International Journal for Uncertainty Quantification, 5 (4): 375–392 (2015)

LOW-COST MULTI-DIMENSIONAL GAUSSIAN PROCESS WITH APPLICATION TO UNCERTAINTY QUANTIFICATION

***Bledar A. Konomi*[1] *& Guang Lin*[2,*]**

[1] *Department of Mathematical Sciences, University of Cincinnati, Cincinnati, Ohio 45221, USA*

[2] *Department of Mathematics and School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907, USA*

Original Manuscript Submitted: 05/30/2013; Final Draft Received: 07/17/2014

Computer codes simulating physical systems often have responses that consist of a set of distinct outputs that evolve in space and time and depend on many uncertain input parameters. The high dimensional nature of these computer codes makes the computations of Gaussian process (GP)-based emulators infeasible, even for a small number of simulation runs. In this paper we develop a covariance function for the GP to explicitly treat the covariance among distinct output variables, input variables, spatial domain, and temporal domain and also allows for Bayesian inference at low computational cost. We base our analysis on a modified version of the linear model of coregionalization (LMC). The proper use of the conditional representation of the multivariate output and the separable model for different domains leads to a Kronecker product representation of the covariance matrix. Moreover, we introduce a nugget to the model which leads to better statistical properties (regarding predictive accuracy) of the multivariate GP without adding to the overall computational complexity. Finally, the prior specification of the LMC parameters allows for an efficient Markov chain Monte Carlo (MCMC) algorithm. Our approach is demonstrated on the Kraichnan-Orszag problem and Flow through randomly heterogeneous porous media.

KEY WORDS: *multivariate Gaussian process, linear model of coregionalization, separability, Markov chain Monte Carlo, computer codes (experiments)*

1. INTRODUCTION

Complex computer codes (or simulators) have been garnering much attention in recent years because of their ability to make useful predictions about real world systems in many fields of science and engineering. To obtain accurate statistics, standard Monte Carlo (MC)-based methods require thousands of simulation runs without taking into consideration the cost of deterministically solving the computer code (simulator). Different uncertainty quantification (UQ) methods based on generalized polynomial chaos (gPC) [1] and Gaussian process (GP) [2] have been proposed to emulate the simulator outputs to a high degree of precision using only a few hundred runs of the simulator. GP is successfully used as an emulator because of its simplicity and non-parametric features [2]. Moreover, when an output is multivariate it can directly model their dependencies as shown in [3–6]. These methods are based on the use of the separable covariance function between the input and the multivariate output [7]. More recently, [8] uses this model and Bayesian tree techniques to build a non-stationarity multivariate covariance function. [9] generalized the separable covariance function to deal with multi-dimensional spatio-temporal computer code output. The proposed covariance function models the dependencies between distinct outputs, the input, spatial domain, and time.

Despite the successful applications and the attractive computational efficiency, in practice the assumptions associated with the separable covariance function may be violated. For example, not all the distinct outputs are likely to

*Correspond to Guang Lin, E-mail: guanglin@purdue.edu

2152–5080/15/$35.00 © 2015 by Begell House, Inc.

have the same correlation parameters over the input, spatial domain or time. Another unrealistic assumption of the separable model is the symmetry of the covariance function; $\mathrm{cov}(\boldsymbol{\eta}_l(\boldsymbol{x}_i), \boldsymbol{\eta}_{l'}(\boldsymbol{x}_j)) = \mathrm{cov}(\boldsymbol{\eta}_{l'}(\boldsymbol{x}_i), \boldsymbol{\eta}_l(\boldsymbol{x}_j))$ for all i, j, l, l'. More generalized covariance functions have been proposed in spatial statistics literature to model the cross-covariances of the multivariate output, see [10, 11] for recent reviews. The linear model of coregionalization (LMC) [12, 13] is a popular covariance function for spatial multivariate output, based on linear transformations of independent latent processes, which overcomes the above restrictions of the separable model. Different variations of LMC have been proposed to deal with the computational difficulties and types of nonstationarity in the variance [14, 15]. [16] uses a different approach to model cross-covariance functions based on latent dimensions. However, the computational cost associated with these models is not affordable for statistical inference in the case where the output is a multi-dimensional spatio-temporal process.

This paper introduces a multi-dimensional spatio-temporal covariance function which allow Bayesian inference at low computational cost and also explicitly models the dependencies between the distinct outputs. We base our approach on the use of a generalized LMC for the covariance function of the multi-dimensional spatio-temporal process. We further investigate ways to simplify the computations. The conditional representation of the LMC and the proper use of a separable covariance function inside each conditional model enable us to write the huge covariance matrix as a Kronecker product of smaller ones leading to efficient algorithms for carrying out inference and predictions. All the distinct outputs in the model have their own input, spatial, and time correlation parameters, in contrast with the separable model which assumes same correlation parameters for all the outputs in the model. Moreover, the new covariance function is not symmetric. In addition, we introduce computational stability in our method by adding nugget effects into the correlation functions. In this way, the computational complexity of the proposed method remains the same. The GP defines a probability measure over the input space and can be used to sample more data and predict the response surface. The uncertainty of the GP model quantifies the lack of information we have about the real response due to the finite number of samples.

The rest of the paper is organized as follows: In Section 2 we describe the multi-dimensional spatio-temporal computer codes. In Section 3 we describe the multivariate GP and the modeling of the covariance function. Section 4 describes the Bayesian inference for parameter estimations and predictions, including details of the implementation. In Section 5 we apply our GP models to quantify the uncertainty in two different problems, Kainchnan-Orszag three-mode problem and the flow through randomly heterogeneous porous media, and compare it to the GP with separable covariance function. Conclusions are presented in Section 6.

2. PROBLEM DESCRIPTION

Let us consider a physical problem with input domain $\mathcal{X}_\xi \subset \mathbb{R}^{k_\xi}$, spatial domain $\mathcal{X}_s \subset \mathbb{R}^{k_s}$, and temporal domain which is expressed as an interval $\mathcal{X}_T = [0, T]$, where k_ξ, k_s are the dimensions of the input and spatial domain. The input domain $\mathcal{X}_\xi$ usually represents a bounded domain and can thus be considered a compact subset of $\mathbb{R}^{k_\xi}$ while the spatial domain $\mathcal{X}_s$ and time domain $\mathcal{X}_T$ can be given intervals on $\mathbb{R}^{k_s} \times \mathbb{R}^+$.

In the computer simulations, we usually fix the spatial and temporal domain and sample the input domain. This give us computational flexibility because we can represent the domain as a tensor product of the input, spatial, and temporal domain. For an input parameter $\boldsymbol{\xi} \in \mathcal{X}_\xi$, the computer simulation returns the (multi-output) response on a given (*a priori* known) set of n_s spatial points $\boldsymbol{X}_s = (\boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n_s})^T \in \mathbb{R}^{n_s \times k_s}$, where $k_s = 1, 2$, or 3 is the number of spatial dimensions, at each one of the n_t timesteps $\boldsymbol{X}_t = (t_1, \ldots, t_{n_t}) \in \mathbb{R}^{n_t \times 1}$. That is, a single choice of the input domain $\boldsymbol{\xi}$ generates a total of $n_s \times n_t$ training samples. Therefore, the response is a matrix in $\mathbb{R}^{(n_s n_t) \times q}$, where q is the number of the output variables of the computer simulation.

For modeling reasons we will represent the problem as a q multivariate response $\boldsymbol{\eta}(\boldsymbol{x}_i) = \boldsymbol{\eta}(\boldsymbol{\xi}_{i_\xi}, \boldsymbol{s}_{i_s}, t_{i_t}) \in \mathbb{R}^q$ given input, spatial, and time point. Each of the $N = n_\xi n_s n_t$ points $(\boldsymbol{\xi}_{i_s}, \boldsymbol{s}_{i_s}, t_{i_t})$ is represented uniquely by $\boldsymbol{x}_i$. We denote the multiple observed output vector as $\boldsymbol{Y} = (\boldsymbol{\eta}(\boldsymbol{x}_1)^T, \ldots, \boldsymbol{\eta}(\boldsymbol{x}_N)^T)^T$ and $\boldsymbol{X} = (\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_N^T)$ as its corresponding input, spatial, and time vector, where $N = n_\xi n_s n_t$ is the total sample size of the new setting. For simplification purpose, we call the input domain, spatial domain, and temporal domain input, space, and time, respectively. Throughout this paper we will collectively denote input of $\boldsymbol{\eta}(.)$ by $\boldsymbol{x} = (\boldsymbol{\xi}, \boldsymbol{s}, t)$ and the space domain by $\mathcal{X} \equiv \mathcal{X}_\xi \times \mathcal{X}_s \times [0, T]$.

GP is a tool that has been successful building surrogate model for the computer simulations. The challenge in GP is to model the mean and the variance. A linear regression model for the mean is usually a good choice for continuous fields. The covariance function is harder to model and usually depends on the computational complexity and the form of $\boldsymbol{\eta}(\cdot)$. In this paper we use a combination of linear model of coregionalizations and the separable model. The Bayesian formulation assumes a GP prior distribution for the function $\boldsymbol{\eta}(\cdot)$, conditional on various hyperparameters. This prior distribution is updated using a preliminary training sample on input domain $\mathcal{X}_{\xi} \times \mathcal{X}_s \times [0, T]$. The goal of this paper is to interpolate within the Bayesian framework a multivariate function $\boldsymbol{\eta}(.)$: $\mathbb{R}^{\xi} \times \mathbb{R}^{k_s} \times \mathbb{R}^{+} \rightarrow R^q$ based on some observations.

3. GENERALIZED LINEAR MODEL OF COREGINALIZATION

The classical multivariate Gaussian process can be written as

$$\boldsymbol{\eta}(\boldsymbol{x}) = \boldsymbol{\mu}(\boldsymbol{x}) + \boldsymbol{w}(\boldsymbol{x}) + \boldsymbol{\epsilon}(\boldsymbol{x}), \tag{1}$$

where $\boldsymbol{\mu}(\boldsymbol{x})$ is the mean which is usually modeled as a *linear regression,* $\boldsymbol{w}(\cdot)$ is the spatial correlation, $\boldsymbol{\epsilon}(\cdot)$ denotes the nugget error. The coregionalization model is based on the representation of $\boldsymbol{w}(\boldsymbol{x})$ in Eq. (1) as

$$\boldsymbol{w}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{v}(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{A}$ is a $q \times r$, with $r \leq q$, non-singular transformation matrix, which explains the association among the q variables, and $\boldsymbol{v}(\boldsymbol{x})$ is a vector of r independent zero mean, unit variance GPs with correlation functions $\rho_1(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\psi}_1), \ldots, \rho_r(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\psi}_r)$ and hyperparameters $\boldsymbol{\psi}_i$. This will lead to a valid covariance function $C_{\boldsymbol{\eta}(\boldsymbol{x}),\boldsymbol{\eta}(\boldsymbol{x}')} = \sum_{j=1}^{r} \rho_j(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\psi}_r)\boldsymbol{\Sigma}_j$. The matrix covariance of the vector $\boldsymbol{Y}$ can be expressed as

$$\boldsymbol{C} = \sum_{j=1}^{r} \boldsymbol{R}_j \otimes \boldsymbol{\Sigma}_j,$$

where $\boldsymbol{\Sigma}_j = \boldsymbol{a}_j \boldsymbol{a}_j'$, with $\boldsymbol{a}_j$ the jth column of $\boldsymbol{A}$ and $\sum_{j=1}^{r} \boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$ represents the "covariance matrix," $\boldsymbol{R}_j$ represents the spatial correlation matrix of $\boldsymbol{v}_j$. In order to obtain a rich, constructive class of multivariate spatial process models; we can set $r = q$ and assume $\boldsymbol{A}$ is full rank. In general we can represent $\boldsymbol{A}$ and $\boldsymbol{R}$ as a function of the input.

We define a multivariate input-space-time GP $\boldsymbol{Y}$ as

$$\boldsymbol{\eta}(\boldsymbol{\xi}, \boldsymbol{s}, t) = \boldsymbol{h}(\boldsymbol{\xi}, \boldsymbol{s}, t)^T \boldsymbol{B} + \boldsymbol{A}\boldsymbol{v}(\boldsymbol{\xi}, \boldsymbol{s}, t) + \boldsymbol{\epsilon}(\boldsymbol{\xi}, \boldsymbol{s}, t), \tag{3}$$

where $\boldsymbol{B}$ are the parameters of the linear model, $\boldsymbol{A}$ is a $q \times q$ weight matrix, $\boldsymbol{v}(\boldsymbol{\xi}, \boldsymbol{s}, t) = (v_1(\boldsymbol{\xi}, \boldsymbol{s}, t), \ldots, v_q(\boldsymbol{\xi}, \boldsymbol{s}, t))$ is a zero mean and unit variance GP and $\boldsymbol{\epsilon}(\xi, s, t)$ is Gaussian white noise as in the simple LMC. This joint model representation becomes computationally intractable as increasing the dimensionality because an $n_{\xi} n_s n_t q \times n_{\xi} n_s n_t q$ dense covariance matrix of $\boldsymbol{Y}$ has to be inverted. In order to deal with this issue we investigate computationally more efficient representations in this work.

