Full scale multi-output Gaussian process emulator with nonseparable auto-covariance functions

Bohai Zhang [a], Bledar A. Konomi [b], Huiyan Sang [c], Georgios Karagiannis [d], Guang Lin [e]

[a] *Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA,* `zhang@stat.tamu.edu`

[b] *Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221-0025, USA,* `alex.konomi@uc.edu`

[c] *Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA,* `huiyan@stat.tamu.edu`

[d] *Computational Sciences and Mathematics Division, Pacific Northwest National Laboratory, 902 Battelle Blvd., P.O. Box 999, MSIN K7-90, Richland, WA 99352, USA,* `georgios.karagiannis@pnnl.gov`

[e] *Department of Mathematics and School of Mechanical Engineering, Purdue University, West Lafayete, IN 47907-2067, USA,* `guanglin@purdue.edu`

Abstract

Gaussian process emulator with separable covariance function has been utilized extensively in modeling large computer model outputs. The assumption of separability imposes constraints on the emulator and may negatively affect its performance in some applications where separability may not hold. We propose a multi-output Gaussian process emulator with a nonseparable auto-covariance function to avoid limitations of using separable emulators. In addition, to facilitate the computation of nonseparable emulator, we introduce a new computational method, referred to as the Full-Scale approximation method with block modulating function (FSA-Block) approach. The FSA-Block approach is very flexible and can apply to partially separable covariance models. We illustrate the effectiveness of our method through simulation studies and compare it with emulator with separable covariances. We also apply our method to a real computer code of the carbon capture system.

Key words: Gaussian process emulator, Uncertainty quantification, Nonseparable covariance function, Full-Scale approximation, Carbon capture unit.

1 Introduction

The computer model plays a crucial role in scientific research for studying behaviors of complex systems through computer experiments. In the context of Uncertainty Quantification (UQ), a key question of interest is to examine how computer model outputs change with different configurations of input parameters controlling physical variables, initial or boundary conditions, and so on. Although a computer model with a fine resolution is desired since it often produces more accurate simulations, it can be computationally prohibitive to produce a large number of fine resolution simulation runs at different input values, especially for computer models with high resolutions. This motivates the use of computationally inexpensive surrogate models to facilitate learning of response surface.

Gaussian process models were first used in [1] and [2] for building surrogate models for computer experiments. [3] later applied Gaussian process emulators for uncertainty quantification under the Bayesian framework. Covariance function is a key ingredient in such models since it determines the dependence structure of the Gaussian process. In the context of Gaussian process emulators, the most widely accepted auto-covariance function is usually stationary and separable in each input dimension; the cross-covariance among outputs is also assumed to be separable for mathematical tractability. For example, [4] proposed a stationary multi-output Gaussian process emulator based on separable cross-covariance. Also based on separable cross-covariance, [5] generalized the work in [6] to a Bayesian Treed Multivariate Gaussian process model, accounting for both the nonstationarity and the multivariate features of the data.

The assumption of separability allows fitting Gaussian process model in each input dimension separately. The separability structure of covariance function can alleviate the computational demand because it reduces the dimension of the covariance matrices that need to be inverted. One such example is in [7], where they introduced a multi-output separable Gaussian process model assuming the auto-covariance function of each output is separable in input, space and time. Then by making use of the properties of Kronecker product, the inverse of the covariance matrix of one output can be decomposed into the Kronecker product of inverses of an input covariance, a purely spatial covariance, and a purely temporal covariance, all of which typically have reduced dimensions so that data likelihood can be evaluated efficiently. Although the separable auto-covariance model has the aforementioned merits, it suffers from several limitations. First, it is lack of flexibility to allow for interactions between different types of correlations. [8] pointed out that if a stationary spatio-temporal covariance function is separable, then the temporal dependence structure can not vary spatially and the spatial dependence

structure can not vary temporally. However, in spatio-temporal statistics, the space-time interaction effect is often of particular interest. Such a limitation is also encountered by the separable emulator; the dependence structure of one input dimension is not allowed to change with other input dimensions. Second, the separable covariance function also has implications on conditional independence of outputs [9]. For instance, given a stationary bivariate Gaussian process $f(\cdot,\cdot)$ with a separable covariance function, it can be shown that $f(\xi,t)$ and $f(\xi',t')$ are independent given $f(\xi,t')$. A more comprehensive discussion of separable model can be found in [10].

