Adaptive Sensing for Learning Nonstationary Environment Models

Sahil Garg, Amarjeet Singh, Fabio Ramos
sahilgar@usc.edu, amarjeet@iiitd.ac.in, fabio.ramos@sydney.edu.au

Abstract—Most environmental phenomena, such as wind profiles, ozone concentration and sunlight distribution under a forest canopy, exhibit nonstationary dynamics i.e. phenomenon variation change depending on the location and time of occurrence. Non-stationary dynamics pose both theoretical and practical challenges to statistical machine learning algorithms aiming to accurately capture the complexities governing the evolution of such processes. In this paper, we address the sampling aspects of the problem of learning nonstationary spatio-temporal models, and propose an efficient yet simple algorithm - LISAL. The core idea in LISAL is to learn two models using Gaussian processes (GPs) wherein the first is a nonstationary GP directly modeling the phenomenon. The second model uses a stationary GP representing a latent space dynamics using information theory quantities to reduce the computational cost during the learning phase. The relevance of LISAL is extensively validated using multiple real world datasets.

I. INTRODUCTION

Understanding the spatio-temporal dynamics of environmental phenomena is of primal importance for sustainable development. From renewable energy generation problems such as wind speed estimation [1], to assessment of pollution impact such as ozone concentration in low altitudes [2], accurate and reliable spatio-temporal prediction is crucial from several aspects, including the development of new policies.

Given the complex dynamics of environmental phenomena (see Figure 1 for illustration), nonparametric Bayesian models have been the method of choice. These models are very flexible in capturing different characteristics of the phenomenon using different kernels, and naturally account for uncertainty in the predictions and noise in the observations. Amongst the nonparametric Bayesian methods, Gaussian Processes (GP) have been very popular in geostatistics and environmental sciences due to the analytical tractability for the posterior and marginal likelihood estimations [3], [4]. A GP is not only resilient to overfitting but also provides confidence levels which can be used to evaluate information metrics used for informative sensing such as entropy, mutual information, or the reduction of predictive variance [5], [6], [7].

The key challenge for robust prediction in environmental monitoring is to deal with varying parameterizations over the input domain necessary to model changes in the dynamics of the phenomenon, i.e. nonstationarity. To this end, nonstationary GP models have been proposed [8], [9], [10], [11], [12]. Of particular relevance for this work is the class of the generic nonstationary GPs (termed as N-GPs) that employs local hyper-parameters $\tau_{V}$ across the entire input space $X_{V} \in \mathbb{R}^{n \times 3}$ to correspondingly represent the latent space dynamics. This work, although applicable to all N-GPs, focuses primarily on two nonstationary models: 1) Process Convolution with Local Smoothing Kernels-PCLSK [10]; and 2) Latent Extension of Input Space-LEIS [12]. PCLSK and LEIS models are chosen for their intuitiveness and flexibility in modeling the environmental phenomena as compared to alternative N-GPs such as Mixture of GPs [11].

Motivated by the work of [13], [14], we propose to employ a stationary GP (latent GP) for modeling the dynamics of local hyper-parameters of a N-GP model. The main contribution is to adaptively sense the latent space dynamics using LISAL— an algorithm that builds accurate representation of the latent dynamics by adaptively learning local hyper-parameters of the N-GP, and the latent GP. LISAL reduces the overall computational cost for learning the N-GPs while providing significant improvement in phenomenon modeling. Extensive empirical validation of LISAL using three environmental sensing datasets (see complex dynamics for the datasets in Figure 1(a), 1(b), 1(c)) demonstrate the relevance of the proposed work.

II. GAUSSIAN PROCESSES

Gaussian Processes (GP) place a multivariate Gaussian distribution over the space of functions, $f(x)$, mapping the input space ($x \in \mathbb{R}^{p}$) to output space: $f(x) \sim \mathcal{N}(m(x), K(x, x'))$, where, $m(x)$ is a mean function and $K(x, x')$ is a positive definite covariance function with hyper-parameters $\theta$ [4]. Observations are modeled as noisy measurements of the function, $y = f(x) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \sigma_{e}^{2})$.