3.1 Separable Model

When $r = 1$ the above model is equivalent to the separable covariance model of a multi-output Gaussian process which has been used by [9]. In this case, when $\boldsymbol{\epsilon}(\boldsymbol{\xi}, \boldsymbol{s}, t) = 0$ the covariance function can be written as

$$c(\boldsymbol{\eta}(\boldsymbol{x}), \boldsymbol{\eta}(\boldsymbol{x}')) = \rho(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\psi}_{\xi})\rho(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\psi}_s)\rho(t, t'; \boldsymbol{\psi}_t)\boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is the $q \times q$ variance matrix of $\boldsymbol{Y}(\boldsymbol{x}) = (\eta_1(\boldsymbol{x}, \ldots, \eta_q(\boldsymbol{x}))$ at any location $\boldsymbol{x}$, $\rho(., .; \boldsymbol{\psi}_{\cdot})$ is a known correlation function (e.g., the power exponential, rational quadratic, and Matérn) and $\boldsymbol{\psi}_{\xi}, \boldsymbol{\psi}_s$, and $\boldsymbol{\psi}_t$ are the parameters associated with the correlation function of the input, space, and time domain, respectively.

The covariance matrix of the vector $\boldsymbol{Y} = (\boldsymbol{\eta}(\boldsymbol{x}_1), \ldots, \boldsymbol{\eta}(\boldsymbol{x}_n))$ can be written as

$$\boldsymbol{C} = \boldsymbol{R} \otimes \boldsymbol{\Sigma} = \boldsymbol{R}_{\xi} \otimes \boldsymbol{R}_s \otimes \boldsymbol{R}_t \otimes \boldsymbol{\Sigma},$$

where $\boldsymbol{R}_{\xi}(i,j) = [\rho_{\xi}(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j; \boldsymbol{\psi}_{\xi})]$ is the correlation matrix of the input X_{ξ}, $\boldsymbol{R}_s(k,l) = [\rho_s(\boldsymbol{s}_k, \boldsymbol{s}_l; \boldsymbol{\psi}_s)]$ is the correlation matrix of the spatial domain X_s, $\boldsymbol{R}_t(h,r) = [\rho_t(t_h, t_r; \boldsymbol{\psi}_t)]$ is the correlation matrix of the time domain X_t.

The covariance matrix facilitates the computations of the likelihood which depends on the determinant and the inverse of $\boldsymbol{C}$. The determinant can be expressed as $|\boldsymbol{C}| = |\boldsymbol{R}_{\xi}|^{n_s n_t q} |\boldsymbol{R}_s|^{n_{\xi} n_t q} |\boldsymbol{R}_t|^{n_{\xi} n_s q} |\boldsymbol{\Sigma}|^{n_{\xi} n_s n_t}$ and the inverse as $\boldsymbol{C}^{-1} = \boldsymbol{R}_{\xi}^{-1} \otimes \boldsymbol{R}_s^{-1} \otimes \boldsymbol{R}_t^{-1} \otimes \boldsymbol{\Sigma}^{-1}$. The likelihood of the matrix $\boldsymbol{Y}$ shown in Eq. (4) can be written as

$$\begin{aligned} \log f(\boldsymbol{Y}) = \text{const} &- \frac{1}{2} \log(|\boldsymbol{R}_{\xi}|^{n_s n_t q} |\boldsymbol{R}_s|^{n_{\xi} n_t q} |\boldsymbol{R}_t|^{n_{\xi} n_s q} |\boldsymbol{\Sigma}|^{n_{\xi} n_s n_t}|) \\ &- \frac{1}{2} \operatorname{tr}(\boldsymbol{\Sigma}^{-1} (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{\mu}}) (\boldsymbol{R}_{\xi}^{-1} \otimes \boldsymbol{R}_s^{-1} \otimes \boldsymbol{R}_t^{-1}) (\tilde{\boldsymbol{Y}} - \tilde{\boldsymbol{\mu}})^T), \end{aligned} \tag{4}$$

where $\tilde{\boldsymbol{Y}} = (\boldsymbol{Y}(\boldsymbol{x}_1), \ldots, \boldsymbol{Y}(\boldsymbol{x}_N))$ is a $q \times N$ matrix and $\tilde{\boldsymbol{\mu}}$ the matrix of the mean of $\tilde{\boldsymbol{Y}}$.

Despite the computational convenience, the above model has some unsatisfying restrictions for real applications. First of all, it is obvious that it assumes same correlation input, space, and time function for each distinct output $\eta_j(\boldsymbol{x})$. For $j = 1, \ldots, q$, the output $\eta_j(\cdot)$ is modeled with correlation function equal to $\rho(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\psi}_{\xi}) \rho(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\psi}_s) \rho(t, t'; \boldsymbol{\psi}_t)$. In simple words, all distinct outputs have the same variation over the input, space, and time. One more weakness of the above covariance model is the symmetry. Clearly, $\operatorname{cov}(\eta_j(\boldsymbol{x}), \eta_{j'}(\boldsymbol{x}')) = \operatorname{cov}(\eta_j(\boldsymbol{x}), \eta_{j'}(\boldsymbol{x}'))$ for all $j, j', \boldsymbol{x}, \boldsymbol{x}'$. To overcome the restrictions of the separable model and the computational demand of the joint model we introduce the conditional representation of coregionalization.

3.2 Conditional Generalized LMC

A special case of the coregionalization model is the conditional representation [12, 13, 17]. The conditional representation is equivalent to the lower triangular $\boldsymbol{A}$ in the joint LMC, i.e., the Cholesky decomposition of $\boldsymbol{\Sigma}$. The conditional model is written as

$$\begin{aligned} \eta_1(\boldsymbol{\xi}, \boldsymbol{s}, t) | \boldsymbol{\theta}_{c_1} &= \tilde{\boldsymbol{h}}_1(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t})^T \boldsymbol{\beta}_1 + \sigma_1 v_1(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t}) + \tau_1 u_1(\boldsymbol{\xi}, \boldsymbol{s}, t), \\ &\vdots \\ \eta_q(\boldsymbol{\xi}, \boldsymbol{s}, t) | \eta_1(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t}), \ldots, \eta_{q-1}(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t}), \boldsymbol{\theta}_{c_q} &= \tilde{\boldsymbol{h}}_q(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t})^T \boldsymbol{\beta}_q + \alpha^{q|1} \eta_1(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t}) + \\ &\quad \cdots + \alpha^{q|q-1} \eta_{q-1}(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t}) + \sigma_q v_q(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t}) + \tau_q u_q(\boldsymbol{\xi}, \boldsymbol{s}, t), \end{aligned} \tag{5}$$

where $\boldsymbol{\theta}_c = (\boldsymbol{\theta}_{c_1}, ..., \boldsymbol{\theta}_{c_q})$ the parameters of the conditional representation of the corregionalization model, $\tilde{\boldsymbol{h}}_j(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t})$ are the basis functions of the *linear regression model* of the input in $\eta_j(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t})$, σ_j^2 is the model variance, τ_q^2 is the nugget variance. To facilitate the representation we denote $\boldsymbol{h}_j(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t})^T = [\tilde{\boldsymbol{h}}_j(\boldsymbol{\xi}, \boldsymbol{s}, t), \eta_1(\boldsymbol{\xi}, \boldsymbol{s}, t), \ldots, \eta_{j-1}(\boldsymbol{\xi}, \boldsymbol{s}, t)]$, for $j = 1, \ldots, q$. We also denote $\boldsymbol{H}_j^T = (\boldsymbol{h}_j(\boldsymbol{\xi}_1, \boldsymbol{s}_1, t_1), \ldots, \boldsymbol{h}_j(\boldsymbol{\xi}_{n_{\xi}}, \boldsymbol{s}_{n_s}, t_{n_t}))$ as the basis matrix, $\boldsymbol{B}_j = (\boldsymbol{\beta}_j, \boldsymbol{\alpha}_j)$ as the linear parameter associated with the $\boldsymbol{h}_j$ basis functions, and m_j as the total number of basis functions. The basis function in each of the conditional GP introduces the dependency between the multivariate output data. In order to enable the conditional and marginal specifications to agree, we will require a common covariate vector or matrix $\tilde{\boldsymbol{h}}_1(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t})^T$ and that $u_1(\boldsymbol{\xi}, \boldsymbol{s}, t) = \cdots = u_{q-1}(\boldsymbol{\xi}, \boldsymbol{s}, t) = 0$.

Despite this simplification, the computational intractability remains since we still have to invert q different high-dimensional correlation matrices of dimension $n_{\xi} n_s n_t \times n_{\xi} n_s n_t$. An important simplification can be achieved if the spatial and the temporal points at which the output is observed remain fixed independent of the input $\boldsymbol{\xi}$ and if we assume that the correlation function is separable. The correlation of each latent variable $v_j(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t})$ in Eq. (5) is modeled as

$$\begin{aligned} \operatorname{corr}(v_j(\boldsymbol{\xi}, \boldsymbol{s}, \boldsymbol{t}), v_j(\boldsymbol{\xi}', \boldsymbol{s}', \boldsymbol{t}')) &= \rho_j(\boldsymbol{x}, \boldsymbol{x}'; \boldsymbol{\psi}) \\ &= \rho_{\xi,j}(\boldsymbol{\xi}, \boldsymbol{\xi}'; \boldsymbol{\psi}_{\xi,j}) \rho_{s,j}(\boldsymbol{s}, \boldsymbol{s}'; \boldsymbol{\psi}_{s,j}) \rho_{t,j}(t, t'; \boldsymbol{\psi}_{t,j}), \end{aligned} \tag{6}$$

where $\rho_{\xi,j}(\boldsymbol{\xi},\boldsymbol{\xi}';\boldsymbol{\psi}_{\xi,j})$, $\rho_{s,j}(\boldsymbol{s},\boldsymbol{s}';\boldsymbol{\psi}_{s,j})$ and $\rho_{t,j}(t,t';\boldsymbol{\psi}_{t,j})$ are valid correlation functions (e.g., the power exponential, rational quadratic, and Matérn) of the jth conditional representation in the input, space, and temporal domain, respectively. Each conditional representation has a separable model for the covariance function. For computational efficiency, we assume $\tau_j = 0$ for $j = 1,\ldots,q$ and introduce the nugget in the correlation function. The covariance matrix can be written as a Kronecker product of smaller covariance matrices for each of the q conditional representations in Eq. (5).

Let $\boldsymbol{Y}^j = (\eta_j(\boldsymbol{\xi}_1,\boldsymbol{s}_1,\boldsymbol{t}_1),\ldots,\eta_j(\boldsymbol{\xi}_{n_\xi},\boldsymbol{s}_{n_s},t_{n_t}))^T$ denote the response vector of the jth conditional representation in Eq. 5, for $j = (1,\ldots,q)$. The correlation matrix of the vector $\boldsymbol{Y}^j$ can be written as

$$\boldsymbol{R}_j = \boldsymbol{R}_{\xi,j} \otimes \boldsymbol{R}_{s,j} \otimes \boldsymbol{R}_{t,j},$$

where $\boldsymbol{R}_{\xi,j} \in \mathbb{R}^{n_\xi \times n_\xi}$ is the correlation matrix generated by $\boldsymbol{X}_\xi$ and $\rho_{\xi,j}(\cdot,\cdot;\boldsymbol{\psi}_{\xi,j})$ as $\boldsymbol{R}_{\xi,j}(k,l) = \rho_{\xi,j}(\boldsymbol{\xi}_k,\boldsymbol{\xi}_l;\boldsymbol{\psi}_{\xi,j})$, $\boldsymbol{R}_{s,j} \in \mathbb{R}^{n_s \times n_s}$ is the correlation matrix generated by $\boldsymbol{X}_s$ and $\rho_{s,j}(\cdot,\cdot;\boldsymbol{\psi}_s)$ as $\boldsymbol{R}_{s,j}(k,l) = \rho_{s,j}(\boldsymbol{s}_k,\boldsymbol{s}_l;\boldsymbol{\psi}_s)$, $\boldsymbol{R}_t \in \mathbb{R}^{n_t \times n_t}$ is the correlation matrix generated by $\boldsymbol{X}_t$ and $\rho_t(t_h,t_r;\boldsymbol{\psi}_t)$ as $\boldsymbol{R}_t(h,r) = \rho_t(t_h,t_r;\boldsymbol{\psi}_t)$, and $\otimes$ corresponds to the Kronecker product.

