Multi-resolution Gaussian processes for modeling multi-scale physical processes

Somya Sharma Chatterjee
University of Minnesota - Twin Cities
USA
sharm636@umn.edu

Xiang Li
University of Minnesota - Twin Cities
USA
lixx5000@umn.edu

Vipin Kumar
University of Minnesota - Twin Cities
USA
kumar001@umn.edu

Snigdhansu Chatterjee
University of Minnesota - Twin Cities
USA
chatt019@umn.edu

ABSTRACT

In physical and earth sciences, several processes can be characterized by multi-scale and multi-level physics. They are formulated as strongly-coupled differential equations and simulating them as such, can be computationally challenging. Even in absence of physical equations or already existing theory, building multiple models to separately model the different levels of information leads to increased resource cost. Using multi-resolution Gaussian processes model we accurately forecast the multi-scale physics in a dynamical, chaotic system. An auto-regressive scheme is used to obtain precise response estimates by leveraging multi-scale and multi-fidelity information about the physical system. With resource efficient performance, the multi-resolution Gaussian processes is able to surpass fine-scale Gaussian processes model on a Lorenz-96 system dataset and a hydrology case study. Through empirical evidence, it is proved that this methodology is able to fully capture the spatial and temporal cross-correlation, which is otherwise not possible using vanilla Gaussian processes on the same datasets.

CCS CONCEPTS

• **Computing methodologies** → **Machine learning**; • **Applied computing** → **Physical sciences and engineering**.

KEYWORDS

Guassian processes, multi-scale physics, sequence modeling

ACM Reference Format:
Somya Sharma Chatterjee, Xiang Li, Vipin Kumar, and Snigdhansu Chatterjee. 2018. Multi-resolution Gaussian processes for modeling multi-scale physical processes. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Physical sciences have recently started benefiting from data-driven machine learning regimes to understand physical processes that historically relied on decades of theory development and carefully crafted simulation models. The benefits of these data-driven methods are rooted in the reduced time and the diminishing data requirements of modern methods. One such domain that benefits from these methods is that of multi-scale process modeling in fields like hydrology, crop yield estimation and cloud formation. Even more so, the advent of climate change, also calls for fast development of these powerful machine learning models in order to rapidly adapt to the various physical changes to the environment. Whether it is modeling lake temperature as a function of multi-scale weather drivers or measuring river stream-flow from surface runoff and lateral flow at basin / sub-basin levels, multitude of multi-scale problems can benefit from machine learning algorithms for parameterization research and data assimilation problems.

Applications like lake temperature modeling by the government to understand the impact of wild life and crop yield estimation by stakeholders to understand the impact on the economic supply chain are some of the scenarios where effective multi-scale physical modeling can be pivotal in improving the outcomes for all stakeholders and for resource efficiency. Even more so, in applications where a physical equation or theory has not already been established, empirical knowledge of varying scales (different temporal and spatial resolution, differing variability) of input drivers can be useful in understanding if a multi-resolution modeling method can be beneficial.

One such case study within the purview of this work is the Lorenz-96 system for understanding the behavior of dynamical, non-linear processes in climate systems. Lorenz-96 system is one of the most studied pedagogical examples and because of its resurgence in problems like parameter estimation and equation recovery in physics-informed machine learning problems [9, 14], we also attempt to study the multi-scale nature of the system in modeling higher fidelities and finer scale processes using information from lower fidelities and coarser scale processes.

For massive climate and earth systems, Gaussian Processes (GPs) allow for modeling of temporal and spatial cross-correlations among disparate data sources [12]. They are well-known for their innate propensity to easily model the spatial and temporal association among samples using their covariance function [8, 13, 19]. More

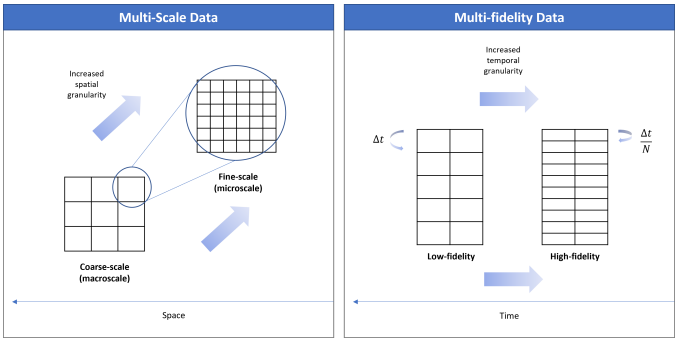


Figure 1: Multi-scale and multi-fidelity data in our experiments.

recent advances in the field have aided in more sparser and efficient estimation of covariance function using variational inference and graph theory [4, 16–18].

A more recent development [3, 6, 11] has enabled the use of autoregressive functional estimation of higher-fidelity processes using lower fidelity processes [5] where each of the physical processes can be a stochastic Gaussian process (GP). This recursive scheme ensures that the performance of inaccurate models based on lower fidelity data can be enhanced by exploiting any of the higher-fidelity data that may, otherwise, be difficult to model on its own [10]. In many applications, the higher fidelity model is required but the data is difficult to obtain due to higher costs associated with discovery or computation, while, a simpler, lower fidelity model is less accurate but also requires cheaper computations, thereby making them more accessible. Therefore, through the joint, intelligent use of both levels of data in multi-fidelity modeling, one can do away with duplicate efforts on two (or multiple) different levels of modeling.

The multi-level modeling framework that is is used in this work can be called multi-resolution GP. The multi-resolution GP aims to capture the spatial and temporal variability in the evolution of a multi-scale processes system (also shown in figure 1). In order to understand the terminology better, we can clearly differentiate between these two types of variability as follows:

(1) Fine scale process (microscale) and coarse scale process (macroscale): For the purposes of this study, we define fine scale process as having higher variability than a coarse scale process in multi-scale model. The fine scale processes tend to also have more information due to higher variability. Often times, spatial granularity is also higher for finer scale process as we need higher spatial resolution to appropriately capture the variability in the processes. In our study also, the spatial granularity is encoded as the inherent data grid structure for the multi-scale processes.
(2) High fidelity data and low fidelity data: The temporal granularity of the data can be encoded as the size of time interval after which consecutive observations are recorded for all processes in the system. Higher fidelity data in our study have more observation with small intervals between two time steps while lower fidelity data has sparsely spaced time steps.

In the experiment section, we will note that depending on the objective of modeling, the different types of variability can be helpful in obtaining better estimation of process outcomes in a multi-scale processes system.

Similar to the framework for investigating the relation among multiple fidelities using Gaussian processes, multi-scale processes can be modeled using a multi-level GP model [5, 11, 12]. The objective of this empirical study is as follows,

- The finer scale processes can be modeled as a function of coarser scale processes. Where the lower fidelity information also benefits the higher fidelity modeling in an intrinsic corregionalization autoregressive procedure.
- Sequence models can be created to study the evolution of finer scale processes. These can also be benefited from the understanding of the evolution of coarser scale processes. This can be done by modeling each scale as a level in the multi-level recursive model [5, 6]. So, a natural extension would be to use the same autoregressive framework to forecast the finer scale process using the sequence model of coarser scale models.
- For both of the above problems, we compare the performance with a standard GP trained only on finer scale / higher fidelity data to investigate the change in performance resulting from accounting for additional (coarser scale / lower fidelity) sources of information.