Since the separable covariance may be restrictive in some cases, it is often desirable to consider a more general class of nonseparable auto-covariance models. In spatio-temporal statistics, much work have been done to construct flexible classes of nonseparable auto-covariance functions in space and time [8,11,12]. Typically the nonseparable space-time model has a parameter $\beta \in [0,1]$, referred to as the spatio-temporal interaction parameter, and the model reduces to be separable when $\beta = 0$. More sophisticated nonseparable covariance model of three or higher input dimensions can be constructed following the work by [13], where they extended methods in [11] to propose a nonseparable cross-covariance model for multivariate random fields. Motivated by these work in spatial statistics, we develop a flexible class of nonseparable auto-covariances for uncertainty quantification of computer models.

For computations, it is well known that the Gaussian process model scales badly with sample size n, requiring $\mathcal{O}(n^3)$ order of computations. Therefore without assuming any structures of the covariance function, e.g. the separability, the computations for the emulator with nonseparable auto-covariance models can be prohibitive when n is large. To overcome the computational bottleneck, we introduced the Full-Scale approximation (FSA) approach to reduce computations [14,15], which applies to both separable and nonseparable covariance structure. The FSA approach combines the ideas of the Gaussian predictive process [16] and covariance tapering [17] to provide a satisfactory approximation of the original covariance, under both large and small dependence scales of the data. Its computational complexity is linear with n, reducing the computations significantly.

The major contributions of this paper have two folds: first we propose to use flexible classes of nonseparable auto-covariance functions for each computer output to model the interaction effect among input, space and time. Second, we introduce the FSA approach to provide efficient computations for nonseparable Gaussian process emulator. Since the FSA approach applies to any given covariance structure of a computer model output, it can also be combined with separable model to further reduce computational cost in the case when certain input dimensions have large sample sizes for simulation accuracy. In this paper, we illustrate our method assuming a stationary covariance function for

each computer model output. We remark that our computational approach directly applies to nonstationary covariance functions as well.

The rest of this paper is organized as follows: in Section 2, we describe the multi-output Gaussian process model for computer code outputs; the discussions of nonseparable auto-covariance functions and the FSA approach are also given in Section 2. In Section 3, we describe Bayesian inference of model parameters and prediction. In Section 4, we compare the proposed nonseparable model with separable models through some simulation examples. In Section 5, we use our proposed method to analyze the computer code outputs of the regenerator device of a carbon capture unit. The potential extensions and some concluding remarks are given in Section 6.

2 Methodology

We 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 in an interval $\mathcal{X}_t = [0, T]$, where k_ξ, k_s are the dimensions of the input and spatial domain. The input domain $\mathcal{X}_\xi$ is usually assumed to bounded and can thus be considered as a compact subset of $\mathbb{R}^{k_\xi}$ while the spatial domain $\mathcal{X}_s$ and time domain $\mathcal{X}_t$ are often assumed to be intervals on $\mathbb{R}^{k_s} \times \mathbb{R}^+$.

In computer simulations, spatial and temporal domain are often fixed while simulations are run at a set of samples from the input domain. Therefore, we can represent the whole 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 $(\mathbf{s}_1, \ldots, \mathbf{s}_{n_s})^T$ and n_t time steps $(t_1, \ldots, t_{n_t})^T$. A single choice from the input domain $\boldsymbol{\xi}$ generates a multi-output response data which can be represented as a $(n_s n_t) \times q$ matrix, where q is the number of the output variables of a computer simulation.

Let $n = n_\xi n_s n_t$ denote for the total sample size. We define $\mathbf{x}_i = (\boldsymbol{\xi}_i, \mathbf{s}_i, t_i)$, denoting for an input, space and time point for $i = 1, \cdots, n$. For modeling reasons we represent the output as a q multivariate response $\mathbf{f}(\mathbf{x}_i) = \mathbf{f}(\boldsymbol{\xi}_i, \mathbf{s}_i, t_i) \in \mathbb{R}^q$. For simplicity, we call the input domain, spatial domain, and temporal domain as input, space, and time respectively throughout this paper. And we will collectively denote input of $\mathbf{f}(\cdot)$ by $\mathbf{x} = (\boldsymbol{\xi}, \mathbf{s}, t)$ and the domain of $\mathbf{x}$ by $\mathscr{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$.