In a Bayesian setup, the hyper-parameters, $\theta$, can be learned from training data, $\{X_{V}, y_{V}\}$, by maximizing the log marginal likelihood (lml) defined as:

$$\frac{1}{2} y_{V}^{T} K_{y}^{-1} y_{V} - \frac{1}{2} \log |K_{y}| - \frac{n}{2} \log 2\pi; K_{y}=K(X_{V}, X_{V}) + \sigma_{n}^{2}I_{n}$$

Using the learned GP model, the joint Gaussian distribution of observed input locations $X_{V} \in \mathbb{R}^{n \times p}$ and test input locations $X_{T} \in \mathbb{R}^{a \times p}$, with zero mean function $m(.)$, can be written as:

$$\begin{bmatrix} y_{V} \\ f_{T} \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X_{V}, X_{V}) + \sigma_{n}^{2}I_{n} & K(X_{V}, X_{T}) \\ K(X_{T}, X_{V}) & K(X_{T}, X_{T}) \end{bmatrix}\right)$$

A popular covariance function between two input locations $x, x' \in \mathbb{R}^{3}$ is the squared exponential defined as:

$$K(x, x') = \sigma_{f}^{2} \exp\left(-\frac{1}{2} (x - x')^{T} \Sigma^{-1} (x - x')\right)$$
where \( \sigma_f \) is the signal variance, and \( \Sigma \) is a matrix with length scale parameters. In the anisotropic case, with eigenvectors aligned along the coordinate axes, \( \Sigma = \text{diag}(\llbracket l_x^2, l_y^2, l_z^2 \rrbracket) \), with \( l_x, l_y, l_z \in \mathbb{R} \).

Next, we describe two nonstationary extensions of stationary Gaussian processes, denoted as N-GPs.

1) **Process Convolution with Local Smoothing Kernels- PCLSK**: Under the process convolution framework [10], [15], a nonstationary extension, \( K_{\text{ST}}^{NS} \), of a stationary spatio-temporal covariance, \( K_{\text{ST}} \), is proposed as [14]:

\[
K_{\text{ST}}^{NS}(x_i, x_j) = \frac{|\Sigma_i|^{1/2} |\Sigma_j|^{1/2}}{|(\Sigma_i + \Sigma_j)/2|^{1/2}} \cdot K_{\text{ST}}(\sqrt{q_i^t}, \sqrt{q_j^t}),
\]

where \( q_i^t = (x_i - x_j)^T \left( \frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} (x_i - x_j) \); \( q_i^t = (t_i - t_j)^2 \left( \frac{\Sigma_i + \Sigma_j}{2} \right)^{-1} \); \( x = [x_s, t] \) (space and time coordinates); \( x_s \in \mathbb{R}^d \) and \( t \in \mathbb{R}^l \); \( \Sigma_i, \Sigma_j \) are local kernel matrices at \( x_i \) and \( x_j \) respectively. In anisotropic covariance functions, the local latent length scale parameters \( \mathbf{I}_{x_s}, \mathbf{I}_{x_t}, \mathbf{I}_{t} \in \mathbb{R}^n \) across input space \( X_V \in \mathbb{R}^{n \times 3} \) model the varying degree of phenomenon dynamics along the corresponding axes.

2) **Latent Extension of Input Space- LEIS**: In this approach, the real input space, \( X_V \), is extended with latent space dimension, \( \mathbf{I}_V \in \mathbb{R}^n \), resulting in a nonstationary covariance:

\[
K_{\text{ST}}^{NS}(x, x') = K_{\text{ST}}(\mathbf{I}_V, \mathbf{I}_V', \mathbf{I}_V) \exp \left(-\frac{1}{2} \frac{(l - l')^2}{l'} \right),
\]

where \( l_l \) is a latent length scale parameter for the latent space. See [12] for details.