The next section outlines the related literature and the methodology that is used in this study. The Experiments section discusses the datasets and followed by the results obtained on those datasets, while the conclusion section provides a summary and scope for future extensions of this work.

2 METHOD

To model strongly-coupled physical processes, linear multi-output, multi-variate Gaussian process model can be used. This model was first introduced as linear multifidelity model [5] where the higher fidelity process is modeled as additive Gaussian processes components for the bias and the lower fidelity processes scaled by the magnitude of linear correlation between the lower and higher fidelity process. There bias or error component is also modeled as a Gaussian process. All the additive components are assumed to be independent processes. If we assume there is only one low fidelity process, this can be formulated as,

$$H(x) = \epsilon_H(x) + \rho\, L(x) \tag{1}$$

Here, $\epsilon(x)$ models the bias in the high fidelity process, ρ is the linear correlation between high and low fidelity processes while $L(x)$ is the Gaussian process (GP) modeling the output of low fidelity process. Using this *intrinsic corregionalization model*, we can model two consecutive fidelity processes with higher accuracy,

$$\begin{pmatrix} L(x) \\ H(x) \end{pmatrix} \sim GP\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K_L & \rho K_L \\ \rho K_L & \rho^2 K_L + K_\epsilon \end{pmatrix} \right) \tag{2}$$

As shown in Equation 2, the assumption of linear correlation among two levels of processes aid in leveraging the multivariate GP for enhanced learning of a sparsely available high fidelity (or a finer scale) process.

While many recent advances in this field have aided in leveraging information from high and low fidelity datasets, allowing for more efficient learning, many atmospheric systems also require such methods for understanding coupled differential equations. This framework can not only be utilized for integrated modeling of multi-fidelity processes but also multi-scale processes.

For instance, modeling evolution of multi-scale processes included in the microphysics processes in clouds, such as feedback from oceans and land, can be further enhanced with inclusion of not only, say, daily changes, but also hourly changes to the data. The auto-regressive scheme offered by the multiresolution GP can model different scale processes and different fidelity data as separate levels.

2.1 Lorenz-96 System

In our study, the multi-resolution is defined by two things - multi-scales of processes and two fidelities of data (obtained via different time step size). For the high fidelity data, we are interested in modeling the fine scale process using the coarse scale process. Also, the lower fidelity data is used to create a multi-variate Gaussian process model that aids in more effective learning of the high fidelity, fine scale process. As opposed to a discriminative method, the multiresolution Gaussian process not only facilitates in the prediction of high fidelity, fine scale process but also the prediction of low fidelity, fine scale process using the coarse scale process as input at both levels.

The multi-resolution Gaussian process can also be used to understand the evolution of a multiscale chaotic system. Such a canonical system is available in the form of multiscale Lorenz-96 system [1]. A recent extension [15] allows for a three scale Lorenz-96 model [7] and can be expressed in three equations as follows,

$$\frac{dX_k}{dt} = X_{k-1}(X_{k+1} - X_{k-2}) + F - \frac{hc}{b}\sum_j Y_{j,k} \tag{3}$$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}(Y_{j+2,k} - Y_{j-1,k}) - cY_{j,k} + \frac{hc}{b}X_k - \frac{he}{d}\sum_i z_{i,j,k} \tag{4}$$

$$\frac{dZ_{i,j,k}}{dt} = dZ_{i-1,j,k}(Z_{i+1,j,k} - Z_{i-2,j,k}) - geZ_{i,j,k} + \frac{he}{d}Y_{j,k} \tag{5}$$

We set the variables to similar levels as a previous deep learning study on Lorenz-96 dataset [1]. That is, $F = 20$, $h = 1$ and $b, c, d, e, g = 10$. X is the coarsest scale process with slow variability. Y is relatively finer scale while Z has the highest variability. Processes X, Y and Z affect each other. When solving the equations 3- 5, initial conditions of the processes Y and Z need to be known. However, in this machine learning framework, the initial conditions are not required and in some experiments we assume that Z does not affect the processes X and Y. This way we can focus solely on the interaction between X and Y using the multiresolution GP. We will be using the Lorenz-96 system as a running example for the purpose of explaining the methodology and the experiments on the dataset.

From the method point of view, multi-fidelity modeling can be employed for flexible and data efficient learning. This method relies on learning the lower fidelity, multi-scale data using standard Gaussian processes regression model. Based on the autoregressive scheme proposed in previous studies [5, 6], higher-fidelity processes can be modeled using lower fidelity processes as given in Equation 1. The cross-correlation among the multiple fidelities may also be non-linear and space-dependent as opposed to a simple encoding like ρ. To incorporate for these complex non-linear patterns, modifications of the kernel can enable learning for multi-fidelity data with non-linear associations among the fidelities [11].

2.2 Lower Fidelity Gaussian Process

Gaussian process regression is a non-parametric, Bayesian approach to modeling the target or output variable as a Gaussian stochastic processes. This allows for modeling the posterior predictive distribution by coupling the observed input data with a Gaussian prior that encodes our prejudice about the unknown functional form of the output. Such a Gaussian process can be fully specified via the mean and covariance function. A function $f : \mathcal{X} \rightarrow \mathbb{R}$ is modeled as a Gaussian Process with mean function m and covariance function k and is of the form [13],

$$f \sim \mathcal{GP}(m(x), k(x)), \tag{6}$$

For any collection of points $x_1, x_2, \ldots, x_m \in \mathcal{X}$, the m-dimensional random variable $\big(f(x_1), \ldots, f(x_m)\big)$ is said to have a multivariate Gaussian distribution with mean $\big(\mu(x_1), \ldots, \mu(x_m)\big)$ and a covariance matrix Σ where the (i, j)-th element is $k(x_i, x_j)$. The mean and the kernel functions are characterized by hyper-parameters θ. The unknown functional form f is given a zero mean Gaussian process prior of the form Equation 6. With a suitable choice of kernel, the prior reflects our prior beliefs about the functions attributes and captures the procedure for measuring closeness of training examples.