We model $\mathbf{f}(\mathbf{x})$ as a q-dimensional Gaussian process:

$$\mathbf{f}(\cdot)|B, \boldsymbol{\theta} = \mathcal{N}_q(\boldsymbol{\mu}(\cdot; B), \Gamma(\cdot, \cdot; \boldsymbol{\theta})), \tag{1}$$

where $\boldsymbol{\mu}(\cdot; B)$ is the mean function and $\Gamma(\cdot, \cdot; \boldsymbol{\theta})$ is the covariance function of the q-dimensional Gaussian process $\mathbf{f}(\mathbf{x})$. A typical choice of the mean function $\boldsymbol{\mu}(\cdot; B)$ is the linear regression model: $\boldsymbol{\mu}(\mathbf{x}) = \mathbf{h}^T(\mathbf{x})B$, where $\mathbf{h}(\mathbf{x})$ is formed by m basis functions evaluated at $\mathbf{x}$ and B is a $m \times q$ unknown regression coefficients matrix. For covariance function $\Gamma(\cdot, \cdot; \boldsymbol{\theta})$, if the cross-covariance is separable [18], then $\Gamma(\cdot, \cdot; \boldsymbol{\theta}) = \rho(\cdot, \cdot; \boldsymbol{\theta})\Sigma$, where Σ is the covariance matrix that models the cross-dependence structure of q distinct components of $\mathbf{f}(\cdot)$, and $\rho(\cdot, \cdot; \boldsymbol{\theta})$ is the auto-correlation of each component. In this work we will assume the cross-covariance among multivariate components and auto-correlation within each component are separable for simplicity. More general nonseparable cross-covariance models can be constructed following the work in [19].

Let $Y = (\mathbf{f}^T(\mathbf{x}_1), \mathbf{f}^T(\mathbf{x}_2), \cdots, \mathbf{f}^T(\mathbf{x}_n))^T$ be the $n \times q$ computer model output matrix, $H = (\mathbf{h}(\mathbf{x}_1), \ldots, \mathbf{h}(\mathbf{x}_n))^T \in \mathbb{R}^{n \times m}$ be the design matrix, and $R = [\rho(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1:n} \in \mathbb{R}^{n \times n}$ be the correlation matrix at a given set of n points $\mathscr{X}$. The data likelihood function is given by the following matrix normal distribution

$$\begin{aligned} Y|B, \Sigma, \boldsymbol{\theta} &\sim \mathcal{N}_{n \times q}(HB, R, \Sigma), \\ &\propto |R|^{-q/2}|\Sigma|^{-n/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}(Y - HB)^T R^{-1}(Y - HB)\right)\right). \end{aligned} \tag{2}$$

To evaluate the above likelihood function, we need to compute the determinant and inverse of the $n \times n$ matrix R. When n is very large, the computation burden can often lead to failures in calculating these quantities. In Section 2.3, we will introduce a covariance approximation method to facilitate the computations for likelihood evaluations.

Squared exponential kernel function is one classical choice of the separable auto-correlation function $\rho(\cdot, \cdot)$,

$$\rho(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^{k_\xi} \frac{(\xi_i - \xi_i')^2}{\phi_i^2} - \sum_{j=1}^{k_s} \frac{(s_j - s_j')^2}{c_j^2} - \frac{(t - t')^2}{a^2}\right),$$

where ϕ_i, c_i and a are dependence range parameters of input, space and time respectively. Since $\rho(\mathbf{x}_i, \mathbf{x}_j) = \rho_\xi(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)\rho_s(\mathbf{s}_i, \mathbf{s}_j)\rho_t(t_i, t_j)$, it assumes the sepa-

rability in input, space and time. When computer model outputs are generated on a $n_\xi \times n_s \times n_t$ regular grid of input,space and time and data are ordered properly, we have $R = R_\xi \otimes R_s \otimes R_t$, where $R_\xi = [\rho_\xi(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)]_{i,j=1,\ldots,n_\xi}$, $R_s = [\rho_s(s_i, s_j)]_{i,j=1,\ldots,n_s}$, $R_t = [\rho_t(t_i, t_j)]_{i,j=1,\ldots,n_t}$. Since $R^{-1} = R_\xi^{-1} \otimes R_s^{-1} \otimes R_t^{-1}$ and $|R| = |R_\xi|^{n_s n_t} |R_s|^{n_\xi n_t} |R_t|^{n_\xi n_s}$, the computations of evaluating the likelihood can be greatly reduced when n is very large.

Although the computer model is deterministic, a small variance term $\tau^2 \delta_{\mathbf{x}=\mathbf{x}'}$ is usually added to $\rho(\mathbf{x}, \mathbf{x}'; \boldsymbol{\theta})$ for numerical stability, where δ is the Kronecker delta function. This small variance term is also referred to as the "nugget" effect in spatial statistics, accounting for the measurement error. We incorporate the nugget effect parameter τ^2 in $\boldsymbol{\theta}$ throughout the paper for simplicity. For auto-correlation model that is separable in input, space and time, three nugget terms $\tau_\xi^2 \delta_{\xi_i=\xi_j}$, $\tau_s^2 \delta_{\mathbf{s}_i=\mathbf{s}_j}$, and $\tau_t^2 \delta_{t_i=t_j}$ are added to $\rho_\xi(\cdot,\cdot)$, $\rho_s(\cdot,\cdot)$, and $\rho_t(\cdot,\cdot)$ respectively. Adding nuggets in this way can preserve the separability so that fast computations can still be achieved.