Based upon the discussion of PCLSK, LEIS, from now onwards, we can consider a general N-GP that employs global hyper-parameters \( \theta_y \) and local hyper-parameters \( z_V \) (i.e. \( \{\mathbf{I}_{x_s}, \mathbf{I}_{x_t}, \mathbf{I}_{t} \} \) for PCLSK and \( \mathbf{I}_V \) for LEIS) across input space \( X_V \). As an illustration, Figure 1(d), 1(e), 1(f) show \( z_V \) learned across \( X_V \) for LEIS model using the proposed adaptive learning algorithm LISAL.

Local hyper-parameters \( z_V \) of N-GP corresponds to the latent space dynamics of an environmental phenomenon. Accuracy of \( z_V \) for representing the latent space dynamics depends upon intuitiveness of N-GP in modeling the specific phenomenon and accurate learning of \( z_V \). Considering the generic N-GPs, and specifics of PCLSK, LEIS, this work is entirely focused upon representative efficient learning of the latent space dynamics of a general class of environmental phenomena, as discussed in Section III and IV.

### III. LEARNING NONSTATIONARY GPs

Modeling environmental phenomena using N-GPs requires estimating local hyper-parameters \( z_V \) (across the input domain \( X_V \in \mathbb{R}^{n \times 3} \)), corresponding to the latent space dynamics, in addition to the global hyper-parameters \( \theta_y \). Considering the typical large expanse of input space (high value of \( n \)), either of learning \( z_V \) by directly maximizing the marginal likelihood (see (1)) or sampling from the posterior (see (2)) is computationally expensive.

\[
\{z_V^*, \theta_y^*\} = \arg \max_{z_V, \theta_y} p(y|X_V, z_V, \theta_y) \tag{1}
\]

\[
\{z_V^*, \theta_y^*\} = \arg \max_{z_V, \theta_y} p(z_V, \theta_y|X_V, y_V) \tag{2}
\]

Therefore, in Section III-A, we introduce approximations from the existing literature that result in an efficient representation of the latent space dynamics. These approximations eventually motivate the need for adaptive learning algorithm LISAL, discussed in Section IV.

#### A. Latent Space Dynamics Representation

While the latent space dynamics are expected to be complex across space and time, the degree of nonstationarity of the latent dynamics is expected to be less than the real observable dynamics (see Figure 1 for comparison on the
real and learned latent dynamics). Following existing work
that uses this intuition [13], we employ a stationary GP with
hyperc-parameters \( \theta_z \) (termed as latent GP and represented
as \( GP_z \)) for modeling the latent space dynamics. We then
learn local hyper-parameters \( z_M \) across a sparse set of input
locations \( X_M \in \mathbb{R}^{m \times d} \) s.t. \( m \ll n \) (termed as latent
locations) so as to infer joint Gaussian distribution \( z_V \) across
\( X_V \) by conditioning \( \theta_z \) on \( [X_M, z_M] \). It is important to note
that modeling the latent space dynamics with a stationary GP
ensures that N-GP model does not result in overfitting with
increase in the number of local hyper-parameters across the
latent locations (i.e. increment on \( m \)), as is also apparent
from our empirical results.

With the introduction of the induced locations \( X_M \), ex-
pressions in (1) and (2) are transformed into optimization of
intractable integrals as shown in (3) and (4) respectively.

\[
\{z_M^*, \theta_y^*, \theta_z^*\} = \arg\max_{z_M, \theta_y, \theta_z} \int p(y_V|X_V, z_V, \theta_y) p(z_V|X_V, X_M, z_M, \theta_z) \, dz_V \tag{3}
\]

\[
\{z_M^*, \theta_y^*, \theta_z^*\} = \arg\max_{z_M, \theta_y, \theta_z} \int p(z_M, \theta_y|X_V, X_M, z_V) p(z_V|X_V, \theta_y) \, dz_V \tag{4}
\]

wherein the predictive joint Gaussian distribution \( z_V \) across
\( X_V \) is marginalized so as to learn local hyper-
parameters \( (z_M) \) across \( X_M \) only.