If $\mathbf{f}$ represents the training set function values and $\mathbf{f}_*$ is the test set function values $\mathbf{X}_* \subset \mathcal{X}$, the joint distribution can be specified as,

$$\begin{bmatrix}\mathbf{f}\\ \mathbf{f}_*\end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix}\mu\\ \mu_*\end{bmatrix}, \begin{bmatrix}\Sigma & \Sigma_*\\ \Sigma_*^T & \Sigma_{**}\end{bmatrix}\Big). \tag{7}$$

Here, Σ_* represents the train-test set covariance while Σ_{**} represents the test set covariance. The conditional distribution of $\mathbf{f}_*$ given $\mathbf{f}$ is,

$$\mathbf{f}_*|\mathbf{f} \sim \mathcal{N}(\mu_* + \Sigma_*^T\Sigma^{-1}(\mathbf{f} - \mu), \Sigma_{**} - \Sigma_*^T\Sigma^{-1}\Sigma_*) \tag{8}$$

The hyper-parameters in θ are tuned by maximizing for the logarithm of the marginal likelihood given as,

$$log\ p(\mathbf{y}|\mathbf{x}, \theta) = -\frac{1}{2}log|\Sigma| - \frac{1}{2}(\mathbf{y} - \mu)^T\Sigma^{-1}(\mathbf{y} - \mu) - \frac{n}{2}log(2\pi). \tag{9}$$

Partial derivatives of Eq. 9 give us the gradient estimate update rules for the hyper-parameters of the mean and covariance functions. And the value of hyper-parameters is usually approximated using an iterative numerical optimization technique. Depending on the different kernel functions, the definition and shape of similarity that is encoded through the kernel function can be changed.

In practice, higher fidelity or finer scale process data are more sparse due to higher cost in obtaining the observations. Due to the slow variability in the coarse scale processes, it is also relatively less challenging to model coarse scale processes over finer scale

processes. Therefore, the lowest fidelity / coarsest scale process can be effectively modeled using a standard Gaussian process model. On N_X number of training examples in coarsest scale data, a GP model m_X is trained on the coarsest scale data X to model the process evolution. On N_y number of training examples in data pertaining to process Y, $\mu_{Y,*} \in \mathbb{R}^{N_y}$ and $\Sigma_{Y,*} \in \mathbb{R}^{N_y \times N_y}$ can further be obtained by passing Y as input to m_X.

2.3 Higher Fidelity Modeling

Gaussian processes can be used to recursively model multi-scale data using coarser scale data [5, 6]. This autoregressive recursion scheme (equation 1) can also be expressed as,

$$f_j(X) = \rho f_{j-1}(X) + \epsilon_j(X) \tag{10}$$

ρ is the linear correlation between the two scales / fidelities. f_j and f_{j-1} are the GPs modeled on data for j^{th} and $j-1^{\text{th}}$ scale, respectively. Equation 10 also follows from the assumption that [5].

$$cov(f_j(x), f_{j-1}(x')|f_{j-1}(x)) = 0, \forall x \neq x' \tag{11}$$

This suggests that the higher level process at x can learn nothing new from lower level process at x' which it has not already learned from x. A modification suggested by Perdikaris et al [11] of this recursive scheme [6] allows for more efficient learning by replacing the prior f_{j-1} with learned posterior of the previous scale / fidelity level.

For modeling the finer scale process, Y, $\mu_{Y,*}$ is used as input along with observations in process data, Y, for modeling the second scale / fidelity. Similarly, for all higher scale / fidelity models, $\mu_{j-1,*} \in \mathbb{R}^{N_{j-1}}$ and $\Sigma_{j-1,*} \in \mathbb{R}^{N_{j-1} \times N_{j-1}}$ are obtained from m_{j-1} model where $\mu_{j-1,*}$ is included as an input in training m_j model for j^{th} scale / fidelity model, where $j \geq 2$. The subscript $*$ suggests the (inexact) estimates obtained on a higher level data using a lower level model.

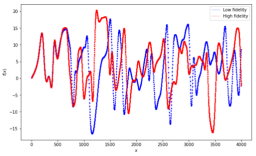
After training m_Y model, in order to train model m_Z using Z inputs, $\mu_{X,*}$ and $\Sigma_{X,*}$ are obtained from posterior prediction function m_X with Z as input data. Depending on how many samples (, say n) we want to draw to ascertain the epistemic uncertainty, $\hat{X}_{Z,*} \in \mathbb{R}^{N_Z,n}$ can be obtained as draws from multivariate normal with mean $\mu_{X,*}$ and covariance matrix $\Sigma_{X,*}$. Similarly, $\mu_{Y,*}$ and $\Sigma_{Y,*}$ are obtained again from m_Y with Z and $\hat{X}_{X,*}$ as input. For each of the examples in N_Z, the means of $\mu_{Y,*}$ and $\Sigma_{Y,*}$ are obtained to get $\mu_{Z,*}$ and $\Sigma_{Z,*}$. m_Z is trained on input Z and $\mu_{Z,*}$. This procedure ensures that a higher scale model benefits from the already learned posterior of a lower scale process model, no matter how inexact it is. This is enabled by consecutively passing Z as input of lower scale models and incorporating that posterior distribution as the input at the next level of modeling.

According to Perdikaris et al [11], the multi-scale posterior probability function $p(f_j|y_j, X_j, f_{j-1,*})$ can be obtained by decoupling the fully coupled autoregressive scheme in Kennedy Hagan [5]. The predictive mean and variance at scale / fidelity j can be expressed as,

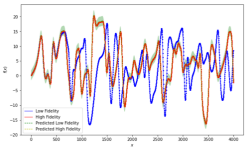
$$\mu_{j,*}(x_*) = \rho\mu_{j-1,*}(x_*) + \mu_{\epsilon_j} + \Sigma_{N_j,*}\Sigma_j^{-1}[y_j - \rho\mu_{j-1,*}(x_j) - \mu_{\epsilon_j}] \tag{12}$$

$$\sigma^2_{j,*}(x_*) = \rho^2\sigma^2_{j-1,*}(x_*) + \Sigma_{**} - \Sigma_{N_j,*}\Sigma_j^{-1}\Sigma^T_{N_j,*} \tag{13}$$

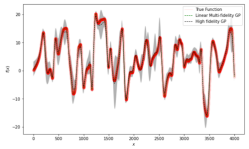
As can be seen, the mean and variance estimators of the conditional posterior predictive distribution on a level rely on the correlation coefficient, and the mean and variance of the previous level.



(a) Observations



(b) Multi-resolution GP predictions (high fidelity test MSE: 1e-18, low fidelity test MSE:4e-5)



(c) High fidelity standard GP predictions (Test MSE: 2.75)

Figure 2: High fidelity Data Predictions - Y values predictions for first cell. The Y values are displayed on Y-axis and X-axis denotes the 2000 time steps. Red solid points are observed high-fidelity data and red line are the high fidelity predictions.

In monitoring the evolution of a multiscale system like the Lorenz-96 system, often times the solving of equations is dependent on several pre-determined factors. In case of the Lorenz-96 system with three coupled differential equations, the equations and the initial conditions have to be known. In interdependent processes like the Lorenz-96, the higher frequency processes, often influence the fineness of the variability in the system along with the storage and resource requirements for computation. To alleviate the compute constraints, sequence deep learning models have been studied [1]. The evolution in the coarsest scale process, X, has been modeled in the absence of Y and Z processes data using RNNs and echo state networks [1]. While, the concern about higher computational costs are valid to eliminate Y and Z data from the experiments, the forecasting capabilities of a sequence model can significantly benefit from additional availability of information about associated processes. Especially, in real-life scenarios where the finer scale or higher fidelity data is often difficult to obtain, the additional information from coarse scale process can benefit the modeling of finer scale process and vice-versa.

