_i, X_i)^{-1}K(X_*, X_i)^T$. The locations X_* can be chosen to be either uniformly spaced or randomly distributed so as to be a representative set for the corresponding dynamics in the environment. A more sophisticated approach such as that of selecting active set in [23] can also be used to chose X_*. The entropy of the phenomena can now be calculated using the log determinant of this predicted covariance matrix; $H(X_*|X_i) \propto \log|K(X_*, X_*|X_i)|$.

Let $\mathbf{x}_{i+1}$ be the next observation location selected at timestep $i + 1$. Since we have assumed $m = 1$, the time component of $\mathbf{x}_{i+1}$ will be $i + 1$. The new set of observed locations is $X_{i+1} = X_i \cup \mathbf{x}_{i+1}$. The new entropy after making observation at $\mathbf{x}_{i+1}$ is given by $H(X_*|X_{i+1}) \propto \log|K(X_*, X_*|X_{i+1})|$. The contribution to the entropy from the new observation location can be computed as:

$$I_{\mathbf{x}_{i+1}} = \log|K(X_*, X_*|X_{i+1})| - \log|K(X_*, X_*|X_i)| = \log\left(\frac{|K(X_*, X_*|X_{i+1})|}{|K(X_*, X_*|X_i)|}\right).$$

To minimize the posterior entropy after making the new observation, the next observation location is selected as:

$$\mathbf{x}_{i+1} = \text{argmin}_{\mathbf{x}_j \in \mathbb{R}^D \times \mathbb{R} \setminus X_i} \log\left(\frac{|K(X_*, X_*|(X_i \cup \mathbf{x}_j)|}{|K(X_*, X_*|X_i)|}\right). \quad (14)$$

Since this is a non-convex optimization, we use gradient ascent to find a local minimum which calculates the local gradient of the objective function. The gradient of the above function can be easily computed both analytically as well as numerically. For our experiments in Section VI, we compute the gradient numerically. If the next selected location is farther than the maximum distance traversable by the robot in a single timestep, the robot is moved in the direction of maximum gradient by the maximum distance.

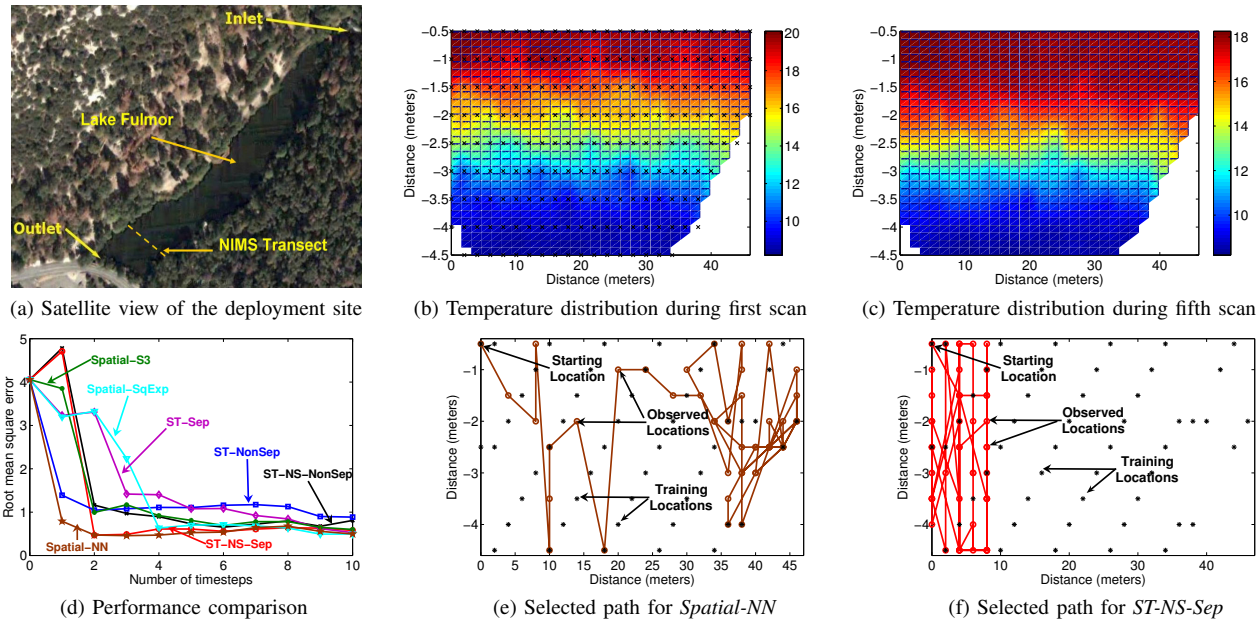
We note that such path planning is an open loop path planning, wherein the next observation location is selected only based on the locations of previously visited locations while the actual observations made at the previously visited locations are not taken into account. Hence such planning can also be performed offline.