The above representation of the covariance matrix facilitates the computations of each conditional likelihood, which depend on the determinant and the inverse of $\boldsymbol{R}_j$. The determinant can be expressed as $|\boldsymbol{R}_j| = |\boldsymbol{R}_{\xi,j}|^{n_s n_t}|\boldsymbol{R}_{s,j}|^{n_\xi n_t}|\boldsymbol{R}_{t,j}|^{n_\xi n_s}|$ and the inverse as $\boldsymbol{R}^{-1} = \boldsymbol{R}_{\xi,j}^{-1} \otimes \boldsymbol{R}_{s,j}^{-1} \otimes \boldsymbol{R}_{t,j}^{-1}$. Each likelihood of the conditional representation $f(\boldsymbol{Y}^j|\boldsymbol{\theta}_{c,j})$ has a separable covariance function for the input, spatial, and time domain

$$\begin{aligned}\log f(\boldsymbol{Y}^j|.) = \text{const} &- \frac{1}{2}\log(|\boldsymbol{R}_{\xi,j}|^{n_s n_t q}|\boldsymbol{R}_{s,j}|^{n_\xi n_t q}|\boldsymbol{R}_{t,j}|^{n_\xi n_s q}|\boldsymbol{\sigma}_j|^{n_\xi n_s n_t}|) \\ &- \frac{1}{2}\operatorname{tr}(\boldsymbol{\sigma}_j^{-1}(\boldsymbol{Y}^j - \boldsymbol{\mu})^T(\boldsymbol{R}_{\xi,j}^{-1} \otimes \boldsymbol{R}_{s,j}^{-1} \otimes \boldsymbol{R}_{t,j}^{-1})(\boldsymbol{Y}^j - \boldsymbol{\mu})),\end{aligned} \tag{7}$$

where $\boldsymbol{\mu}$ is the matrix of the mean of $\boldsymbol{Y}^j$ as shown in Eq. (5). The likelihood of $\boldsymbol{Y}$ is

$$f(\boldsymbol{Y};\boldsymbol{\theta}_c) = f(\boldsymbol{Y}^1;\boldsymbol{\theta}_{c,1})\ldots f(\boldsymbol{Y}^q|\boldsymbol{Y}^1,\ldots,\boldsymbol{Y}^{q-1};\boldsymbol{\theta}_{c,q}).$$

With these simplifications, we can carry out inference in the likelihood avoiding the use of the full covariance matrix and facilitating the computations significantly. The above form with the right priors can be applied to parallel computing for the separate conditional representations.

To generalize our approach, we can combine two different domains (e.g., spatial and temporal domains) and use a non-separable covariance function which in sequence can be combined with the third domain in a separable model. We can also combine different conditional representations in a separable way if we have the prior information of the same dependence on the input, spatial, or time domains. This is the same as assuming $r \leq q$ in Eq. (2) and can simplify the model. Despite the interesting features of these models they are out of the scope of the present paper.

If we include a nugget error u_j in the model we have to invert an $n_\xi n_s n_t \times n_\xi n_s n_t$ covariance matrix in each of the conditional representations which we want to avoid due to its high dimensionality. However, we need to include small quantities of an error term in each of the above correlation matrices due to the ill-conditioned matrix which are common in practice. Below we give a solution to avoid this computational complexity similar to [18], but our approach is extended to the multivariate case.

3.3 Choosing the Correlation Function

The models described in previous subsections require the specification of three covariance functions. Because the computational simplification of the Kronecker product requires the nugget error to be zero, in practice we may experience computation instabilities. A remedy for the ill-conditioned matrix is to assume in every separate correlation matrix a positive quantity in the diagonals. Depending on the problem and the correlation matrix, the diagonal positive quantity can vary usually from 10^{-10} to 10^{-1}. Another, more automatic and sophisticated way of establishing computational stabilities is to assume a nugget random parameter for each correlation function that has to be estimated. The chosen form of the three correlation functions for each conditional representation of $\boldsymbol{v}_j(\cdot)$ is

$$\rho_{\xi,j}(\boldsymbol{\xi},\boldsymbol{\xi}';\boldsymbol{\psi}_{\xi,j}) = \tilde{\rho}_{\xi,j}(\boldsymbol{\xi},\boldsymbol{\xi}';\boldsymbol{\lambda}_{\xi,j}) + g_{\xi,j}^2\delta_{\xi,\xi'},$$

$$\rho_{s,j}(\boldsymbol{s},\boldsymbol{s}';\boldsymbol{\psi}_{s,j}) = \tilde{\rho}_{s,j}(\boldsymbol{s},\boldsymbol{s}';\boldsymbol{\lambda}_{s,j}) + g_{s,j}^2\delta_{s,s'},$$
$$\rho_{t,j}(t,t';\boldsymbol{\psi}_{t,j}) = \tilde{\rho}_{t,j}(t,t';\boldsymbol{\lambda}_{t,j}) + g_{t,j}^2\delta_{t,t'},$$

where $\boldsymbol{\psi}_{\xi,j} = (\boldsymbol{\lambda}_{\xi,j}, g_{\xi,j})$, $\boldsymbol{\psi}_{s,j} = (\boldsymbol{\lambda}_{s,j}, g_{s,j})$, $\boldsymbol{\psi}_{t,j} = (\boldsymbol{\lambda}_{t,j}, g_{t,j})$, $\boldsymbol{\lambda}_{\xi,j} = (\lambda_{\xi,j,1},\ldots,\lambda_{\xi,j,k_\xi})$, and $\boldsymbol{\lambda}_{s,j} = (\lambda_{s,j,1},\ldots,\lambda_{s,j,k_s})$ represent the correlation strength vector of $\boldsymbol{v}_j(\cdot)$ in the input and spatial domain, respectively, and $\lambda_{\xi,j,k}, \lambda_{s,j,k}$ represents the correlation strength of $\boldsymbol{v}_j(\cdot)$ in the kth direction in the input and spatial domain, respectively, and $\lambda_{t,j}$ represent the correlation strength of $\boldsymbol{v}_j(\cdot)$ in time. The $g_{\xi,j}$, $g_{s,j}$, and $g_{t,j}$ are the nugget quantities used for the stability of the input, space and time correlation matrix. We use "nugget" quantities in the correlation function and not in the covariance function to avoid the computational cost. We observe better numerical results when using the nugget as random than fixing it to an arbitrary quantity.

The choice of each correlation function $\tilde{\rho}_{\xi,j}, \tilde{\rho}_{s,j}$, and $\tilde{\rho}_{t,j}$, can be chosen independent from each other and depend solely on the information of the scientist or the computational simplification associated with it. The correlation functions can be chosen from any well known correlation family such as exponential, square exponential, and Matérn. In this paper we choose to work with the power exponential family because of its popularity and simplicity:

$$\tilde{\rho}_{\xi,j}(\boldsymbol{\xi},\boldsymbol{\xi}';\boldsymbol{\lambda}_{\xi,j}) = \exp\left\{-\frac{1}{2}\sum_{k=1}^{k_\xi}\frac{||\xi_k-\xi_k'||^{\zeta_\xi}}{\lambda_{\xi,j,k}^{\zeta_\xi}}\right\},$$
$$\tilde{\rho}_{s,j}(\boldsymbol{s},\boldsymbol{s}';\boldsymbol{\lambda}_{s,j}) = \exp\left\{-\frac{1}{2}\sum_{k=1}^{k_s}\frac{||s_k-s_k'||^{\zeta_s}}{\lambda_{s,j,k}^{\zeta_s}}\right\},$$
$$\tilde{\rho}_{t,j}(t,t';\lambda_{t,j}) = \exp\left\{-\frac{1}{2}\frac{||t-t'||^{\zeta_t}}{\lambda_{t,j}^{\zeta_t}}\right\},$$

where ζ_ξ, ζ_s, and ζ_t are values in the interval $(0,2]$. Each conditional representation has a separable model for the covariance function. For computational efficiency, we assume $\tau_j = 0$ for $j = 1,\ldots,q$ and introduce the nugget in the correlation function. The covariance matrix can be written as a Kronecker product of smaller covariance matrices for each of the q conditional representations in Eq. (5). Furthermore, when for a particular domain the data are sampled in a grid, the correlation matrix of that domain can be also expressed as a Kronecker product of one-dimensional matrices.

4. BAYESIAN INFERENCE

Prior distributions. Let $\theta_c = (\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{g})$ denote the parameters of the conditional representation in Eq. (6). In order to carry out computations in a Bayesian way we assign prior distributions into the parameters and hyper-parameters which represent prior knowledge about the computer simulator that might be available for non-informative prior distribution. In general we may choose priors which may facilitate the evaluation of the posterior distribution of these parameters. The set of parameters to be estimated in the above model is $\theta_c = \{\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\sigma}^2,\boldsymbol{\lambda},\boldsymbol{g}\}$, where $\boldsymbol{\sigma}^2 = (\sigma_1^2,\ldots,\sigma_q^2)$ has q parameters, $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_{\xi,1},\ldots,\boldsymbol{\lambda}_{\xi,q},\boldsymbol{\lambda}_{s,1},\ldots,\boldsymbol{\lambda}_{s,q},\boldsymbol{\lambda}_{t,1},\ldots,\boldsymbol{\lambda}_{t,q})$ has $q\times(k_\xi+k_s+1)$ parameters, $\boldsymbol{g} = (\boldsymbol{g}_\xi,\boldsymbol{g}_s,\boldsymbol{g}_t) = (\boldsymbol{g}_{\xi,1},\ldots,\boldsymbol{g}_{t,q})$ has $3q$ parameters, $\boldsymbol{\beta}^T = (\beta_1,\ldots,\beta_q)$ has qM parameters, and $\boldsymbol{\alpha}^T = (\alpha^{2|1},\ldots,\alpha^{q|(q-1)})$ has $(q-1)q/2$ parameters.

In order to fit q models separately in the Bayesian inference we use a prior which has an independent parameter for each conditional model of Eq. (5), $\pi(\theta_c) = \prod_{j=1}^q \pi(\theta_{c_j}) = \prod_{j=1}^q \pi(\boldsymbol{\beta}_j,\boldsymbol{\alpha}_j,\sigma_j^2,\boldsymbol{\lambda}_j,\boldsymbol{g}_j)$. This will facilitate our Bayesian inference. The problem can be seen as q different GP regressions each of which can be fitted separately.

Standard method can be used to determine the prior distributions of the parameters in the model. For simplicity, conjugate prior distributions can be used for the parameters associated with the mean $\boldsymbol{B}_j = (\boldsymbol{\beta}_j,\boldsymbol{\alpha}_j)$ and variance $\boldsymbol{\sigma}_j^2$. We suggest non-informative priors for correlation hyper-parameters since the choice of conjugate priors is not possible in practice. We can assign $\pi(\boldsymbol{B}_j,\sigma_j^2)$ to be a product of inverse gamma with a normal distribution $\pi(\boldsymbol{B}_j,\sigma_j^2) \equiv IG(\sigma_j;r,\omega)\times \mathrm{N}(\boldsymbol{B}_j;0,\tilde{\boldsymbol{\omega}})$. When there is no prior information about $\boldsymbol{B}_j$ and $\boldsymbol{\sigma}_j$, we can consider non-informative prior $p(\boldsymbol{B}_j,\sigma_j^2)\propto\sigma_j^{-2}$ which will lead to a closed form of the marginal posterior distribution of $\boldsymbol{\lambda}_j$ [6].

This is the equivalent of using inverse Wishart with diagonal parameter matrix for $\boldsymbol{\Sigma}$, [17], in the joint representation of the coregionalization model. In order to ensure positive support on the values of $\boldsymbol{\lambda}_{\xi,j}, \boldsymbol{g}_{\xi,j}, \boldsymbol{\lambda}_{s,j}, \boldsymbol{g}_{s,j}, \boldsymbol{\lambda}_{t,j}$, and $\boldsymbol{g}_{t,j}$ we assign exponential prior distributions with parameters depending on the problem. The posterior distribution of the above parameters can be derived using methods which are similar in computational cost to the separable model.

Posterior distributions. We use the special form of the above conjugate prior for $\boldsymbol{B}_j, \sigma_j^2$ which is $p(\boldsymbol{B}_j, \sigma_j^2) \propto \sigma_j^{-2}$, which when combined with the likelihood of the jth conditional GP, leads to further computational simplifications.