The multivariate Gaussian process model is able to utilize the association between the processes, even in sparser datasets to obtain higher precision. This is studied in more detail in the next sections. In a reasonably sized dataset, the number of tunable hyperparameters in GPs are also far less than a deep learning framework.

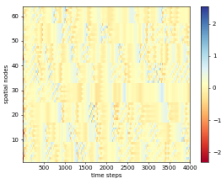
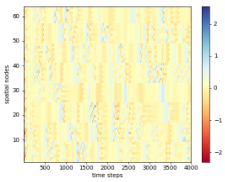
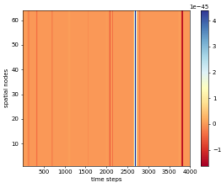
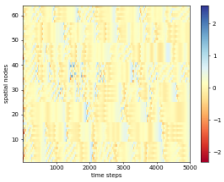
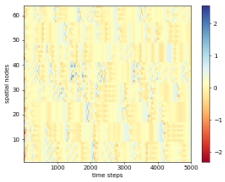
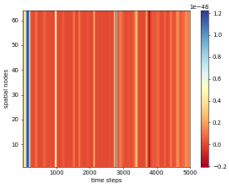
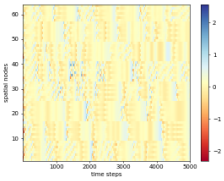
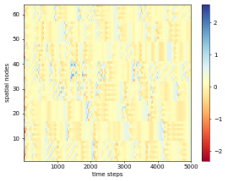
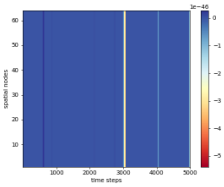
Model	Observed Outcome	Multi Resolution Estimate	Standard GP Estimate
$Y_{t+\Delta t} = f(X_{t+\Delta t})$; $\Delta t = 0.0025$; $f(X_{t+\Delta t}) = \rho f(X_{t+0.005t}) + \epsilon$	 (a)	 (b)	 (c)
$Y_{t+\Delta t} = f(X_{t+\Delta t})$; $\Delta t = 0.0020$; $f(X_{t+\Delta t}) = \rho f(X_{t+0.0025t}) + \epsilon$	 (d)	 (e)	 (f)
$Y_{t+\Delta t} = f(X_{t+\Delta t})$; $\Delta t = 0.0020$; $f(X_{t+\Delta t}) = \rho f(X_{t+0.005t}) + \epsilon$	 (g)	 (h)	 (i)

Figure 3: Multi-scale, multi-fidelity modeling using multi-resolution GP. Spatial nodes on Y axis and Time steps on X axis. Color represents the process value.

The closed form poster predictive function also ensures that complexity of the dataset is learned via the kernel function. Moreover, as opposed to a deterministic ANN, a Gaussian process based model enables estimating uncertainty in the posterior predictive function. For a fairly large dataset, where an exact Gaussian process implementation might lead to $O(n^3)$ computational costs, sparser representations through inducing variables and variational inference [4, 16] can also be explored. In our experiments, the GP implementation is based on latent variable Gaussian processes multi-outcome model [2].

3 RESULTS

We provide experiment results on one pedagogical illustration on Lorenz-96 system. Due to its mass applicability in multiple dynamical, chaotic physical systems, this case study can provide evidence of efficacy of multi-resolution Gaussian processes in multi-scale physical systems. Furthermore, we also discuss results on a hydrology dataset consisting of streamflow and lateral flow information from Headwater root river sub-basin. Such results can potentially shed light on the quality and quantity of surface and ground water run off in understanding the environmental impacts of human land use and intervening land management practices on climate change.

The first two subsections address the experimental results on the Lorenz-96 datasets followed by the results in the hydrology example.

3.1 Multi-scale, multi-fidelity modeling

The multi-level GP models can be directly utilized to estimate a physical process from another process in a multi-scale system. In the Lorenz-96 system, we will approximate finer scale processes (higher spatial resolution) as function of coarser scale processes. In a multi-resolution GP model, we can estimate the processes at higher fidelities (higher temporal resolution) as autoregressive function of lower fidelity process data.

We can first look at the results on estimating Y process as a function of X process. Higher fidelity comprises of $\Delta t = 0.0025$ while the lower fidelity data comprises of $\Delta t = 0.005$. In the first experiment, we can estimate the Y process at the same spatial resolution as X by

Model	Multi-resolution GP MSE	Standard GP MSE	Multi-resolution GP Uncertainty	Standard GP Uncertainty
$Y_{t+\Delta t} = f(X_{t+\Delta t}); \Delta t = 0.0025; f(X_{t+\Delta t}) = \rho f(X_{t+0.005t}) + \epsilon$	4e-10	0.05	0.06	0.02
$Y_{t+\Delta t} = f(X_{t+\Delta t}); \Delta t = 0.0020; f(X_{t+\Delta t}) = \rho f(X_{t+0.0025t}) + \epsilon$	2e-9	0.09	0.07	0.02
$Y_{t+\Delta t} = f(X_{t+\Delta t}); \Delta t = 0.0020; f(X_{t+\Delta t}) = \rho f(X_{t+0.005t}) + \epsilon$	5e-12	0.21	0.07	0.02

Table 1: Performance evaluation in multi-fidelity modeling scenarios in Lorenz-96 system using multi-resolution GP.

Model	Multi-resolution GP MSE	Standard GP MSE	Multi-resolution GP Uncertainty	Standard GP Uncertainty
$Y_{t+\alpha\Delta t} = f(Y_t); \alpha = 100; f(Y_t) = \rho f(X_t) + \epsilon$	1.8e-9	1.27	0.03	0.04
$Y_{t+\alpha\Delta t} = f(Y_t); \alpha = 3000; f(Y_t) = \rho f(X_t) + \epsilon$	1.8e-9	1.06	0.04	0.03
$Z_{t+\alpha\Delta t} = f(Z_t); \alpha = 3000; f(Z_t) = \rho f(X_t) + \epsilon$	4e-10	6.98	6e-5	1e-3
$Z_{t+\alpha\Delta t} = f(Z_t); \alpha = 3000; f(Z_t) = \rho f(Y_t) + \epsilon$	4e-10	0.06	6e-5	1e-3
$Z_{t+\alpha\Delta t} = f(Z_t); \alpha = 3000; f(Z_t) = \rho f(Y_t) + \epsilon; f(Y_t) = \rho' f(X_t) + \epsilon'$	0.01	0.80	6e-5	1e-3

Table 2: Performance evaluation in sequence modeling scenarios in Lorenz-96 system using multi-resolution GP.