[16] proposed Gaussian approximation that uses the law of
iterated expectations and conditional variance, to analytically
approximate the predictive joint Gaussian distribution in
terms of predictive mean and predictive variance. To avoid
complications of accounting for local hyper-parameters pre-
diction uncertainty using predictive variance, it is proposed
in [13], [14] to approximate the joint Gaussian distribution
\( z_V \) with only predictive mean \( z_V^m \). So we apply the Gaussian
approximation on the intractable integrals, shown in (3) and
(4), by approximating \( z_V \) with \( z_V^m \), to create computationally
tractable closed form expressions, as shown in (5) and (6)
respectively.

\[
\{z_M^*, \theta_y^*, \theta_z^*\} = \arg\max_{z_M, \theta_y, \theta_z} \int p(y_V|X_V, z_V^m, \theta_y) p(z_V^m|X_V, X_M, \theta_z) \, dz_V \tag{5}
\]

\[
\{z_M^*, \theta_y^*, \theta_z^*\} = \arg\max_{z_M, \theta_y, \theta_z} \int p(z_M, \theta_y|X_V, X_M, z_V^m) p(z_V^m, \theta_y|X_V, \theta_z) \, dz_V \tag{6}
\]

wherein \( z_V^m \) is the predictive mean for local hyper-
parameters, inferred by conditioning \( \theta_z \) on \( [X_M, z_M] \), \( z_V^m \)
is employed as local hyper-parameters across \( X_V \), in place of
\( z_V \), to the N-GP, and corresponds to the latent space
dynamics of the modeled environmental phenomenon.

Clearly, for \( m \ll n \), computational cost for optimization
in (5) and (6), will be significantly smaller than optimization
in (1) and (2) respectively. However, we need to establish if
the concept of induced latent locations, compromises on the
accurate representation of latent space dynamics.

Algorithm 1 LISAL

1: Input: \( \{X_V \in \mathbb{R}^{m \times d}, y_V \in \mathbb{R}^m\}, m_1, m_2, c \)
2: Output: \( \theta_z^*, \theta_y^*, X_M^* \in \mathbb{R}^{m \times c}, z_M^* \in \mathbb{R}^m \)
3: \( \theta_z^* = \arg\max_{\theta_z} p(y_V|X_V, \theta_z) \)
4: \( X_M^* = \arg\max_{X_M \subset X_V} \|X_M; X_V\setminus X_M|\theta_z^*\)
5: \( \theta_y^*, \theta_z^*, z_M^* = \text{snsLt}(X_V, y_V, X_M^*, \theta_y^*, \theta_z^*, z_M^*) \)
6: for \( i = 1 \rightarrow c \) do
7: \( X_M^*_{i+1} = \arg\max_{X_M \subset X_V} \|\{X_M_{i+1}, X^*_M\}; X_V\setminus X_M^*_{i+1}|\theta_y^*, \theta_z^*, z_M^*\)
8: \( \{\theta_y^*, \theta_z^*, z_M^*\} = \text{snsLt}(X_V, y_V, X_M^*_{i+1}, \theta_y^*, \theta_z^*, z_M^*_{i+1}) \)
9: end for

B. Limitations of Induced Points Representation

In reference to (5), if the selection of \( X_M \) is not repre-
sentative of \( X_V \) in the context of latent dynamics, the predictive
mean \( z_V^m \) (inferred by conditioning \( \theta_z \) on \( [X_M, z_M] \)),
will also inaccurately represent the latent dynamics across
\( X_V \). Inaccurate prediction of \( z_V^m \) eventually constrains the
learning of \( z_V^m \) while marginalizing the real observations \( y_V \).
Further, inaccurate learning of \( z_V^m \) leads to inaccurate learn-
ing of \( \theta_z \) and \( z_M \). Similar argument applies to optimization
in (6) as well. How do we then select induced local locations
\( X_M \) that are informative about latent dynamics across \( X_V \),
so as to perform accurate learning of N-GP?