VI. Experiments

Prediction using Gaussian Process depends on effective modeling of the covariance structure for the phenomena of interest. As discussed in the motivating applications, to represent the spectrum of modeling approaches from simple spatial models to complex non-stationary, non-separable spatio-temporal models, following classes of covariance functions were empirically compared :

1) Stationary spatial covariance functions as given in Eq. (4) (*Spatial-SqExp*) and Eq. (10) (*Spatial-S3*).
2) Non-stationary spatial covariance function (*Spatial-NN*), as given in Eq. (5).
3) Stationary, non-separable, spatio-temporal covariance function (*ST-NonSep*) as given in Eq. (9).
4) Stationary, separable, spatio-temporal covariance function (*ST-Sep*) as given in Eq. (11) (with $c = 0$).
5) Non-stationary, non-separable, spatio-temporal covariance function (*ST-NS-NonSep*) as given in Eq. (12).
6) Non-stationary, separable, spatio-temporal covariance function (*ST-NS-Sep*) as given in Eq. (13).

Real world sensing datasets were used to compare these different classes of covariance functions both in simulation and using a real robotic system. For each of the dataset and the class of covariance function, we first learned the corresponding set of hyper-parameters from a subset of collected dataset. A timestep is considered as instance of the environment during which the observed phenomena is assumed to be static. In the case of spatial covariance functions, we used the mean temperature (from all the



(a) Satellite view of the deployment site

(b) Temperature distribution during first scan

(c) Temperature distribution during fifth scan

(d) Performance comparison

(e) Selected path for *Spatial-NN*

(f) Selected path for *ST-NS-Sep*

Fig. 2: Results from lake temperature measurement (data collected using NIMS [24] at Lake Fulmor, San Jacinto Mountain Reserve).

timesteps) at the selected subset of observation locations while for learning the spatio-temporal covariance functions we used the data across all the individual timesteos. Using the learned parameters (trained model), we then perform path planning in the continuous domain as discussed in Section V. Root mean square error between the predicted observations and the actual observations at a set of grid locations, specific to each dataset, was then computed and used as the parameter for performance comparison.

A. Simulation experiments

Real world temperature measurements made in a cross-section of Lake Fulmor, located in a sub alpine coniferous forest within the San Jacinto mountains of Southern California were used to perform simulation experiments. Fig. (2a) displays satellite view of Lake Fulmor marked with the observed cross-section. The motivation for studying temperature distribution in a lake environment comes from the dependency of phytoplankton growth on temperature. Phytoplankton form the largest source of oxygen and the foundation of the food web in most aquatic ecosystems. Understanding the phytoplankton dynamics is dependent on characterizing several related parameters of which temperature is one of the most critical. Dense scans were performed using Networked Info-Mechanical System (NIMS), a tethered robotic system [24].