Model	Multi-resolution GP MSE	Standard GP MSE	Lower resolution
Streamflow(t $+\alpha\Delta t$) ~ Streamflow(t $+\alpha\Delta t$); α = 3000 days; Streamflow(t) = f (Lateral Flow(t))	7.38	10.17	8.79
Streamflow(t $+\Delta t$) ~ Lateral Flow(t $+\Delta t$); Δt = 1 day; Streamflow(t $+\Delta t$) = f (Streamflow(t $+\Delta_2 t$)); Δ_2 = 50 days	1.49	7.36	6.20

Table 3: First row: Performance in sequence modeling on streamflow data using multi-resolution GP. Second row: Performance in multi-scale multi-fidelity modeling on Streamflow data using multi-resolution GP.

averaging over the 8 sub-cells in each of the 8 cells. The holdout set MSE and posterior predictive functions' standard error are shown in Fig. 2. Multi-fidelity GP enables a more accurate estimation of Y process using the multi-fidelity, multi-scale processes data. Since, the lower fidelity data is more coarsely spaced in time, it is easier to estimate the variability in the Y process with less uncertainty. The uncertainty estimates as measured are 1e-4 for higher fidelity multi-resolution GP results, 2e-5 on lower fidelity data, and 1.5 using standard GP. This suggests that the recursive modeling regime ensures lower uncertainty for a reasonably simple experiment as this one. The individual mean predictions and standard error in predictions are shown in figure 2. The Y process target at the two fidelities is visualized in the first sub-plot 2a, where the red solid points represent the higher fidelity observations while the blue solid points represent the lower fidelity data. In the sub-plot 2b, it is evident that the standard error in lower fidelity dataset is relatively lower than that of the higher fidelity prediction estimates and also lower than the standard GP standard error values on the higher fidelity data. Also, the prediction bias and standard errors are lower in multi-resolution GP results in sub-plot 2b

We further experiment with predicting Y process as a function of X process using multi-fidelity data. Figure 3 showcases the test set prediction results for all 64 cells in the response array and the table 1 outlines the performance measures. The first column in the figure outlines the autoregressive scheme. The first row corresponds with $\Delta t = 0.0025$ in the higher fidelity data and $\Delta t = 0.005$ in lower fidelity data. We can see that the standard GP is unable to fully learn the variability in the response and underestimates the response values. The results from the multi-resolution model suggest the the predictive response profile is more closely estimated to the observed response. In other higher fidelity cases (row two and three), $\Delta t = 0.0020$ for higher fidelity data. The lower fidelity data consists of $\Delta t = 0.0025$ and $\Delta t = 0.005$ for row two and three, respectively. Similar to first case, multi-resolution is able to fully approximate the spatial granularity in the response. The standard GP method is unable to obtain that level of spatial granularity using just a single fidelity of information and so it again underestimates the response. From table 1, we can see that the efficacy in estimating spatially accurate data enables the multi-resolution model to obtain higher model performance in terms of lower mean squared error and lower standard error in posterior predictive distribution.

3.2 Multi-scale sequence modeling

Extrapolation or forecasting problems are more challenging as they require good approximation behavior on out-of-distribution data. For this reason, a more interesting case to compare multi-resolution GP and standard GP model on a multi-scale data is in studying sequence modeling. The dynamical behavior seamlessly allows for sequence modeling as we are studying evolution of the multi-scale system over time. We estimate a process at a future time step using the current time step. We further, utilize a coarser scale process sequence model for more effectively approximating the finer scale future time steps. A similar example to the first experiment in multi-scale, multi-fidelity modeling section produces a test MSE of 3.75e-19 on fine scale process multi-resolution forecasting and 49.48 for the standard GP based forecasting (appendix section, in figure 6). The results pertain to experiment where we predict Y process, 100 steps into the future. The Y process values are again averaged over the 8 sub-cells in 8 cells to obtain 8 features, similar

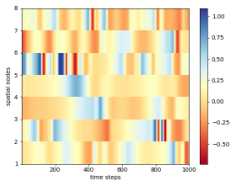
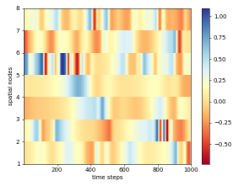
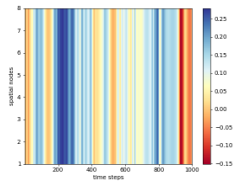
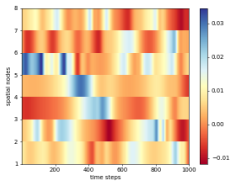
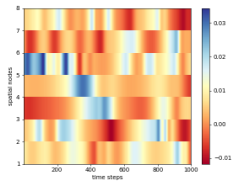
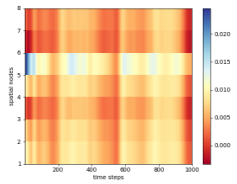
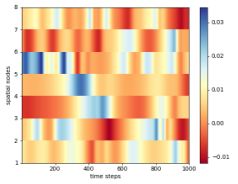
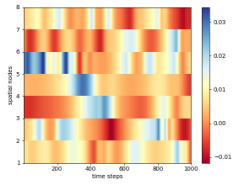
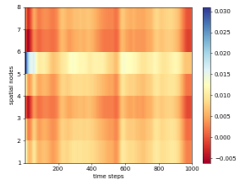
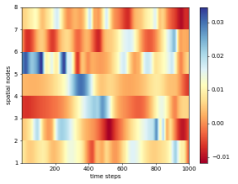
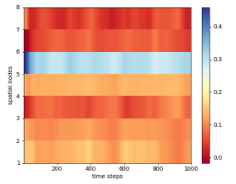
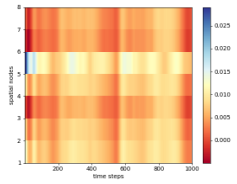
Model	Observed Outcome	Multi Resolution Estimate	Standard GP Estimate
$Y_{t+\alpha\Delta t} = f(Y_t)$; $\alpha = 3000$; $f(Y_t) = \rho f(X_t) + \epsilon$	 **(a)**	 **(b)**	 **(c)**
$Z_{t+\alpha\Delta t} = f(Z_t)$; $\alpha = 3000$; $f(Z_t) = \rho f(X_t) + \epsilon$	 **(d)**	 **(e)**	 **(f)**
$Z_{t+\alpha\Delta t} = f(Z_t)$; $\alpha = 3000$; $f(Z_t) = \rho f(Y_t) + \epsilon$	 **(g)**	 **(h)**	 **(i)**
$Z_{t+\alpha\Delta t} = f(Z_t)$; $\alpha = 3000$; $f(Z_t) = \rho f(Y_t) + \epsilon$; $f(Y_t) = \rho' f(X_t) + \epsilon'$	 **(j)**	 **(k)**	 **(l)**

Figure 4: Multi-scale sequence modeling using multi-resolution GP.

to X. Multi-resolution GP is able to perform reasonably well on the average Y process value sequence data.