Mutual information criterion has been a favorable choice
amongst the other Bayesian criteria, such as entropy and
reduction of predictive variance, for the problem of infor-
mative sensing of environmental dynamics [17], [18], [19].
In [14], it is proposed to quantify mutual information on real
observations \( y_V \) (see (7)) for learning latent locations \( X_M \).

\[
X_M^* = \arg\max_{X_M \subset X_V} \|X_M; X_V\setminus X_M^*|\theta_z^*; \theta_y^* p(y_V|X_V, \theta_z^*) \tag{7}
\]

However, it is hard to establish that latent locations se-
lected using mutual information on real observable var-
iance will also be informative about the latent space dynamics.
Therefore, we propose LISAL for adaptive informative sens-
ing of the latent space dynamics.

IV. LATENT INFORMATIVE SENSING WITH ADAPTIVE
LEARNING- LISAL

LISAL incrementally decides on the informative subset of
latent locations \( (X_M \subset X_V) \) using adaptive sensing of latent
space dynamics.

We first learn a stationary GP model, with hyper-
parameters \( \theta_z^* \), using the real observations (Line 3 in
Algorithm 1). Similar to the overall approach followed in
[14], LISAL then learns a preliminary set of latent locations
\( X_M^* \in \mathbb{R}^{m \times d} \) by greedily maximizing mutual information
with \( \theta_z^* \) using the algorithm proposed in [20] (Line 4 in
Algorithm 1, \( m_1 \) can be a small number s.t. \( m_1 \ll n \)).
The greedy algorithm exploits submodularity of the mutual
information criterion, providing approximation guarantee of
\(1 - \frac{1}{z}\) of the optimum, with computational cost \(O(mn)\) for selection of \(m\) out of \(n\) locations.

The next step involves learning the local hyper-parameters \(z_{M_1}\) across the latent locations \(X_{M_1}\) (see Line 5). The local hyper-parameters \(z_{M_1}\) are optimized jointly with global hyper-parameters \(\theta_g\) of the N-GP and hyper-parameters \(\theta_{z_1}\) of the stationary latent GP- \(GP_z\). Joint optimization of the parameters set \(\{\theta_{M_1}, \theta_g, \theta_{z_1}\}\) can be done either by (1) maximizing log of Gaussian approximation in (8) of marginal likelihood \(p(y_{y_1} | X_V, \theta_{y_1}, \theta_{z_1}, z_{M_1})\); or (2) Bayesian inference from Gaussian approximation in (9) of posterior \(p(\theta_{y_1}, \theta_{z_1}, z_{M_1} | X_V, y_{y_1}, X_{M_1})\) using Monte Carlo [21] or Variational Methods [22], [23]. Hereafter, whenever we refer to joint optimization of local and global hyper-parameters using marginal likelihood, it is assumed that it can also be done using posterior sampling.

\[
\{\theta_{y_1}^*, \theta_{z_1}^*, z_{M_1}^*\} = \arg\max_{\theta_{y_1}, \theta_{z_1}, z_{M_1}} \{p(y_{y_1} | X_V, \theta_{y_1}, z_{M_1}^*)
\cdot p(z_{M_1}^* | X_V, X_{M_1}, z_{M_1}, \theta_{z_1})\}
\]

\[
\theta_{y_1}^*, \theta_{z_1}^*, z_{M_1}^* = \arg\max_{\theta_{y_1}, \theta_{z_1}, z_{M_1}} \{p(z_{M_1}^* | X_V, \theta_{y_1}, \theta_{z_1})
\cdot p(z_{M_1}^* | X_{M_1}, \theta_{z_1})\}
\]

wherein \(z_{M_1}^*\) is predictive mean for local hyper-parameters across \(X_V\), inferred by conditioning \(\theta_{z_1}\) on \(z_{M_1}\). \(z_{M_1}^*\) corresponds to, although suboptimally, latent space dynamics of the environmental phenomenon.