The joint posterior distribution of $\boldsymbol{B}$ and $\boldsymbol{\sigma}$ is $p(\boldsymbol{B}, \boldsymbol{\sigma}^2|\cdot) = \prod_{j=1:q} p(\boldsymbol{B}_j|\boldsymbol{\sigma}_j^2, \boldsymbol{Y}, \boldsymbol{\lambda}_j, \boldsymbol{g}_j) p(\boldsymbol{\sigma}_j^2|\boldsymbol{Y}, \boldsymbol{\lambda}_j, \boldsymbol{g}_j)$, where

$$p(\boldsymbol{B}_j|\boldsymbol{Y}, \sigma_j^2, \boldsymbol{\lambda}_j, \boldsymbol{g}_j) \equiv \mathrm{N}(\boldsymbol{H}_j \hat{\boldsymbol{B}}_j, \sigma_j^2 (\boldsymbol{H}_j^T \boldsymbol{R}_j^{-1} \boldsymbol{H}_j)), \tag{8}$$

where

$$\hat{\boldsymbol{B}}_j = (\boldsymbol{H}_j^T \boldsymbol{R}_j^{-1} \boldsymbol{H}_j)^{-1} \boldsymbol{H}_j^T \boldsymbol{R}_j^{-1} \boldsymbol{Y}^j,$$

and

$$p(\sigma_j^2|\boldsymbol{Y}, \boldsymbol{\lambda}_j, \boldsymbol{g}_j) \equiv \mathrm{InvGam}\big[\frac{N-1}{2}, \frac{(N-m_j-2)\hat{\sigma}_j^2}{2}\big], \tag{9}$$

where

$$\hat{\sigma}_j^2 = \frac{\boldsymbol{Y}^{jT} (\boldsymbol{R}_j^{-1} - \boldsymbol{R}_j^{-1} \boldsymbol{H}_j (\boldsymbol{H}_j^T \boldsymbol{R}_j^{-1} \boldsymbol{H}_j)^{-1} \boldsymbol{H}_j^T \boldsymbol{R}_j^{-1}) \boldsymbol{Y}^j}{N - m_j - 2}.$$

Integrating out $\boldsymbol{B}_j$ and σ_j^2 from the posterior of $\boldsymbol{\lambda}_j, \boldsymbol{g}_j, \sigma_j^2, \boldsymbol{B}_j|\boldsymbol{Y}$, it can be shown that:

$$p(\boldsymbol{\lambda}_j|\boldsymbol{Y}, \boldsymbol{g}_j) \propto \pi(\boldsymbol{\lambda}_j)\pi(\boldsymbol{g}_j)|\boldsymbol{R}_j|^{-1/2}|\boldsymbol{H}_j^T \boldsymbol{R}_j^{-1} \boldsymbol{H}_j|^{-1/2} (\hat{\sigma}_j^2)^{(N-m_j)/2}, \tag{10}$$

and

$$p(\boldsymbol{g}_j|\boldsymbol{Y}, \boldsymbol{\lambda}_j) \propto \pi(\boldsymbol{\lambda}_j)\pi(\boldsymbol{g}_j)|\boldsymbol{R}_j|^{-1/2}|\boldsymbol{H}_j^T \boldsymbol{R}_j^{-1} \boldsymbol{H}_j|^{-1/2} (\hat{\sigma}_j^2)^{(N-m_j)/2}. \tag{11}$$

The integration of $\boldsymbol{B}_j$ and σ_j^2, in Eqs. (10) and (11), is done using probability density function (pdf) properties. We first integrate out $\boldsymbol{B}_j$ with the help of the normal pdf and then integrate out σ_j using an inverse-Gamma pdf. Both the posterior distributions (10) and (11) are intractable and the inference is carried out with MCMC computations techniques. Using $\boldsymbol{R}_j = \boldsymbol{R}_{\xi,j} \otimes \boldsymbol{R}_{s,j} \otimes \boldsymbol{R}_{t,j}$ facilitates the posterior distribution in the above representation since we apply the Kronecker product techniques to compute the determinant and inverse the matrix. Yet, integrating over $\boldsymbol{B}_j$ and $\boldsymbol{\sigma}_j^2$ can improve the mixing of the MCMC [19, 20]. This is crucial since the MCMC we applied in our problem is a combination of Metropolis-Hasting within Gibbs sampling [21, 22], which requires many of iterations. The computational cost for the conditional model is q times more expensive than the separable model. However, the model used here is more general and as we will show, in the numerical example section, it gives better results.

4.1 Predictive Distribution

In this section, we calculate the predictive distribution of $\boldsymbol{\eta}(\boldsymbol{x}') \in \mathbb{R}^q$ at a new point $\boldsymbol{x}' \in \mathbb{R}^k$. The main problem is usually to obtain an analytical representation of the response surface for arbitrary input, spatial, or time values. The predictive distribution is used to predict the response surface and the error associated with it.

Given the data and all the parameters of the GP, the distribution of $\boldsymbol{\eta}(\cdot)$ is

$$p(\eta_j(\boldsymbol{x}')|\boldsymbol{B}_j, \boldsymbol{\sigma}_j, \boldsymbol{\lambda}_j, \boldsymbol{g}_j, \boldsymbol{Y}) \equiv \mathcal{N}\left(\boldsymbol{m}_j^*(\boldsymbol{x}'), c_j^*(\boldsymbol{x}', \boldsymbol{x}'; \boldsymbol{\lambda}_j, \boldsymbol{g}_j)\right), \tag{12}$$

where

$$\boldsymbol{m}_j^*(\boldsymbol{x}') = E(\eta_j(\boldsymbol{x}')|\boldsymbol{Y}, \boldsymbol{B}_j, \boldsymbol{\sigma}_j, \boldsymbol{\lambda}_j, \boldsymbol{g}_j) = \boldsymbol{\mu}_j(\boldsymbol{x}') + \boldsymbol{r}_j(\boldsymbol{X}, \boldsymbol{x}')^T \boldsymbol{R}_j^{-1} (\tilde{\boldsymbol{Y}}^j - \mu(\boldsymbol{Y}^j),$$
$$c_j^*(\boldsymbol{x}', \boldsymbol{x}'; \boldsymbol{\lambda}_j, \boldsymbol{g}_j, \boldsymbol{\sigma}_j) = \mathrm{Var}(\boldsymbol{\eta}(\boldsymbol{x}')|\boldsymbol{Y}^j, \boldsymbol{\lambda}_j, \boldsymbol{g}_j, \boldsymbol{\sigma}_j) = (\boldsymbol{r}_j(\boldsymbol{x}', \boldsymbol{x}') - \boldsymbol{r}_j(\boldsymbol{X}, \boldsymbol{x}')^T \boldsymbol{R}_j^{-1} \boldsymbol{r}_j(\boldsymbol{X}, \boldsymbol{x}')) \boldsymbol{\sigma}_j^2.$$

The above representation can be further simplified by the use of the conditional distribution of $\eta_j(\cdot)|\boldsymbol{Y}^j, \boldsymbol{\lambda}_j, \boldsymbol{g}_j$. Given that we have chosen the priors specification of $\pi(\boldsymbol{B}_j, \boldsymbol{\sigma}_j) \propto |\boldsymbol{\sigma}_j|^{-(q+1)/2}$ and integrating out both $\boldsymbol{B}_j$ and σ_j

the distribution of $\eta_j(\cdot)|\boldsymbol{\lambda}_j, \boldsymbol{g}_j, \boldsymbol{Y}_j$ is a multivariate $\boldsymbol{t}$-student, [6].

$$p(\eta_j(\boldsymbol{x}')|\boldsymbol{\lambda}_j, \boldsymbol{g}_j, \boldsymbol{Y}, \eta_1(\boldsymbol{x}'), \ldots, \eta_{j-1}(\boldsymbol{x}')) \equiv \mathcal{T}_1\left(\boldsymbol{m}_j^{**}(\boldsymbol{x}'), \boldsymbol{r}_j^{**}(\boldsymbol{x}', \boldsymbol{x}'; \boldsymbol{\lambda})\hat{\sigma}_j^2; N - m_j\right), \tag{13}$$

with $N - m_j$ degrees of freedom and

$$\begin{aligned}
\boldsymbol{m}_j^{**}(\boldsymbol{x}') &= \boldsymbol{h}_j(\boldsymbol{x}')^T\hat{\boldsymbol{B}}_j + \boldsymbol{r}(\boldsymbol{X}, \boldsymbol{x}')^T\boldsymbol{R}_j^{-1}(\tilde{\boldsymbol{Y}}^j - \boldsymbol{H}_j\hat{\boldsymbol{B}}_j),\\
r_j^{**}(\boldsymbol{x}', \boldsymbol{x}'; \boldsymbol{\lambda}_j, \boldsymbol{g}_j) &= (\boldsymbol{r}_j(\boldsymbol{x}', \boldsymbol{x}') - \boldsymbol{r}_j(\boldsymbol{X}, \boldsymbol{x}')^T\boldsymbol{R}_j^{-1}\boldsymbol{r}_j(\boldsymbol{X}, \boldsymbol{x}')),\\
\hat{\boldsymbol{B}}_j &= (\boldsymbol{H}_j^T\boldsymbol{R}_j^{-1}\boldsymbol{H}_j)^{-1}\boldsymbol{H}_j^T\boldsymbol{R}_j^{-1}\boldsymbol{Y}^j,\\
\hat{\sigma}_j^2 &= \frac{1}{N - m_j}(\boldsymbol{Y}^j - \boldsymbol{H}_j\hat{\boldsymbol{B}}_j)\boldsymbol{R}_j^{-1}(\boldsymbol{Y}^j - \boldsymbol{H}_j\hat{\boldsymbol{B}}_j).
\end{aligned}$$

In practice we do not observe directly $\eta_1(\boldsymbol{x}'), \ldots, \eta_{j-1}(\boldsymbol{x}')$ and the covariance parameters $(\boldsymbol{\lambda}_j, \boldsymbol{g}_j)$. However, they can be sampled iteratively from MCMC algorithm.

The Bayesian predictive density function $\eta(\cdot)|\boldsymbol{Y}$ is calculated as

$$p(\eta_1(\boldsymbol{x}'), ..., \eta_q(\boldsymbol{x}')|\boldsymbol{Y}) = \int_{\boldsymbol{\lambda}, \boldsymbol{g}} p(\eta_q(\boldsymbol{x})|\boldsymbol{\lambda}_q, \boldsymbol{g}_q, \boldsymbol{Y}^q, \eta_1(\boldsymbol{x}'), ..., \eta_{q-1}(\boldsymbol{x}'))...p(\eta_1(\boldsymbol{x}')|\boldsymbol{Y}^1)\pi(\boldsymbol{\lambda}, \boldsymbol{g}|\boldsymbol{Y})d\boldsymbol{\lambda}d\boldsymbol{g}. \tag{14}$$

Formally speaking, the posterior distribution of $\boldsymbol{\eta}(\boldsymbol{x})|\boldsymbol{Y}$ should be regarded as the emulator. Integration of $\boldsymbol{\lambda}$ and $\boldsymbol{g}$ in Eq. (14) needs to be done numerically since it does not have a closed form. We use MCMC techniques to approximate the above integral. More specifically, the approximation of $p(\boldsymbol{\eta}(\boldsymbol{x}')|\boldsymbol{Y})$ is given by

1. For $j = 1, \ldots, q$, generate MCMC samples $(\boldsymbol{\lambda}_j^{(1)}, \boldsymbol{g}_j^{(1)}), \ldots, (\boldsymbol{\lambda}_j^{(M)}, \boldsymbol{g}_j^{(M)})$ from $p(\boldsymbol{\lambda}_j, \boldsymbol{g}_j|\boldsymbol{Y})$ as we described in Section 4.