The main forecasting performance results are mentioned in table 2 and individual response for the first sub-cell in Y process is chosen to be the response here. This sub-sampling ensures that the Y and X processes have matched dimensionality to utilize the auto-regressive GP procedure. The individual observed response values and mean predictions are shown in figure 4. For these experiments, α is set to 3000 while other α values also produce good results in the appendix section. The choice of α as 3000 is arbitrary and can be changed - but the main reason for choosing 3000 time steps in future is to provide some empirical evidence for good extrapolation behavior of the posterior predictive function when using multi-resolution GP; which was a cause for concern for deep learning sequence models on Lorenz-96 data [1]. In the experiments, we keep finer scale processes as the highest level that we are trying to approximate and lower level models are built for coarser scale processes.

From these experiments, we can study a few different cases: (a) How well can Z be approximated from Y? (b) How well can Z be approximated from X directly? (c) How well can a multi-level model help in utilizing X and Y in approximating Z? Looking at the table 2, we can see that as compared to a standard GP learning only the highest level data, a multi-resolution model produces the lowest MSE in all but one case, when we want to approximate Z as a response in a multi-level model. Predicting Z directly from Y as opposed to X ensures that the uncertainty estimates are also lower for the multi-resolution model. In a subset of the cases, standard

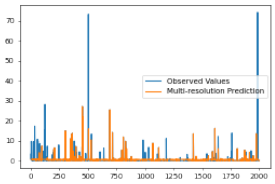
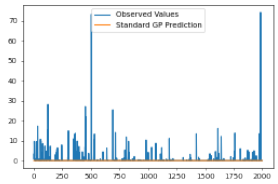
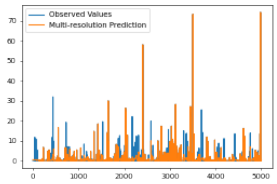
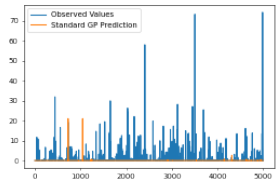
Multi Resolution Estimate (forecasting)	Standard GP Estimate (forecasting)	Multi Resolution Estimate (multi-fidelity)	Standard GP Estimate (multi-fidelity)
			
(a)	(b)	(c)	(d)

Figure 5: Multi-scale sequence modeling using multi-resolution GP in Streamflow Forecasting. Sub-plot (a) forecasts 3000 days into the future using Streamflow data and lower scale data available for Lateral flow. In sub-plot (b) the standard GP model to forecast only using streamflow data is unable to learn a good approximation for the streamflow data. Multi-scale, Multi-fidelity modeling. In sub-plot (c) several of the extreme streamflow values are well approximated using the lateral flow information at multiple fidelities. The sub-plot (d) compares the observed values with a standard Gaussian process learned using only the higher fidelity dataset.

GP is able to achieve lower uncertainty than in a multi-resolution model, suggesting the added complexity of the autoregressive routine increases the uncertainty estimates and the dispersion in the predictive distribution. In figure 4, the experiment where Y is estimated from X is able to see the highest gain in model performance when using a multi-level model over a single level GP model. This may point towards the individual functional relation of Y and X - a high amount of variance in Y can be approximated by the variance in the X process. Here, using a standard GP model results is loss of spatial granularity in the response estimates. Moreover, the worse-off performance case of multi-resolution GP in the three level dataset is evident in a very coarse temporal granularity in the prediction estimates. Also, despite the better MSE score, the standard GP response estimates in this case also lack the required variability for reasonable estimation of Z for several spatial nodes. The multi-resolution predictions are smoother and less fine than the observed response values. This may point towards the complex interdependent relation among the three processes which are being smoothed out by a data-driven method like Gaussian processes as opposed to a more accurate and compute intensive simulation model.

3.3 A case study in hydrology

In the domain of hydrology, predicting river streamflow is an imperative problem, not only because of the complex spatio-temporal serial correlations but also the complex dependency of waterflow on multitude of factors that lead to generation of flow like atmospheric temperature, solar radiation, precipitation, surface runoff, base flow and soil water content. This is also a challenging problem since snow hydrology based processes are drastically different from the streamflow generation process in flooding seasons in the summer. Moreover, this intricate multi-scale system depends on other processes like evaporation, evapotranspiration, inflow from lakes, water cycle and condensation. However, for the purpose of estimating uncertainty, we only investigate the effects of lateral flow contributed to the stream on streamflow or water yield. This means that the bias in the response estimator is likely to be higher due to the absence of multitude of information. However, this can be representative of real-life scenarios where access to all features might be limited, such as in case of developing countries where there is a dearth of ample data resources. In our case study, we try to approximate the streamflow process in the head water sub-basin based on the lateral flow process as measured in millimeters. The dataset contains daily readings for 10000 days starting on first January in the year 1902. Similar to the previous experiments, 50% data is reserved as holdout set.

3.3.1 Forecasting Streamflow. One potential problem that can be addressed through multi-resolution GP in dynamical systems is that of forecasting. Using the data, we can forecast streamflow process as a function of previous time step. For the purposes of the experiment, we forecast 3000 days in future based on current time step. Therefore, $\alpha = 3000$. Since, we assume an autoregressive scheme, streamflow can also be approximated as a function of lateral flow contributed to the stream. The autoregressive scheme for the experiments are specified in table 3. When comparing a standard GP only utilizing streamflow sequence data with a multi-resolution GP in table 3, slight improvement in bias in terms of test set MSE can be observed. The test set predictions are further studied in figure 5 also. Sub-figure 5c shows that the recursive modeling scheme enables the approximation of more variability in the response estimates as opposed to a standard GP devoid of lateral flow information when modeling streamflow. While, lateral flow in itself is not sufficient for modeling the extreme values in streamflow, even when using multi-resolution GP, this can still provide some empirical proof for the limited potential of Gaussian processes in extrapolation settings. The multi-resolution modeling also allows for pooled computational resources when modeling lower and finer scale process instead of separately creating models for the two. While the primary modeling objective was estimating streamflow, we are also able to obtain approximation of lateral flow in the same model. The modeling performance in terms of test MSE is mentioned in the last column of table 5.

3.3.2 Multi-scale, multi-fidelity Modeling. The evolution of streamflow process over several days can also be approximated using lateral flow as the input feature. In this case, we obtain two datasets - the first dataset has observations on daily level while the second dataset consists of observations from ever 50th other day. As opposed to the forecasting model, not only do we have access to the relation between progression of streamflow and lateral flow while observing their evolution, but we also have access to this functional relation on multiple fidelities. Therefore, we are able to leverage more information as opposed to a forecasting model where the primary objective of extrapolation is more challenging. In table 3, using a multi-resolution GP model, we obtain a lower test MSE than a standard GP implementation. Individual daily predictions of streamflow in the holdout set are also presented in figure 5. As opposed to extrapolation setting in the forecasting problem, the access to daily lateral flow process information enables the muli-tresolution model to reasonably approximate the extreme streamflow values as well. In the absence of this information Standard GP is unable to fully estimate the variability in the streamflow values.