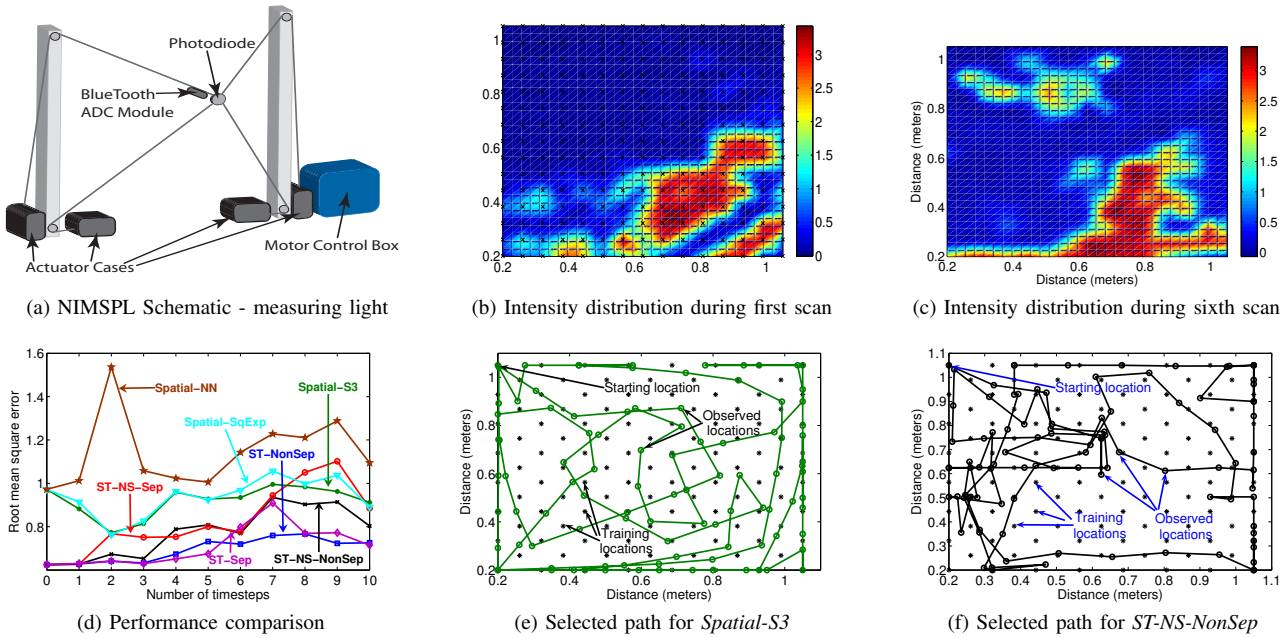
During the deployment, NIMS observed temperature at 197 locations across the lake cross-section (illustrated in Fig. (2b)). Each dense scan of the cross-section took approximately 1.5 hours and 10 such scans were taken sequentially from midnight to 4 pm. Fig. (2b) and Fig. (2c) display the temperature distribution during the first and the fifth scan. Since measurements in this dataset are only available at specific locations, the selected locations during path planning (as per Eq. (14)) in continuous domain is approximated to the nearest location for which a data sample was obtained. This constraint does not come from the proposed path planning method but is only due to the limited sensing dataset available. Such a constraint does not occur while performing path planning using actual systems (as discussed in Section VI-B).

For learning the hyper-parameters, we selected a subset of 50 uniformly spaced observation locations from the total of 197 locations. These locations are illustrated as points in Fig. (2e). For the path planning algorithm, we assumed the starting location to be one end of the cross-section (for practical purpose). We further assumed that the path planning algorithm can observe at a maximum of 5 locations for each timestep. Fig. (2d) compares all the classes of covariance models in terms of their prediction accuracy. X-axis in the plot represent the timestep and Y-axis represent the root mean square error for the corresponding timestep. Since there is little variation across the different timesteps, which is learnt by the spatio-temporal models during the process of learning the parameters, prediction accuracy for simple spatial models and complex spatio-temporal models is comprable.

Fig. (2e) and Fig. (2f) display a path selected for *Spatial-NN* and *ST-NS-Sep* covariance functions respectively during the testing phase (after the parameters have been learned from the subset of the data). It is interesting to note that since the temperature distribution is much more uniform across horizontal direction (c.f. Fig. (2b) and Fig. (2c), path planning based on both the spatial and spatio-temporal covariance functions select more locations along the vertical direction, with path planning based on the spatio-temporal covariance functions even preferring to remain on one side of the lake and only making observations along the depth.

B. System experiments

In addition to using real world sensing dataset in simulation, we also compared the performance of different



(a) NIMSPL Schematic - measuring light

(b) Intensity distribution during first scan

(c) Intensity distribution during sixth scan

(d) Performance comparison

(e) Selected path for *Spatial-S3*

(f) Selected path for *ST-NS-NonSep*

Fig. 3: Results from system experiments - measuring light intensity using NIMSPL.

covariance functions on an actual robotic system for a critical application of monitoring light intensity in the forest understory. Networked InfoMechanical System for Planar actuation (NIMS-PL), a four cable based robotic system, was used for performing the system experiments [25]. NIMS-PL consists of four tension controlled cables capable of actuating a generic sensor to provide planar spatial coverage. A schematic of NIMS-PL system actuating a light sensor in a vertical plane is shown in Fig. (3a).

A critical issue of global climate change in the world today is that of understanding and modeling carbon sources and sinks. One of the most significant components of carbon pools and fluxes is the role of tropical forests (and in particular the small plants growing in the forest understory), which hold large amounts of biomass [26]. A key challenge in understanding the CO_2 sink (or the photosynthetic rate) of plants in the understory is to characterize the dynamic light intensity patterns that exist in that environment. We used our study of spatio-temporal covariance functions to analyze the effectiveness of modeling light intensity as Gaussian Processes with spatio-temporal covariance functions.