2. Approximate $p(\boldsymbol{\eta}(\boldsymbol{x}')|\boldsymbol{Y})$ by

$$\hat{p}(\boldsymbol{\eta}(\boldsymbol{x}')|\boldsymbol{Y}) = \frac{1}{M}\sum_{k=1}^{M} p(\eta_q(\boldsymbol{x}')|\boldsymbol{\lambda}_q^{(k)}, \boldsymbol{g}_q^{(k)}, \boldsymbol{Y}^q, \eta_1(\boldsymbol{x}'), \ldots, \eta_{q-1}(\boldsymbol{x}')) \ldots p(\eta_1(\boldsymbol{x}')|\boldsymbol{\lambda}_1^{(k)}, \boldsymbol{g}_1^{(k)}, \boldsymbol{Y}^1).$$

In computer experiments, we are interested in obtaining an analytical representation of the response surface for random input and by maintaining fixed the spatial and temporal values. The input, spatial, and time values which we would like to predict are in the form of $\boldsymbol{X}_P = \boldsymbol{x}_{\xi,P} \times \boldsymbol{X}_s \times \boldsymbol{X}_t$. In this case we can write

$$p(\boldsymbol{\eta}_j(\boldsymbol{X}_P)|\boldsymbol{\lambda}_j, \boldsymbol{g}_j, \boldsymbol{Y}, \boldsymbol{\eta}_1(\boldsymbol{X}_P), \ldots, \boldsymbol{\eta}_{j-1}(\boldsymbol{X}_P)) \equiv \mathcal{T}_{n_s n_t \times 1}\left(\boldsymbol{m}_j^{**}(\boldsymbol{X}_P), \boldsymbol{r}_j^{**}(\boldsymbol{X}_P, \boldsymbol{X}_P; \boldsymbol{\lambda})\hat{\sigma}_j^2\right), \tag{15}$$

where the mean and variance of the predictions are

$$\begin{aligned}
\boldsymbol{m}_j^{**}(\boldsymbol{X}_P) = \boldsymbol{h}_q(\boldsymbol{X}_P)^T\hat{\boldsymbol{B}}_j + \hat{\alpha}^{j|1}\eta_1(\boldsymbol{X}_P) +\\
\cdots + \hat{\alpha}^{j|j-1}\eta_{j-1}(\boldsymbol{X}_P) + \boldsymbol{r}_{\xi,j}(\boldsymbol{X}_\xi, \boldsymbol{x}_{\xi,P})^T\boldsymbol{R}_{\xi,j}^{-1} \otimes \boldsymbol{I}_s \otimes \boldsymbol{I}_t(\boldsymbol{Y}^j - \mu_{\boldsymbol{Y}^j})
\end{aligned}$$

and

$$\boldsymbol{r}_j^{**}(\boldsymbol{X}_P, \boldsymbol{X}_P; \boldsymbol{\lambda}) = \left\{\boldsymbol{r}_{\xi,j}(\boldsymbol{X}_{\xi,P}, \boldsymbol{X}_{\xi,P}) - \boldsymbol{r}_{\xi,j}(\boldsymbol{X}_\xi, \boldsymbol{x}_{\xi,P})^T\boldsymbol{R}_{\xi,j}^{-1}\boldsymbol{r}_{\xi,j}(\boldsymbol{X}_\xi, \boldsymbol{x}_{\xi,P})\right\} \otimes \boldsymbol{R}_{s,j} \otimes \boldsymbol{R}_{t,j}.$$

The proposed GP model can also be employed to calculate statistical moments, i.e., mean, variance, and the pdf.

5. NUMERICAL EXAMPLES

5.1 Kraichnan-Orszag Three-Mode Problem

Consider the system of ordinary differential equations as shown in [23]

$$\frac{dy_1}{dt} = y_1 y_3,$$
$$\frac{dy_2}{dt} = -y_2 y_3,$$
$$\frac{dy_3}{dt} = -y_1^2 + y_2^2,$$

which are subject to random initial conditions at $t = 0$. The deterministic solver we use is a fourth order Runge-Kutta method as implemented in the GNU Scientific Library [24]. For the two-dimensional problem, the stochastic initial conditions are defined by

$$y_1(0) = 1,\ y_2(0) = 0.1\xi_1,\ y_3(0) = \xi_2,$$

where

$$\xi_i \sim U([-1, 1]),\ i = 1, 2.$$

The input variables $\boldsymbol{\xi}$ here represent the initial conditions. The output consist of three distinct variables ($q = 3$) that are functions of time ($k_s = 0$). For convenience, we choose to work with a constant prior mean by electing $\boldsymbol{h}_{\boldsymbol{\xi}}(\boldsymbol{\xi}) = 1$ and $\boldsymbol{h}_t(t) = 1$.

We fix the sample size, $n_{\boldsymbol{\xi}}$, and gather the input data $\boldsymbol{X}_{\boldsymbol{\xi}} \in \mathbb{R}^{n_{\boldsymbol{\xi}} \times k_{\boldsymbol{\xi}}}$ from a Latin hyper-cube design described in [25]. We solve the system for the time interval $[0, 10]$ and record the response at 10 equidistant time steps, i.e., $\boldsymbol{X}_t \in \mathbb{R}^{n_t}$ with $n_t = 10$.

We take $n_{\boldsymbol{\xi}} = 100$ and $n_{\boldsymbol{\xi}} = 300$ samples with a Latin hyper-cube design and try to evaluate the statistics of each variable with the separable model and the conditional LMC. The priors for $\lambda_{\zeta,j}, g_{\zeta,j}$ are specified by setting the parameter of the exponential prior to 0.05 and 10^{-3}, respectively. The proposals of the Metropolis-Hastings are selected to be a log-normal random walk and step which is determined from a pilot study. For each method, we ran 30,000 iterations to collect posterior samples after a burn-in period of 5,000 iterations. Good convergence of the respective marginal distributions is indicated by the trace plots of parameters. The MCMC distribution of some crucial parameter in the the conditional LMC is shown in the Appendix, Fig. A.1.

After the burn-in iterations we numerically compute the Bayesian predictive density function, presented in Section 4.1, at 120 input grid points and the existing time steps. We compute the mean square prediction error (MSPE) integrated over the whole input space and time for each variable and present them in Table 1. The MSPE is calculated as the mean square error of the true value in comparison with the mean of the Bayesian predictive density function. The LMC for the covariance function gives better results in both cases. When $n_{\xi} = 300$, the MSPE for all the outputs is close to half of the MSPE when we use the separable model. The MSPE of y_1 using LMC is almost three times smaller than the computed MSPE using the separable covariance function. Great differences are also observed in the case where only 100 input samples are taken. The assumptions associated with the separable model seem to be violated in this set of data. For comparison, we have also used Matérn correlation model as a correlation function. The results are very similar to the one appearing in Table 1, suggesting less sensitivity on selecting the correlation function for this system of ordinary differential equations.

Using the same setting we compute the response surface of the mean of Bayesian predictive density function for the two models. For $n_{\xi} = 300$ and $t = 10$ we show the predicted response surfaces in Fig. 1 at a 120×120 input grid. The differences we observed in the table can also be seen in Fig. 1. The conditional coregionalization model gives better representation of the predicted surface. The response surface of y_1 shows great differences between the LMC and the separable covariance function.

TABLE 1: MSPE for the two different cross-covariance functions and $n_{\xi} = (100, 300)$

Variable	**LMC** $n_{\xi} = 100$	**Separable** $n_{\xi} = 100$	**LMC** $n_{\xi} = 300$	**Separable** $n_{\xi} = 300$
y_1	0.0142	0.0201	0.0022	0.0069
y_2	0.0116	0.0160	0.0051	0.0094
y_3	0.0341	0.0509	0.0103	0.0184

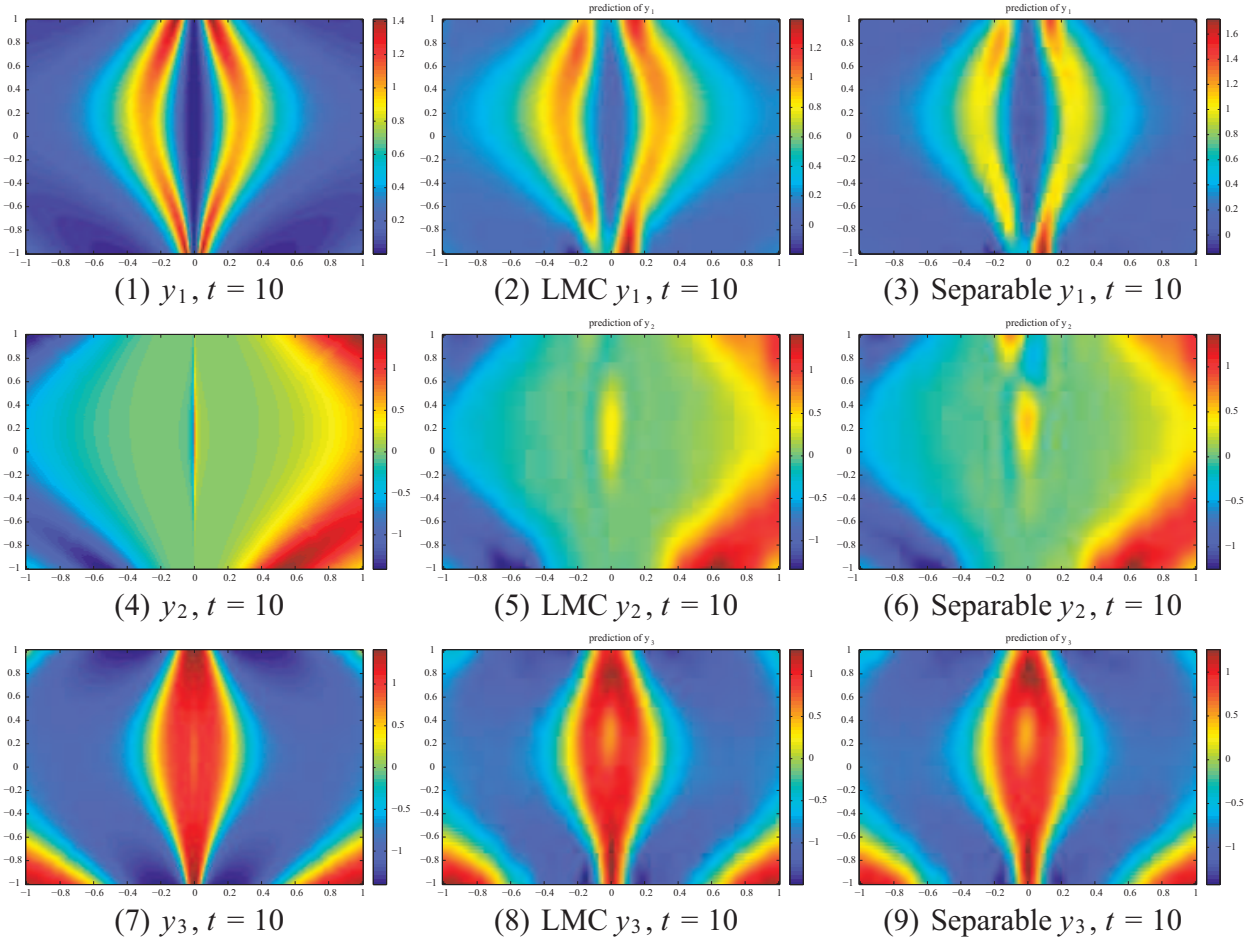

FIG. 1: Two-dimensional K-O problem for $n_\xi = 300$: The first column shows the exact realization of the three output for time $t = 10$. The second column shows the predictive Bayesian mean when we model the covariance function with the conditional linear model of coregionalization and the third column when we use the separable model.

Since we use a Bayesian hierarchical model, we can also provide the prediction distribution or the prediction intervals for the output at each input and time points. Figure 2 shows the estimated prediction densities of (y_1, y_2, y_3) for sample size $n = 100$ (green dashed line) and $n = 300$ (black dash-dot line), in input points $\boldsymbol{\xi}_1^* = (-0.3333, -0.5556)$, $\boldsymbol{\xi}_2^* = (-0.3333, 0.3333)$, $\boldsymbol{\xi}_3^* = (0.1111, 0.1111)$, $\boldsymbol{\xi}_4^* = (0.3333, -0.1111)$, and time $t^* = 10$. We use the last 25,000 MCMC samples to estimate the prediction densities. As expected, the prediction distribution becomes narrower as we increase the sample size. Moreover, the posterior mean moves closer to the real value of $\boldsymbol{y}(\cdot)$ (red star). Finally, we compare the output prediction distributions of the proposed model with the output prediction distributions of separable model. Figure 3 shows the estimated prediction densities for the previous input points $\boldsymbol{\xi}^*$ and time $t^* = 10$ using the two different models. The prediction density of the separable model is shown by blue solid lines and the prediction density of the LMC is shown by black dash-dot lines. The proposed model gives better results overall in terms of the mean and variance of the prediction distribution. As we can see from the graphs the dominance of LMC is not in all the input points. However, better results are observed in most cases when using LMC.