Since \(\theta_{z_1}\) is learned on \(z_{M_1}^*\), it represents the latent space dynamics better than \(\theta_{y_1}\), learned using the real dynamics. Correspondingly, the next set of latent locations \(X_{M_2}\), learned by greedily maximizing mutual information using \(\theta_{z_1}^*\), will be more informative than learning \(X_{M_2}\) using \(\theta_{y_1}^*\). Since the informativeness of the latent locations directly corresponds to learning accuracy of hyper-parameters (see Section III-B), improved informative learning of \(X_{M_2}\) will then help in accurate learning of \(z_{M_2}\).

This improved learning concept can then be extended to iteratively learn the informative latent locations in an adaptive manner (line 7 and 8 in Algorithm 1). At the \(i^{th}\) iteration, mutual information is maximized greedily to learn the next set of latent locations \(X_{M_{i+1}}\) using \(\theta_{z_1}^*\) that is conditioned upon latent locations \(\{X_{M_1}, \cdots, X_{M_i}\}\), as shown in (10).

\[
X_{M_{i+1}} = \arg\max_{X_{M_{i+1}}} \{\{X_{M_{i+1}} \in \{X_{M_1}, \cdots, X_{M_i}\}; X_V \backslash M_{i+1}\}
\mid \theta_{z_1}^*, \{X_{M_1}, \cdots, X_{M_i}\}\}
\]

Thereafter, for sensing the latent space dynamics across \(X_{M_{i+1}}\) (line 8 in Algorithm 1), joint optimization of hyper-parameters \(\{\theta_{y_{i+1}}, \theta_{z_{i+1}}, z_{M_{i+1}}\}\) is performed by maximizing marginal likelihood, as shown in (11).

\[
\arg\max_{\theta_{y_{i+1}}, \theta_{z_{i+1}}, z_{M_{i+1}}} \{p(y_{y_{i+1}} | X_V; \theta_{y_{i+1}}, z_{M_{i+1}})
\cdot p(z_{M_{i+1}} | X_V, \theta_{y_{i+1}}, \theta_{z_{i+1}}, z_{M_{i+1}}^*)\}
\]
prepared by varying the number of timesteps (i.e. images) from 6 to 48 for the selected 36 locations.

A. Experimental Setup

We simulated a mobile robot path planning application, wherein the robot learned the next most informative location, by maximizing mutual information, using the N-GP considered, for sensing the real observation across the spatial expanse at a given timestep (a year for ozone data and a day for wind data, represented a timestep). Considering non-separable spatio-temporal covariance to be a more accurate choice for modeling environmental dynamics, examples 1 and 3 in [26] were used as covariance functions for deriving each of the N-GPs.

To simulate time delay for path traversal, the robot was constrained to make 6 observations at every timestep. These observations, together with the underlying GP model were used to predict the real phenomenon for all the test locations at a given timestep. Root Mean Square (RMS) error, between the predicted and the actual phenomenon value, was used to compare the accuracy of the underlying N-GP models.

B. Empirical Results

We first evaluated the performance of LISAL algorithm for varying values of parameters $m_1$, $m_2$ and $c$, (as used in Algorithm 1 and discussed in Section IV), using the ozone and the wind training datasets. After each iteration of learning in Algorithm 1, starting from iteration 0 (line 5 in Algorithm 1), the learned LISAL model was evaluated within the experimental setup and the mean RMS error for the test run was calculated. Note that the $0^{th}$ iteration corresponds to the offline approach proposed in [14]. Referring to Figure 2(a) and 2(b), number of latent locations ($m$) is calculated as $m_1 + m_2$. High mean RMS value for the offline approach (i.e. iteration 0), was observed. However, in a few iterations of adaptive learning with LISAL (increasing the number of latent locations $m$), mean RMS error significantly decreased, and eventually converged. From analysis of the mean RMS plots for LISAL with varying values of $m_1$, $m_2$ in Figure 2(a), 2(b), we observe that LISAL performs well even for small values of $m_1$ and $m_2$. As a result, we fixed both $m_1$ and $m_2$ to a small value of 6 for the remaining experiments.