A series of 10 images were collected to capture the light distribution under a tree canopy in San Jacinto mountains reserve (Southern California) [27] using a downlooking camera. These images were captured approximately every 10 minutes from circa 8:40 am to 10:10 am. Then, we projected these images onto a screen to be sampled using a light sensor attached to NIMS-PL. There is an ongoing work in correlating light intensity data collected using such approaches with photosynthetically active radiation that is important for understanding the photosynthetic rate. We collected light intensity data at a uniform grid of 15x15 for each projected image. The data is transmitted over bluetooth to the motor controller, to synchronize the sensor location and sensor data.

Fig. (3b) and Fig. (3c) illustrate the interpolated intensity data, at each of the 225 locations, collected using NIMSPL during first and sixth scan respectively. All the observed locations are also marked in Fig. (3b). As illustrated by the two intensity distribution plots (corresponding to light intensity distribution as captured after a separation of 50 minutes), the phenomena displays considerable dynamics and hence it is of potential interest to quantify the effectiveness of spatio-temporal covariance functions in Gaussian Process regression.

We used a subset of 112 of these 225 locations to learn the hyper-parameters of the covariance functions. These 112 locations are marked in Fig. (3e). In this case, observations were allowed to be obtained anywhere in the continuous domain and were not constrained to grid locations. Since the path planning algorithm is optimizing a non-convex optimization problem, if it is not able to find an appropriate next location to observe, we move it to the nearest unobserved location from the 225 grid locations. The path planning algorithm was allowed to observe a maximum of 10 locations for each timestep (making a total of 100 observations for 10 timesteps).

Fig. (3d) compares the root mean square error (Y-axis) for each covariance function for the corresponding timestep (X-axis). The spatio-temporal covariance functions - *ST-Sep, ST-NonSep, ST-NS-Sep, ST-NS-NonSep* provide better prediction accuracy compared to spatial covariance functions - *Spatial-S3, Spatial-SqExp, Spatial-NN*. Within the spatio-temporal covariance functions, the stationary non-separable covariance function (*ST-NonSep*) provides the best performance.

Fig. (3e) and Fig. (3f) display the path selected for *Spatial-S3* and *ST-NS-NonSep* covariance functions respectively. With *ST-NonSep* covariance function learning the time correlations, more number of observations are made in the region with high variance (light spots arriving for a short duration during the complete time duration represented by the upper

left corner as illustrated in Fig. (3c)). The simple spatial covariance function, *Spatial-S3*, makes uniform observations across the complete area. Further, both the paths tend to take a significant number of observations along the edge. Since we used the actual light sensor to measure the intensity of the projected image, the locations along the edge had significant noise from the ambient light. Large noise, increase the entropy of such locations along the edge and hence the increase in the corresponding reward achieved by visiting such locations (by minimizing the overall entropy as per Eq. (14)).

VII. Conclusion

This paper presented a detailed study on modeling spatio-temporal dynamics using Gaussian Process regression. In particular, the following contributions were made:

1) A generic framework was developed to address one of the most difficult problems in GP regression - creating *valid* covariance functions that exploit space and time inter-dependencies. The proposed approach can be used to create complex - non-stationary, non-separable, covariance functions for accurately modeling the environments while avoiding overfitting (learning the functions using marginal likelihood naturally handles overfitting by possessing the Occam Razor's principle).

2) A greedy algorithm for path planning in continuous domain was proposed. The algorithm selects observation locations that increase information gain the most (i.e. reducing the overall entropy by the maximum amount).

3) Extensive empirical evaluations were performed, comparing several classes of covariance functions - stationary spatial, non-stationary spatial, stationary separable spatio-temporal, stationary non-separable spatio-temporal, non-stationary separable spatio-temporal and non-stationary non-separable spatio-temporal. Real world sensing dataset (observing temperature distribution in a lake environment), was used to compare the performance of the path planning algorithm with these classes for modeling the observed environment. Further, a real robotic system, NIMSPL was used to run the path planning algorithm for the critical application of monitoring light intensity in the forest understory. The application of light monitoring demonstrates complementary characteristics of high temporal variation compared to our lake temperature dataset with low temporal variation. High temporal variation results in a different class of covariance functions performing well for this application. This further necessitates the applicability of such an approach that can provide covariance functions suited for specific application characteristics.

We believe that the methods presented here have a broad area of application in both robotics and sensor networks and will contribute to the development and analysis of efficient modeling and path planning algorithms for complex problems in dynamic and continuous domains. The study of spatial-temporal statistical models is of key importance for environment monitoring and this paper is an initial step towards unifying decision making and statistical modeling in robotic systems.