4 CONCLUSION

In this work, the multi-resolution Gaussian processes (GP) model are used to investigate the benefits of an autoregressive, multi-level modeling regime over a standard, one-level model. In physical sciences, this can be especially more useful to study due to the existence of several multi-scale physical processes that are usually modeled as coupled differential equations. This multi-level GP modeling scheme strikes a balance between the conventional use of compute intensive simulation models and completely data-driven methods capable of modeling only at a single level. As a fair comparison, a multi-level GP and standard, one-level GP are used to obtain predictions and uncertainty estimates in the posterior predictive distribution. Higher-fidelity data may be difficult to obtain and therefore might have sparser dimension while a lower-fidelity (or lower-level) data may be inexact but easier to obtain. A multi-level GP model is able to utilize multi-fidelity dataset to model higher fidelity data by jointly modeling the multiple levels. Moreover, multiple scales of processes can also aid in forecasting future time steps from the current time steps.

In our experiments, we showcase that the multi-resolution GP is easily able to leverage the multi-fidelity data and surpasses the standard, one-level GP method. In the multi-scale sequence modeling problem, we are successfully able to forecast 3000 time steps in advance for the Lorenz-96 data and 3000 days in advance for the streamflow data. Despite the positive trend in the performance, it would be beneficial to look at a bigger real-world dataset. For instance, in the streamflow problem, the response estimation for greatly benefit from additional weather drivers and inclusion of surface runoff information. This can easily reduce any epistemic uncertainty due to lack of data and provide more accurate picture about the performance comparison of the methods. In the multiresolution GP model, the epistemic uncertainty is also incorporated in the mean inexact response evaluation of a fine-scale input data at a lower level GP model. The effect of fine-tuning the number of evaluations for estimating this uncertainty on the final multi-resolution model's aleatoric (and overall non-deterministic) behavior can also be interesting. Similar to a spatio-temporal multi-scale physics system, in fields like computer vision, cross-correlations between adjoining columns of pixels can also be used to create an autoregressive framework for higher-order tensor analysis. This can not only be used to encode the spatial correlations but also cross-correlations over channels and temporal scale. Previous tensor regression studies have accounted for the underlying higher dimensionality of a tensor data through modifications to the product kernel function in Gaussian processes. Similar Bayesian treatment can be achieved by tensor decomposition of multiple pixel columns or channels as multiple levels and utilizing the associations among them for efficient modeling through a similar multi-level GP modeling.

REFERENCES

[1] Ashesh Chattopadhyay, Pedram Hassanzadeh, and Devika Subramanian. 2020. Data-driven predictions of a multiscale Lorenz 96 chaotic system using machine-learning methods: reservoir computing, artificial neural network, and long short-term memory network. *Nonlinear Processes in Geophysics* 27, 3 (2020), 373–389.

[2] Zhenwen Dai, Mauricio Álvarez, and Neil Lawrence. 2017. Efficient modeling of latent information in supervised learning using gaussian processes. *Advances in Neural Information Processing Systems* 30 (2017).

[3] Loic Le Gratiet. 2012. Recursive co-kriging model for Design of Computer experiments with multiple levels of fidelity with an application to hydrodynamic. *arXiv preprint arXiv:1210.0686* (2012).

[4] James Hensman, Nicolo Fusi, and Neil D Lawrence. 2013. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835* (2013).

[5] Marc C Kennedy and Anthony O'Hagan. 2000. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* 87, 1 (2000), 1–13.

[6] Loic Le Gratiet and Josselin Garnier. 2014. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification* 4, 5 (2014).

[7] Edward N Lorenz. 1996. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, Vol. 1.

[8] Laura Martínez-Ferrer, María Piles, and Gustau Camps-Valls. 2020. Crop Yield Estimation and Interpretability With Gaussian Processes. *IEEE Geoscience and Remote Sensing Letters* (2020).

[9] Soukayna Mouatadid, Pierre Gentine, Wei Yu, and Steve Easterbrook. 2019. Recovering the parameters underlying the Lorenz-96 chaotic dynamics. *arXiv preprint arXiv:1906.06786* (2019).

[10] Benjamin Peherstorfer, Karen Willcox, and Max Gunzburger. 2018. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *Siam Review* 60, 3 (2018), 550–591.

[11] Paris Perdikaris, Maziar Raissi, Andreas Damianou, Neil D Lawrence, and George Em Karniadakis. 2017. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473, 2198 (2017), 20160751.

[12] Paris Perdikaris, Daniele Venturi, and George Em Karniadakis. 2016. Multifidelity information fusion algorithms for high-dimensional systems and massive data sets. *SIAM Journal on Scientific Computing* 38, 4 (2016), B521–B538.

[13] Carl Edward Rasmussen. 2003. Gaussian processes in machine learning. In *Summer school on machine learning*. Springer, 63–71.

[14] Stephan Rasp. 2019. Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations. *arXiv preprint arXiv:1907.01351* (2019).

[15] Tobias Thornes, Peter Düben, and Tim Palmer. 2017. On the use of scale-dependent precision in Earth system modelling. *Quarterly Journal of the Royal Meteorological Society* 143, 703 (2017), 897–908.

[16] Michalis Titsias. 2009. Variational learning of inducing variables in sparse Gaussian processes. In *Artificial Intelligence and Statistics*. 567–574.

[17] Michalis Titsias and Neil D Lawrence. 2010. Bayesian Gaussian process latent variable model. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 844–851.

[18] Dustin Tran, Rajesh Ranganath, and David M Blei. 2015. The variational Gaussian process. *arXiv preprint arXiv:1511.06499* (2015).

[19] Thomas van Klompenburg, Ayalew Kassahun, and Cagatay Catal. 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture* 177 (2020), 105709.

A EXPERIMENTS

A.1 Reproducibility

The code to create Lorenz-96 dataset at multiple scales and fidelities are given here link. Initial conditions and parameter values are set similar to the experimental conditions in related paper link. The hydrology dataset was derived from the following link. Following the methodology in [11], all models were created using the GPy package. All code for experiments will be made publicly available if the paper is accepted for publication.

A.2 Dataset

As mentioned in the previous sections, we use Lorenz-96 system as our case study to evaluate how multi-scale modeling can be done for a multi-fidelity, multi-scale dataset and for multi-scale sequence modeling. Equations 3-5, specify this multi-scale chaotic system. Again, X is the process with slowest variability, Y has relatively small amplitude and Z has the highest frequency variability and smallest amplitudes. Each of these processes have an inherent spacial structure. X has been specified using 8 nodes. The spatial resolution of the finer scale processes is higher and is captured in the number of spatial dimensions. Since, Y has a higher spatio-temporal variability, the finer information can be captured by a grid of 64 cells. Similarly, for the Z process, we have 512 cells. Therefore, X can be indexed by $k = 1, ..., 8$. Y is indexed by $k, j = 1, ..., 8$ and Z is indexed by $i, j, k = 1, ..., 8$.