LISAL model with the offline approach for latent space dynamics (as proposed in [14]) was then learned using entropy and mutual information criteria. These learned models were then compared with LISAL model learned with (adaptive) LISAL algorithm. Comparison of the three LISAL models for Ozone and Wind datasets is presented in Figure 2(c) and 2(d) respectively. For the $0^{th}$ iteration (with $m = 6$), since LISAL is similar to the offline approach with mutual information, corresponding mean RMS error values were same. With increase in $m$, mean RMS error for the offline approach fluctuated whereas, mean RMS error for LISAL decreased in the initial few iterations (with adaptive increment in $m$), and then converged to an (possibly local) optima.

Figure 2(e) illustrates the computational time\(^1\) for learning the LISAL model under LISAL for all the three datasets (with training samples ranging from 432 to 480). We observe that while the learning time increased linearly with increase in number of latent locations, overall time taken was small even for a large number of latent locations. Figure 2(f) evaluates the corresponding learning time variation with varying number of training samples size (from 216 to 1728), using the sunlight dataset. As expected, learning time increased polynomially with increase in the number of training samples (learning time for a GP is of the order of $O(n^3)$ for $n$ training samples). However, increase in

---

\(^1\)Experiments done using an i3 2.1 GHz CPU with 4 GB RAM
TABLE I
MEAN RMS ERROR COMPARISON FOR LEIS, PCLSK, MGPs, AND CORRESPONDING STATIONARY GPs.

| Samples S | LEIS | GP | Ozone | Wind | Sunlight |
|-----------|------|----|-------|------|---------|
| 216       | S-Ex.3 | 5.65 | 2.98 | 49.30 |
| 432       | LEIS-Ex.3 | 5.08 | 2.86 | 45.31 |
| 648       | PCLSK-Ex.3 | 5.31 | 2.70 | 43.32 |
| 864       | MGPs-Ex.3 | 5.72 | 3.02 | 47.01 |
| 1080      | MGPs-Ex.3 | 5.67 | 2.93 | 45.38 |
| 1296      | MGPs-Ex.3 | 5.81 | 2.93 | 45.38 |
| 1512      | LEIS-Ex.1 | 5.73 | 4.04 | 49.51 |
| 1728      | LEIS-Ex.1 | 4.98 | 2.86 | 48.83 |

VI. CONCLUSION AND FUTURE WORK

Proposed LISAL algorithm is an efficient approach for learning non-stationary dynamics associated with complex environmental phenomena, using Gaussian Process modeling. Extensive empirical evaluation with real world sensing datasets show that adaptively sensing the latent space dynamics, under LISAL, results in improved model learning (demonstrated using corresponding reduction in the RMS error) with very little overhead in the computation cost. Experiments with the complex spatio-temporal sunlight dataset of sample size up to 1700 demonstrate the scalability and applicability of LISAL for environmental applications. In the future, we plan to further improve upon the computation cost to efficiently learn N-GPs on very large datasets.