Acknowledgment

This work is supported by the ARC Centres of Excellence programme funded by the Australian Research Council (ARC) and the New South Wales State Government.

References

[1] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.
[2] D. Estrin, G. Pottie, , and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Internatinal Conference on Acoustics Speech and Signal Processing*, 2001, pp. 2033–2036.
[3] A. Singh, A. Krause, C. Guestrin, W. J. Kaiser, and M. A. Batalin, "Efficient planning of informative paths for multiple robots," in *IJCAI*, 2007, pp. 2204–2211.
[4] M. Rahimi, R. Pon, W. J. Kaiser, G. S. Sukhatmes, D. Estrin, and M. Srivastavan, "Adaptive sampling for environmental robotics," in *IROS*, April, 2004, pp. 3537–3544.
[5] S. T. Dmitri Dolgov, "Detection of principle directions in unknown environments for autonomous navigation," in *RSS*, 2008.
[6] D. F. Maxim Likhachev, "Planning long dynamically-feasible maneuvers for autonomous vehicles," in *RSS*, 2008.
[7] R. H. Jones and Y. Zhang, "Models for continuous stationary space-time processes," in *Modelling Longitudinal and Spatially Correlated Data, Lecture Notes in Statistics*, no. 122, 1997, pp. 289–298.
[8] P. J. Brown, N. D. Lee, and J. V. Zidek, "Multivariate spatial interpolation and exposure to air pollutants," *The Canadian Journal of Statistics*, vol. 22, no. 4, pp. 489–509, 1994.
[9] G. Matheron, "The intrinsic random functions and their applications," *Advances in Applied Probability*, vol. 5, pp. 439–468, 1973.
[10] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.
[11] A. Meliou, A. Krause, C. Guestrin, and J. M. Hellerstein, "Nonmyopic informative path planning in spatio-temporal models," in *AAAI*, 2007.
[12] R. Stranders, A. Rogers, and N. R. Jennings, "A decentralized, on-line coordination mechanism for monitoring spatial phenomena with mobile sensors," *Second International Workshop on Agent Technology for Sensor Networks*, 2008.
[13] C. Stachniss, C. Plagemann, A. Lilienthal, and W. Burgard, "Gas distribution modeling using sparse gaussian process mixture models," in *RSS*, 2008.
[14] B. Ferris, D. Haehnel, and D. Fox, "Gaussian processes for signal strength-based location estimation," in *RSS*, 2006.
[15] K. Mardia, C. Goodall, E. Redfern, and F. Alonso, "The kriged kalman filter," *Journal of the Spanish Society of Statistics and Operations Research*, vol. 7, no. 2, pp. 217–282, 1998.
[16] S. K. Sahu and K. V. Mardia, "A bayesian kriged kalman model for short-term forecasting of air pollution levels," *Journal Of The Royal Statistical Society Series C*, vol. 54, no. 1, pp. 223–244, 2005.
[17] N. Cressie, *Statistics for Spatial Data*. Wiley, 1993.
[18] P. K. Kitanidis, *Introdcution to Geostatistics: Applications in Hydrogeology*. Cambridge University Press, 1997.
[19] C. K. I. Williams, "Computation with infinite neural networks," *Neural Computation*, pp. 1203–1216, 1998.
[20] N. Cressie and H. Huang, "Classes of nonseparable, spatio-temporal stationary covariance functions," *Journal of the American Statistical Association*, vol. 94, pp. 1330–1340, 1999.
[21] S. Bochner, *Harmonic Analysis and the Theory of Probability*. University of California Press, 1955.
[22] C. Ma, "Nonstationary covariance functions that model space-time interactions," *Statistics and Probability Letters*, vol. 61, pp. 411–419, 2003.
[23] N. Lawrence, M. Seeger, and R. Herbrich, "Fast sparse gaussian process methods: The informative vector machine," in *Neural Information Processing Systems (NIPS)*, 2003.
[24] B. L. Jordan, M. A. Batalin, and W. J. Kaiser, "NIMS RD: A rapidly deployable cable based robot," in *ICRA*, 2007.
[25] P. H. Borgstrom, A. Singh, B. L. Jordan, G. S. Sukhatme, M. A. Batalin, and W. J. Kaiser, "Energy based path planning for a novel cabled robotic system," in *IROS*, 2008.
[26] R. A. Houghton, "Aboveground forest biomass and the global carbon balance," *Global Change Biology*, vol. 11, no. 6, pp. 945–958, 2005.
[27] (2005) James san jacinto mountains reserve. [Online]. Available: www.jamesreserve.edu