Using this definition, we can monitor the spatio-temporal evolution of the three processes, $X(t)$, $Y(t)$ and $Z(t)$. As can be seen in Equations 3-5, the evolution of one process is interdependent on other processes in the system. This multi-scale system dataset is generated via a fourth-order Runge-Kutta solver.

In our study, we experiment with multiple fidelities and scales. In order to achieve that, we create multiple datasets with varying Δt as the time interval between consecutive time steps. We try $\Delta t = 0.005, 0.0025$ and 0.0020. For each of these time step intervals, we obtain X, Y and Z. The number of time steps for each of these time step intervals are 4000, 8000 and 10000, respectively. In an interpolation setting, 50 percent of data is reserved as a holdout set. And all results shown in the results section is on the unseen, holdout set.

A.3 Experiment Setup

As we have noted above, $X \in \mathbb{R}^8$, $Y \in \mathbb{R}^{8\times8}$ and $Z \in \mathbb{R}^{8\times8\times8}$. For ease of interpretation, this translates to 8 features for process X, 64 features for process Y and 512 features for process Z. In order to study the evolution of the Lorenz-96 system, $X(t+\alpha\Delta t)$, $Y(t+\alpha\Delta t)$ and $Z(t+\alpha\Delta t)$ can be approximated from $X(t)$, $Y(t)$ and $Z(t)$, respectively. While Δt determines the interval time between two time steps, α is the number of time steps in the future that can be estimated from the current time step. As pedagogical examples, we first experiment with $\alpha = 100$ and then $\alpha = 3000$. This forecasting problem of estimating the same process in future as a function of current time step relies on the assumption that the dimensionality of the predictor set for multiple resolution remains the same. This suggests that inherent spatio-temporal resolution has to be made uniform for modeling in the forecasting setting. In the multi-resolution setting, we forecast a future time step in a finer scale process using current time step and also utilize the association with coarser-scale processes to improve the modeling for finer scales.

Notwithstanding the clear functional relation between two time steps in the evolution of a process, the interdependence amongst the processes also allow for modeling a process as a functional approximation of another. Owing to this relation, we can also model the evolution of a process as the evolution of related processes in that time step, irrespective of the inherent spatial resolution of the two processes. It is also interesting to note, that in the Lorenz-96 system, since a finer scale process also has an inherent fine scale spatial representation in our current setup, we are approximating a higher dimension outcome as a function of a lower dimension predictor set. In this setting, the lower dimensional input might not be able to fully approximate the spatial fineness of the finer scale process. To overcome this, lower fidelity data pertaining to the same system can aid in imbuing the spatial granularity in the finer scale process.

We also compare the use of a standard GP implementation that assumes that we do not have access to more information for multi-resolution modeling. Therefore, in case of the sequence modeling problem, the standard GP assumes that we do not have access to information from other scale processes and have to rely on the one scale process for forecasting future outcomes for the process. In the second case of multi-scale interdependence modeling, standard GP assumes that we do not have the multi-scale system information at lower fidelities or lower temporal granularity.

For the standard GP and the multi-resolution GP, we evaluate and compare the results in terms of the test mean squared error values and the aleatoric uncertainty estimate as estimated from the standard error in the posterior predictive function. Epistemic uncertainty can be estimated from repeated measurements of response estimates to account for the deficiencies in the model and input data that the model in unaware of. For the scope of the current study, we only limit to evaluating the aleatoric non-deterministic behavior of Gaussian processes which can be obtained fairly easily from the posterior covariance function estimate.

B FORECASTING EVOLUTION OF LORENZ 96 SYSTEM

Some more experiment results can be observed in Fig. 6. Similar to the earlier results on sequence modeling, we notice that the multi-resolution GP outperforms the standard GP implementation. For different α values, some experimental results suggest that standard GP is unable to capture the spatial granularity of the response values. The results in the last row of Fig. 6 suggest that while multiresolution GP is capable of capturing spatial resolution, there might be some evidence that it is not fully able to capture temporal resolution of the data when modeling the response estimates.

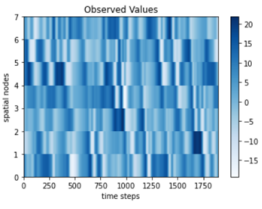
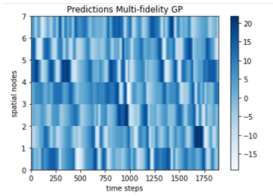
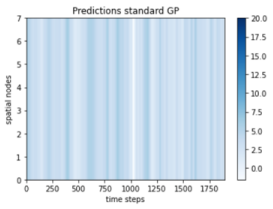
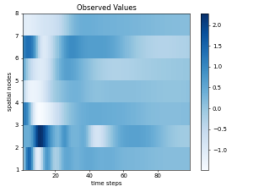
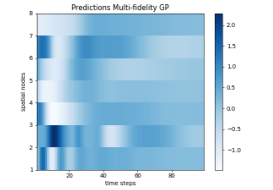
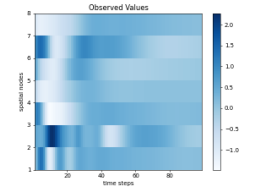
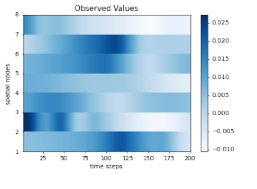
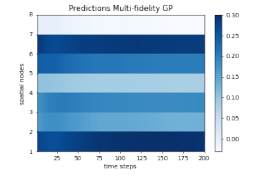
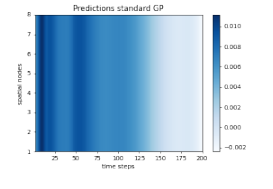
Model	Observed Outcome	Multi Resolution Estimate	Standard GP Estimate
$Y_{t+\alpha\Delta t} = f(Y_t)$; $\alpha = 100$; $f(Y_t) = \rho f(X_t) + \epsilon$; Test set size = 2000	 **(a)**	 **(b)**	 **(c)**
$Y_{t+\alpha\Delta t} = f(Y_t)$; $\alpha = 1$; $f(Y_t) = \rho f(X_t) + \epsilon$; Test set size = 100	 **(d)**	 **(e)**	 **(f)**
$Z_{t+\alpha\Delta t} = f(Z_t)$; $\alpha = 2000$; $f(Z_t) = \rho f(Y_t) + \epsilon$; $f(Y_t) = \rho' f(X_t) + \epsilon'$	 **(g)**	 **(h)**	 **(i)**

Figure 6: Multi-scale sequence modeling using multi-resolution GP for $\alpha = 1, 100, 2000$.