Despite the success of the proposed method, we observe certain aspects that require further investigation. First of all the data predictive distribution does not seem to capture the discontinuity in the data for values of $x_1 = 0$. Moreover, the Latin hyper-cube design for sampling does not seem to be appropriate since the MSPE is bigger in

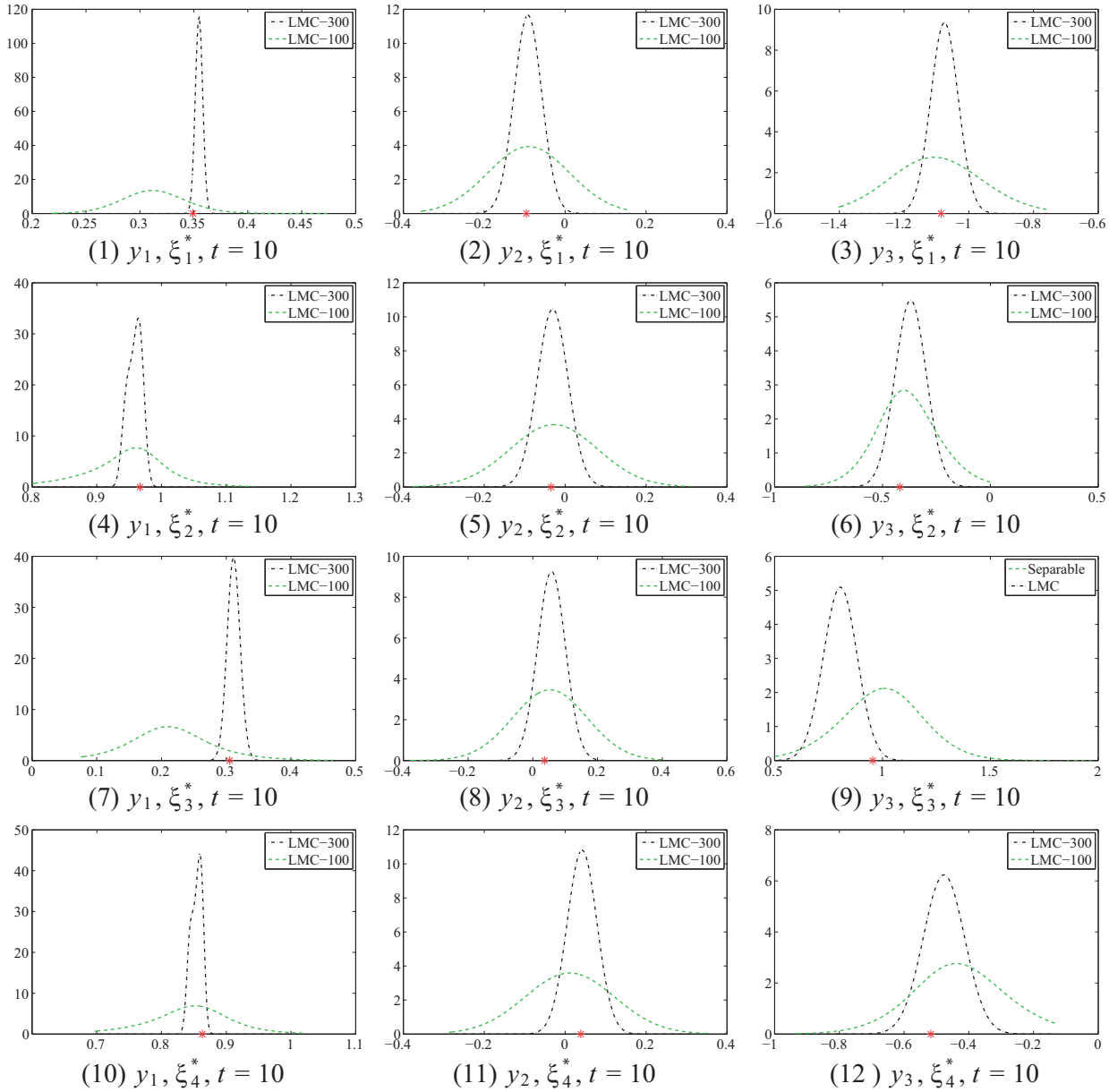

FIG. 2: Prediction distribution of the three outputs at $\boldsymbol{\xi}_1^* = (-0.3333, -0.5556), \boldsymbol{\xi}_2^* = (-0.3333, 0.3333), \boldsymbol{\xi}_3^* = (0.1111, 0.1111), \boldsymbol{\xi}_4^* = (0.3333, -0.1111)$, and $t^* = 10$ for sample size $n_\xi = 100$ (green dashed line) and $n_\xi = 300$ (black dash-dot line). Each column represents one of the three outputs and each row represents one of the four different input points. The red star denotes the true value of the output.

certain areas. Both of these issues can be solved with adaptive sequential design of experiment via active learning [8, 26–28].

The nugget quantity in the multivariate Gaussian process introduce numerical stability and usually leads to better predictions. For example if we fix the nugget effect to 0.05 when $n_\xi = 300$ the MSPE of y_1 is 0.0025, the MSPE of y_2 is 0.0057, and the MSPE of y_3 is 0.0104, which are bigger than the corresponding MSPEs when we include the random nugget effect. The same observation is true for the separable model.

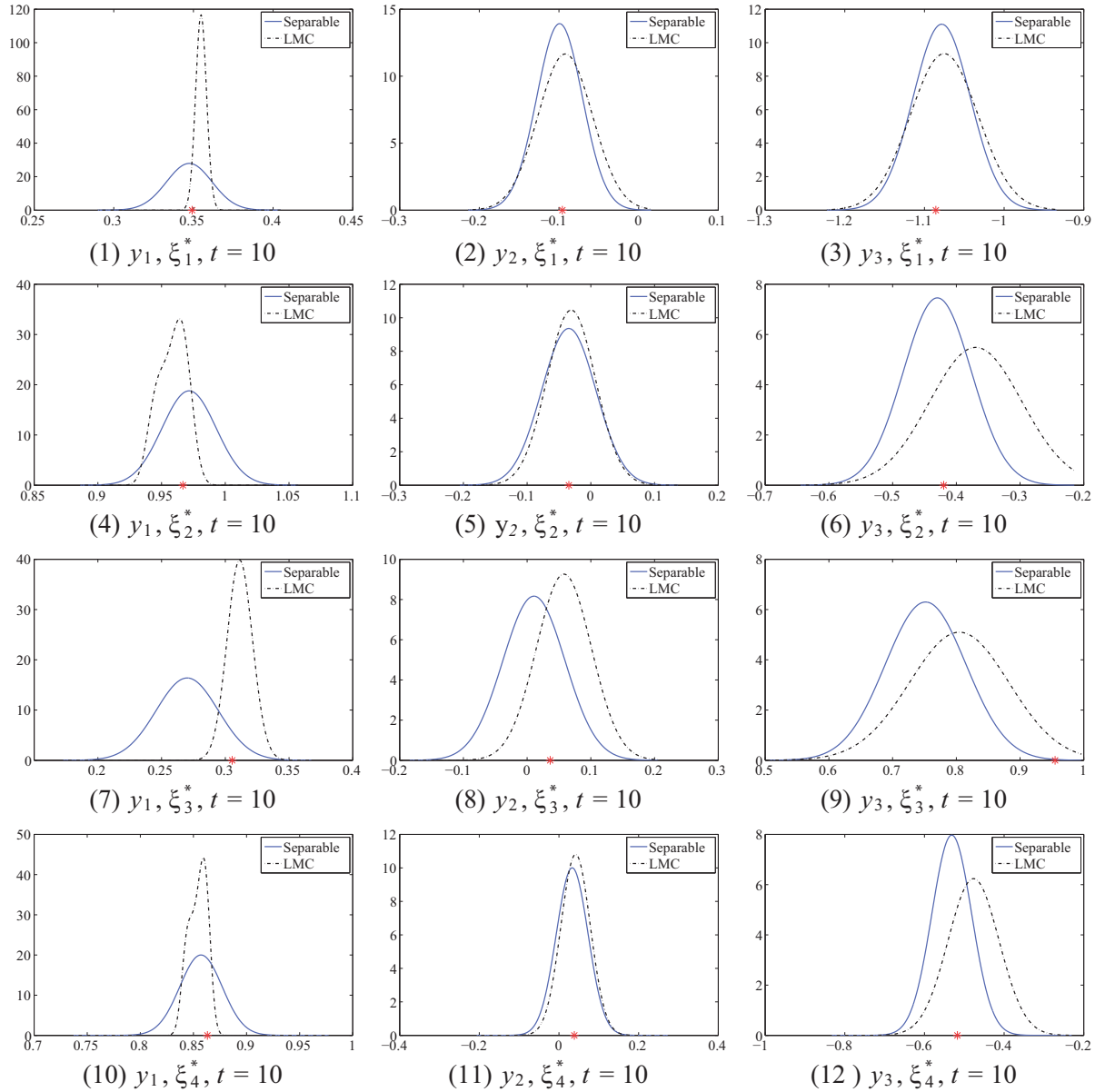

FIG. 3: Prediction distribution of the three outputs for $\boldsymbol{\xi}_1^* = (-0.3333, -0.5556), \boldsymbol{\xi}_2^* = (-0.3333, 0.3333), \boldsymbol{\xi}_3^* = (0.1111, 0.1111), \boldsymbol{\xi}_4^* = (0.3333, -0.1111)$, $t^* = 10$, and sample size $n_\xi = 300$ using separable model (blue solid line) and the proposed LMC (black dash-dot line). Each column represents one of the three outputs and each row represents one of the four different cases. The red star denotes the true value of the output.

5.2 Flow through Porous Media

We now apply our proposed methods to a two-dimensional, single-phase, steady-state flow through a random permeability field. The mathematical models through porous media are well established and can be found in numerous textbooks. A good review of this model can be found in [9, 29]. We follow these to specify the problem. The spatial domain $\mathcal{X}_s$ is chosen to be the unit square $[0, 1]^2$, representing an idealized oil reservoir. Let us denote with p the pressure and $\boldsymbol{u} = (u_x, u_y)$ the velocity fields of the fluid in the x and y spatial direction, respectively.

We set up the physical problem so that there is an injection well on the bottom left corner and a projection well on the top right corner, while we impose no-flux boundary conditions (model square wells). Let $\boldsymbol{K}$ denote the permeability diagonal tensor that models the easiness with which the liquid flows through the reservoir. We restrict ourselves to an isotropic permeability tensor $K_{ij} = K\delta_{ij}$ and $K(\boldsymbol{s}) = \exp\{G(\boldsymbol{s})\}$. The logarithm of the permeability, $G(\boldsymbol{s})$, is modeled with a Gaussian random field as

$$G(\cdot) \sim N(m_G, c_G),$$

with constant mean m_G and exponential covariance function given by

$$c_G(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma_G^2 \exp\left\{-\sum_{k=1}^{k_s} \frac{|s_{1,k} - s_{2,k}|}{l_k}\right\},$$

The parameters l_k are the correlation length of the field in the k direction, and $\sigma_G^2 > 0$ is its standard deviation. For our study we choose $m_G = 0, \sigma_G = 1$, and $l_k = 0.1$ for all k. We obtain a finite-dimensional representation of it by employing its Karhunen-Loéve representation and retaining 50 terms accounting for 95% of the field's energy

$$G(\boldsymbol{w}; \boldsymbol{s}) = m_G + \sum_{k=1}^{k_\xi} w_k \psi_k(\boldsymbol{s}),$$

where $\boldsymbol{w}$ is a vector of independent standardized Gaussian random variables and $\psi_k(\boldsymbol{s})$ are the eigenfunctions of the exponential covariance function. Using a suitable transformation, the final field can be described by 50 independent uniform random variables. The boundary value problem is solved using the mixed finite element formulation.

The above problem can be translated as a forward problem with input dimensions $k_\xi = 50$, spatial dimension $k_s = 2$, and output dimension $q = 3$. More accurately, we define the response of the physical model as

$$\mathbf{f} : \boldsymbol{X}_\xi \times \boldsymbol{X}_s \to \mathbb{R}^q,$$

where $\boldsymbol{X}_\xi = [0,1]^{50}$, $\boldsymbol{X}_s = [0,1]^2$, and

$$\mathbf{f}(\boldsymbol{x}) = \mathbf{f}(\boldsymbol{\xi}, \boldsymbol{s}) = (p(\boldsymbol{\xi}, \boldsymbol{x}_s), u_x(\boldsymbol{\xi}, \boldsymbol{s}), u_y(\boldsymbol{\xi}, \boldsymbol{s}))$$

are the solution of the boundary problem at the spatial point $\boldsymbol{s}$ for a permeability field.

The boundary value problem is solved using the mixed finite element formulation. We use first-order Raviart-Thomas elements for the velocity described in [30], and zero-order discontinuous elements for the pressure from [31]. The spatial domain is descretized using a 64×64 triangular mesh. The solver was implemented using the Dolfin C++ library from [32]. The eigenfunctions of the exponential random field used to model the permeability were calculated via Stokhos, which is part of Trilinos [33].