REFERENCES

[1] J. Haslett and A. E. Raftery, “Space-time modelling with long-memory dependence: Assessing Ireland’s wind power resource,” Journal of the Royal Statistical Society, pp. 1–50, 1989.
[2] L. Li, X. Zhang, and R. Piringer, “A spatiotemporal database for ozone in the conterminous u.s.,” in Proceedings of the Thirteenth International Symposium on Temporal Representation and Reasoning, 2006, pp. 168–176.
[3] N. Cressie, Statistics for Spatial Data. Wiley, 1991.
[4] C. Rasmussen and C. Williams, Gaussian processes for machine learning. MIT Press, 2006.
[5] M. Shevry and H. Wynn, “Maximum entropy sampling,” Journal of Applied Statistics, pp. 165–170, 1987.
[6] T. M. Cover and J. A. Thomas, Elements of Information Theory. Wiley Interscience, 1991.
[7] A. Das and D. Kempe, “Algorithms for subset selection in linear regression,” in Proceedings of the 40th annual ACM symposium on Theory of computing, 2008, pp. 45–54.
[8] P. Sampson and P. Guttorp, “Nonparametric estimation of nonstationary spatial covariance structure,” Journal of the American Statistical Association, pp. 108–119, 1992.
[9] P. Goldberg, C. Williams, and C. Bishop, “Regression with input-dependent noise: A gaussian process treatment,” NIPS, pp. 493–499, 1998.
[10] D. Higdon, J. Swall, and J. Kern, “Non-stationary spatial modeling,” in Bayesian Statistics 6, 1999, pp. 361–708.
[11] V. Tresp, “Mixtures of gaussian processes,” in NIPS, 2001, pp. 654–660.
[12] A. Meliou, A. Krause, C. Guestrin, and J. Hellerstein, “Nonmyopic informative path planning in spatio-temporal models,” in ICML, 2007.
[13] A. Krause, A. Singh, and C. Guestrin, “Near-optimal sensor placement in gaussian processes: Theory, efficient algorithms and empirical studies,” JMLR, pp. 235–284, 2008.
[14] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, “Efficient informative sensing using multiple robots,” JAIR, pp. 707–755, 2009.
[15] A. Guestrin, A. Krause, and A. Singh, “Near-optimal sensor placement in gaussian processes,” in Proc. ICML, 2005.
[16] A. Krause, R. De Freitas, A. Doucet, and M. Jordan, “An introduction to mcmc for machine learning,” Machine learning, pp. 5–43, 2003.
[17] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, “An introduction to variational methods for graphical models,” Machine learning, pp. 183–233, 1999.
[18] M. Hoffman, D. Blei, C. Wang, and J. Paisley, “Stochastic variational inference,” arXiv preprint arXiv:1206.7051, 2012.
[19] S. Garg, A. Singh, and F. Ramos, “Learning non-stationary space-time models for environmental monitoring,” Proc. AAAI, 2012, pp. 299–294.
[20] C. Paciorek and M. Schervish, “Nonstationary covariance functions for gaussian process regression,” NIPS, pp. 273–280, 2004.
[21] A. Girard, C. Rasmussen, J. Quinonero-Candela, and R. Murray-Smith, “Gaussian process priors with uncertain inputs- application to multiple-step ahead time series forecasting,” NIPS, 2003.
[22] A. Meliou, A. Krause, C. Guestrin, and J. Hellerstein, “Nonmyopic informative path planning in spatio-temporal models,” in Proc. AAAI, 2007.
[23] A. Krause, A. Singh, and C. Guestrin, “Near-optimal sensor placement in gaussian processes: Theory, efficient algorithms and empirical studies,” JMLR, pp. 235–284, 2008.
[24] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, “Efficient informative sensing using multiple robots,” JAIR, pp. 707–755, 2009.
[25] C. Guestrin, A. Krause, and A. Singh, “Near-optimal sensor placement in gaussian processes,” in Proc. ICML, 2005.
[26] A. Krause, R. De Freitas, A. Doucet, and M. Jordan, “An introduction to mcmc for machine learning,” Machine learning, pp. 5–43, 2003.
[27] M. Jordan, Z. Ghahramani, T. Jaakkola, and L. Saul, “An introduction to variational methods for graphical models,” Machine learning, pp. 183–233, 1999.
[28] M. Hoffman, D. Blei, C. Wang, and J. Paisley, “Stochastic variational inference,” arXiv preprint arXiv:1206.7051, 2012.
[29] S. Garg, A. Singh, and F. Ramos, “Efficient space-time modeling for informative sensing,” in Proc. of Sixth International Workshop on Knowledge Discovery from Sensor Data, ser. SensorKDD ’12. ACM, 2012, pp. 52–60.
[30] D. Bzdzik, A. Singh, M. Batalin, and W. Kaiser, “Multiscale sensing with stochastic modeling,” in Proceedings of the 2009 IEEE/RJS International conference on Intelligent robots and systems. IEEE Press, 2009, pp. 4657–4643.
[31] C. Rasmussen and Z. Ghahramani, “Bayesian monte carlo,” NIPS, pp. 489–496, 2003.