For each stochastic input $\boldsymbol{\xi}$, the three responses are observed on a 32×32 square spatial grid. A Latin hyper-cube design is used to select 120 observations of the solver. Three of these simulations are shown in Fig. 4. To be able to compare our GP model with the deterministic solver, 10 of these samples are left out for prediction evaluations and 110 samples are used to make Bayesian inference. We sample the posterior of the parameters of the separable and conditional model 20,000 times and make Bayesian prediction. Good convergence of the respective marginal distributions are indicated by the trace plots of the model parameters. The 5,000 iterations trace plot of the separable model parameters is given in Fig. A.2 and the 5,000 iterations trace plot of the conditional linear model of coregionalization parameters is given in Fig. A.3. We conduct predictions for the 10 left-out simulations and compare them with the real computer code results. Fig. 5 shows the real and the predicted images for the three distinct outputs in a 32×32 spatial grid for the two different covariance models. There is an obvious similarity of the computer code and the Gaussian predictions with both covariance models. The predicted values using the conditional LMC model give predictions closer to the real computer simulation. To better see this difference in terms of numbers we compute the MSPE using the 10 left out simulations. The computed MSPEs for the two different covariance models and for three distinct outputs $(u_x(\boldsymbol{\xi}, \boldsymbol{x}_s), u_y(\boldsymbol{\xi}, \boldsymbol{x}_s), p(\boldsymbol{\xi}, \boldsymbol{x}_s))$ are given in Table 2. The MSPE for the conditional LMC is significantly smaller than the MSPE using the separable model. For the pressure p, the MSPE using conditional generalized LMC is almost

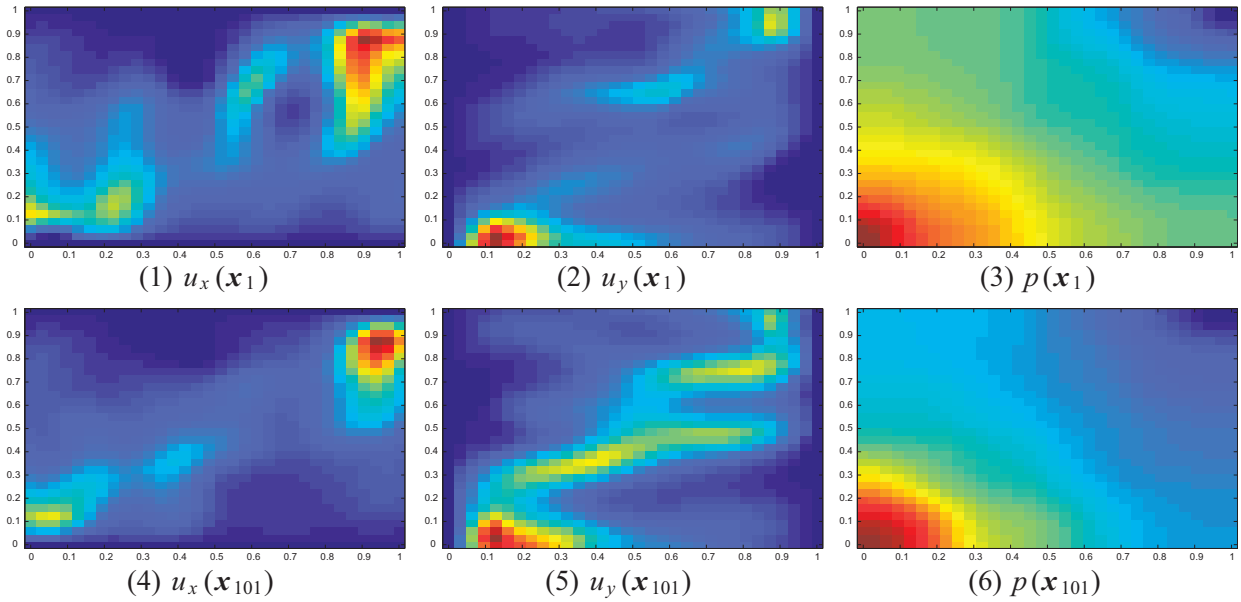

FIG. 4: Output for two different exact computer simulation realizations. Different columns represent different output and different rows represent different input.

half, comparing to the MSPE using the separable model. This is a strong indication that the separable model may not always be the best choice to model real problem using computer code simulations.

With the same predictive Bayesian method, we can compute the predictive probability density for each point in the input domain. Moreover, we can infer the distribution of the mean and variance of the flow through porous media problem.

6. CONCLUDING REMARKS AND EXTENSIONS

We developed a computationally efficient UQ tool based on a multi-dimensional GP which explicitly models the correlation between distinct outputs, input domain, spatial domain, and time. We used covariance structures which enable the posterior computations. More explicitly, we build a generalized LMC covariance for the multivariate outputs, where the dependence on different domains (input, space, and time) is considered to be separable. The conditional representation of the LMC in combination with the separable model for different domains leads to highly efficient computation in terms of of both storage memory and CPU time.

The prior specification of the LMC parameters leads to an efficient algorithm, since only the parameters of the correlation function need to be updated in each MCMC iteration. Moreover, we introduce numerical stability in the covariance function by adding a nugget term, without increasing the computational complexity of the model through maintaining the Kronecker product structure of the covariance matrix.

Finally, we applied the conditional generalized LMC to the Kraichnan-Orszag three-mode problem as well as to the flow through porous media problem and compared it with the separable model. Strong support of the use of the conditional generalized LMC is shown for both problems by comparing the mean-square prediction error. Also, the use of the random nugget effect improves the overall prediction error.

In practice, the conditional separable form we chose to work, may be too simplistic. For example, problems of discontinuity and localized features may be present in the computer simulation. For issues related with non-stationarity we suggest the use of different tree-based techniques already proposed in the literature. The Bayesian tree multivariate Gaussian process [8], the independent Bayesian tree Gaussian process (BTGP) [34], or the multi-output treed GP (MOTGP) [28] can be used to partition the space into a tree form and can also reduce the computational cost of the proposed model. To further reduce the computational cost full-scale approximation methods for the covariance matrix

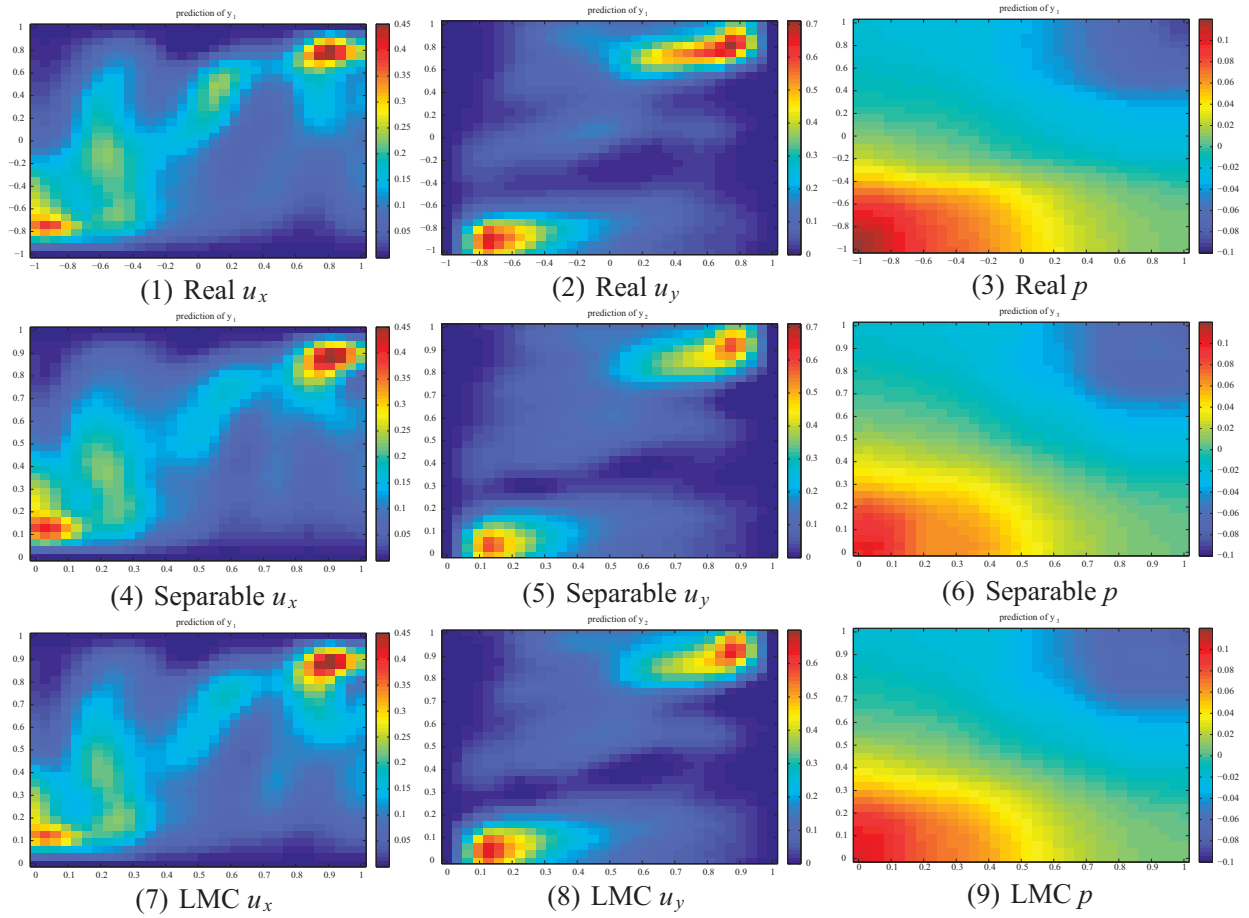

FIG. 5: Real and predicted value for 1 of the 10 training observation. First row is the real simulation, second row is the predicted mean simulation with the separable model, and the third row shows the three output predicted surfaces using LMC.

TABLE 2: MSPE for the two different cross-covariance functions

Variables	**MSPE for** $n_\xi = 110$ **Separable**	**MSPE for** $n_\xi = 110$ **LMC**
$Y_1 = u_x$	5.7250e-005	3.6252e-005
$Y_2 = u_y$	6.2589e-005	4.5979e-005
$Y_3 = p$	5.5921e-006	2.9670e-006

as they are suggested in [35] can be used. A more comprehensive study of the non-stationarity will be investigated in future research.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program as part of the Multifaceted Mathematics for Complex Energy Systems (M^2ACS) project and part of the Collaboratory on Mathematics for Mesoscopic Modeling of Materials project. This work was also partially supported by NSF Grant DMS-1115887. We thank Dr. Ilias Bilionis and Prof. Nicholas Zabaras for providing the data and the preprint of their work on the Multi-output Local Gaussian process.

REFERENCES

1. Ghanem, R. G. and Spanos, P., *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.
2. Cressie, N., *Statistics for Spatial Data. 2nd edition*, John Wiley and Sons Inc, New York, 1993.
3. O'Hagan, A., Kennedy, M. C., and Oakley, J. E., Uncertainty analysis and other inference tools for complex computer codes (with discussion), *Bayesian Stat.*, 6:503–524, 1999.
4. Kennedy, M. and O'Hagan, A., Bayesian calibration of computer models (with discussion), *J. R. Stat. Soc. (Ser. B)*, 68:425–464, 2001.
5. Oakley, J. and O'Hagan, A., Bayesian inference for the uncertainty distribution of computer model outputs, *Biometrika*, 89(4):769–784, 2002.
6. Conti, S. and O'Hagan, A., Bayesian emulation of complex multi-output and dynamic computer models, *J. Stat. Planning Inference*, 140:640–651, 2010.
7. Mardia, K. V. and Goodall, C. R., Spatial temporal analysis of multivariate environmental monitoring data, *Multiv. Env. Stat.*, 1:347–386, 1993.
8. Konomi, B., Karagiannis, G., Sarkar, A., and Lin, G., Bayesian treed multivariate gaussian process with adaptive design: Application to a carbon capture unit, *Technometrics*, 56(2):145–158, 2014.
9. Bilionis, I., Zabaras, N., Konomi, B., and Lin, G., Multi-output separable gaussian process: Towards an efficient, fully bayesian paradigm for uncertainty quantification, *J. Comput. Phys.*, 241:212–239, 2013.
10. Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P., *Handbook of Spatial Statistics*, Chapman & Hall-CRC, Boca Raton, FL, 2010.
11. Cressie, N. and Wickle, C., *Statistics for Spatial-Temporal Data*, John Wiley and Sons Inc, New York, 2011.
12. Grzebyk, M. and Wackernagel, H., Multivariate analysis and spatial/temporal scales: Real and complex models, *Proc. of XVIIth Int. Biometric Conf.*, Vol. 1, Ontario, Canada, pp. 19–33, 8–12 Aug., 1994.
13. Wackernagel, H., *Multivariate Geostatistics: An Introduction with Applications*, 2nd ed., Springer, Berlin, 2003.
14. Schmidt, A. M. and O'Hagan, A., Bayesian inference for non-stationary spatial covariance structure via spatial deformations, *J. R. Stat. Soc., Ser. B*, 65:743–758, 2003.
15. Gelfand, A. E., Schmidt, A., Banerjee, S., and Sirmans, C., Nonstationary multivariate process modeling through spatially varying coregionalization, *TEST*, 13:263–312, 2004.
16. Apanasovich, T. V. and Genton, M. G., Cross-covariance functions for multivariate random fields based on latent dimensions, *Biometrika*, 97:15–30, 2010.
17. Banerjee, S., Carlin, B., and Gelfand, A., *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall-CRC, Boca Raton, FL, 2004.
18. Gramacy, R. B. and Lee, H. K. H., Cases for the nugget in modeling computer experiments, *Stat. Comput.*, 22:713–722, 2012.
19. Liu, J., Wong, W., and Kong, A., Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes, *Biometrika*, 81(1):27–40, 1994.
20. Berger, J. O., Oliveira, V. D., and Sanso, B., Objective Bayesian analysis of spatially correlated data, *J. Am. Stat. Association*, 96:1361–1374, 2001.
21. Mueller, P., Alternatives to the Gibbs sampling scheme, Tech. Rep., Institute Statistics and Decision Sciences, Duke University, 1993.
22. Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL, 2004.
23. Wan, X. and Karniadakis, G. E., An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, *J. Comput. Phys.*, 209:617–642, 2005.
24. Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Alken, P., Booth, M., and Rossi, F., *GNU Scientific Library Reference Manual*, 2009.
25. Mckay, M. D., Beckman, R. J., and Conover, W. J., A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 42(1):202–208, 2000.

26. Seo, S., Wallat, M., Graepel, T., and Obermayer, K., Gaussian process regression: Active data selection and test point rejection, *IEEE. Proc. of the Int. Joint Conf. on Neural Networks*, Como, Italy, pp. 241–246, 2000.

27. Gramacy, R. B. and Lee, H. K. H., Adaptive design and analysis of supercomputer experiments, *Technometrics*, 51:130–145, 2009.

28. Bilionis, I. and Zabaras, N., Multi-output local Gaussian process regression: Applications to uncertainty quantification, *J. Comput. Phys.*, 231:5718–5746, 2012.

29. Aarnes, J. E., Kippe, V., Lie, K.-A., and Rustad, A. B., *Modelling of Multiscale Structures in Flow Simulations for Petroleum Reservoirs*, Springer, 2007.

30. Ghanem, R. G. and Spanos, P. D., *Stochastic Finite Elements: A Spectral Approach*, Wiley Series in Applied Mathematics, Springer-Verlag, 1991.

31. Brezzi, F., Hughes, T. J. R., Marini, L. D., and Masud, A., Mixed discontinuous Galerkin methods for Darcy flow, *J. Sci. Comput.*, 22(17):119–145, 2005.

32. Logg, A., Wells, G. N., and Hake, J., *DOLFIN: A C++/Python Finite Element Library*, Springer, Berlin, 2012.

33. Heroux, M. A., Bartlett, R. A., Howle, V. E., Hoekstra, R. J., Hu, J. J., Kolda, T. G., Lehoucq, R. B., Long, K. R., Pawlowski, R. P., Phipps, E. T., Salinger, A. G., Thornquist, H. K., Tuminaro, R. S., Willenbring, J. M., Williams, A., and Stanley, K. S., An overview of the trilinos project, *ACM Trans. Math. Softw.*, 31(3):397–423, 2005.

34. Gramacy, R. B. and Lee, H. K. H., Bayesian treed Gaussian process models with an application to computer modeling, *J. Am. Stat. Associat.*, 103:1119–1130, 2008.

35. Konomi, B., Sang, H., and Mallick, B., Adaptive bayesian nonstationary modeling for large spatial datasets using covariance approximations, *J. Comput. Graphical Stat.*, 23(3):802–829, 2014.

APPENDIX: MCMC RESULTS

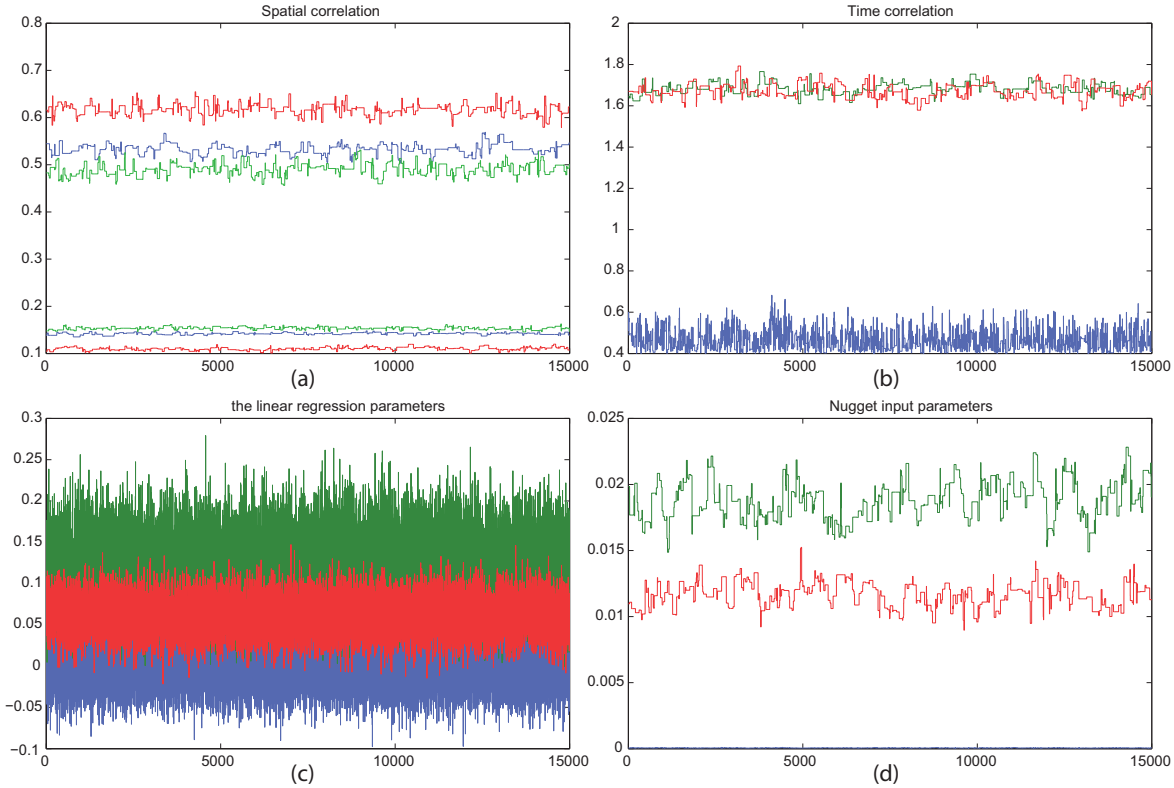

FIG. A.1: MCMC for the last 15,000 iterations.

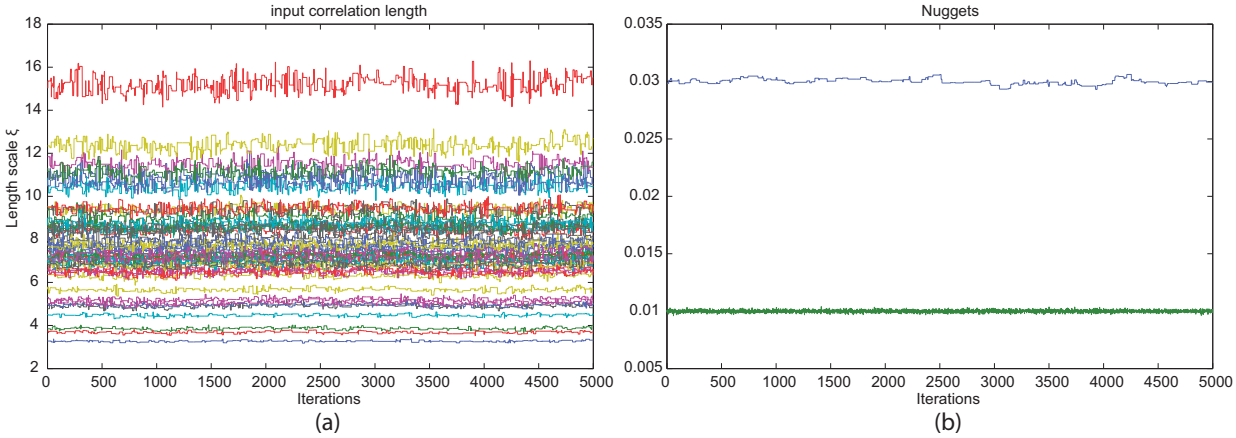

FIG. A.2: Trace plot of 5,000 MCMC iteration when we use separable model: (a) the input correlation parameters of 50 different dimentions and (b) the nugget parameters.

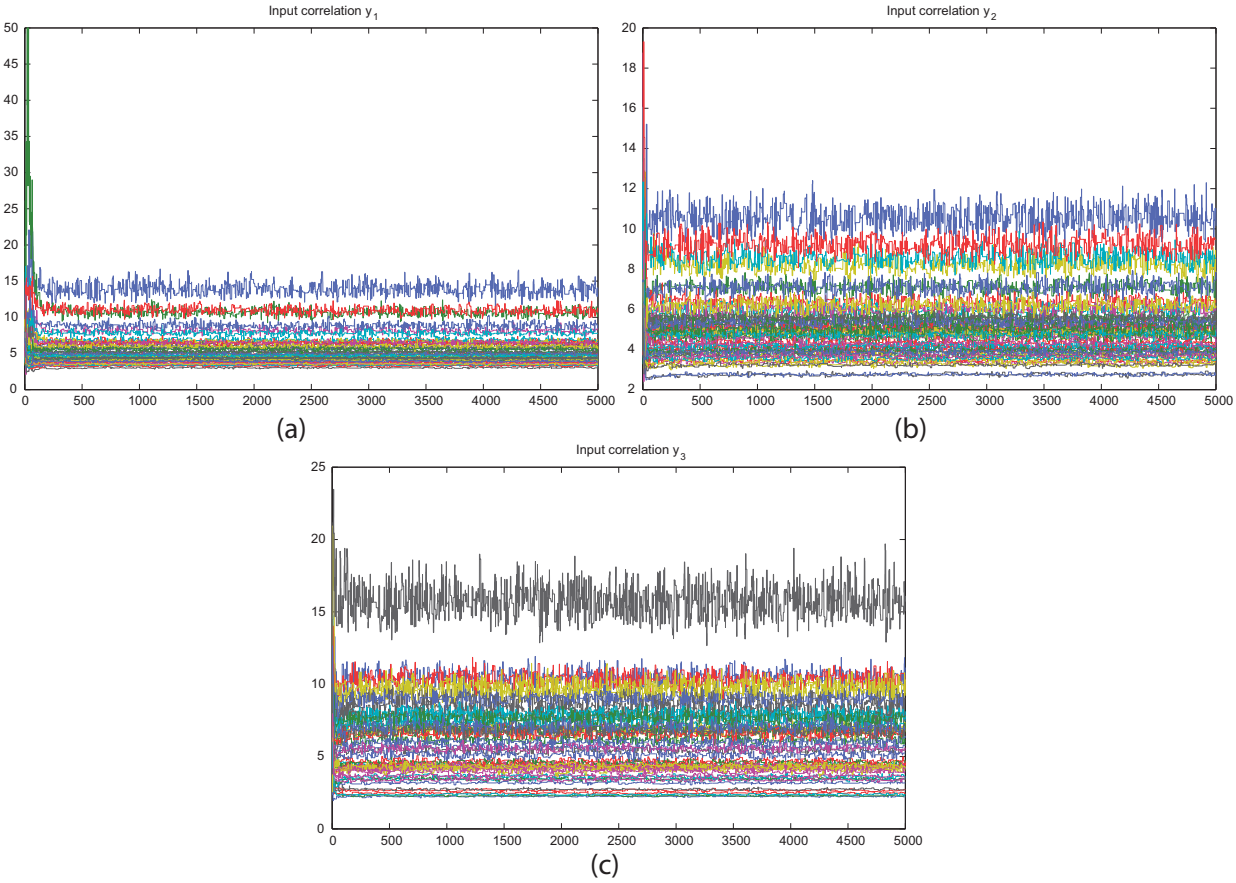

FIG. A.3: Trace plot of 5,000 MCMC iterations of the input correlation length of the three distinct conditional models.