2.2 Nonseparable auto-correlation models

Although the separable auto-correlation model is easy to construct and leads to reduced computational costs, it may be too restrictive in some applications due to its implications on the covariance dependence structures. Take a computer model output $f(\cdot,\cdot)$ with two dimensional input parameters (ξ, t) as an example, the separable covariance implies that

$$\mathrm{Corr}((f(\xi,t), f(\xi',t'))|f(\xi,t')) = 0,$$

under the Gaussian process assumption [9,10]. This implication on conditional correlation may be too restrictive in some applications. Besides, the separability structure also implies that

$$\mathrm{Corr}(f(\xi,t), f(\xi,t')) = \rho_t(t,t'),$$

which means the correlation structure of dimension t can not vary over the dimension ξ. [10] showed that process $f(\xi,t)$ has a separable covariance function if and only if $f(\xi,t) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} Z_{i,j} f_i^{(1)}(\xi) f_j^{(2)}(t)$, where $Z_{i,j}$ are random variables such that $E(Z_{i,j}) = 0$ and $E(Z_{i,j} Z_{i',j'}) = \delta_{i=i'}\delta_{j=j'}$, in which δ is the Kronecker delta function. [10] also gives more comprehensive discussions of the implications of separable covariance function for emulation.

Therefore, the nonseparable auto-covariance model may be preferred when these implications are not true for the dataset. We propose to use the nonseparable covariance functions [8,11,12] for the Gaussian process emulator, since

they are more general and include separable covariance functions as special cases. One example of nonseparable correlation functions in input and time is

$$\rho(\mathbf{x},\mathbf{x}') = \left(\frac{|v|^{2\alpha}}{a}+1\right)^{-\frac{k_\xi}{2}} \exp\left(-\frac{\sqrt{\sum_{i=1}^{k_\xi}|u_i|^2/\phi_i^2}}{\left(\frac{|v|^{2\alpha}}{a}+1\right)^{\beta/2}}\right), \tag{3}$$

where $u_i = |\xi_i - \xi_i'|$, $v = |t-t'|$; $a > 0$ is dependence range parameter in time, $\phi_i > 0$ is dependence range parameter for i^{th} input dimension, $\alpha \in (0,1]$ is the smoothness parameter and $\beta \in [0,1]$ is the input-time interaction parameter. When $\beta = 0$, it reduces to the separable case. More sophisticated nonseparable model in input, space and time can be constructed following the work in [13]. Following [13], a totally nonseparable correlation model in input, space and time $\rho(\mathbf{x},\mathbf{x}')$ is

$$\begin{aligned}\rho(\mathbf{h},\mathbf{u},v) &= \left(a_1\left(\frac{\|\mathbf{u}\|^2}{(a_4|v|^{\alpha_4}+1)^{\beta_4}}\right)^{\alpha_1}+1\right)^{-\beta_1 k_\xi/2}(a_2|v|^{2\alpha_2}+1)^{-\beta_2 k_s/2}\\ &\times(a_3\|\mathbf{h}\|^{2\alpha_3}+1)^{-\beta_3/2}(a_4|v|^{2\alpha_4}+1)^{-\beta_4 k_s/2}\exp\left(-\frac{c_1\|\mathbf{h}\|^{2\gamma_1}}{\left(a_1\left(\frac{\|\mathbf{u}\|^2}{(a_4|v|^{2\alpha_4}+1)^{\beta_4}}\right)^{\alpha_1}+1\right)^{\beta_1\gamma_1}}\right)\\ &\times\exp\left(-\frac{c_2\|\mathbf{u}\|^{2\gamma_2}}{(a_2|v|^{2\alpha_2}+1)^{\beta_2\gamma_2}}-\frac{c_3|v|^{2\gamma_3}}{(a_3\|\mathbf{h}\|^{2\alpha_3}+1)^{\beta_3\gamma_3}}\right),\end{aligned} \tag{4}$$

where $\mathbf{h} = |\boldsymbol{\xi} - \boldsymbol{\xi}'|, \mathbf{u} = |\mathbf{s} - \mathbf{s}'|$ and $v = |t-t'|$; $c_i > 0, \gamma_i \in (0,1]$ for $i = 1,2,3$ and $a_j > 0, \alpha_j \in (0,1], \beta_j \in [0,1]$ for $j = 1,\ldots,4$. $\alpha_1,\ldots,\alpha_4$ and $\gamma_1,\gamma_2,\gamma_3$ can be interpreted as smoothness parameters; $a_1,\ldots,a_4$ and c_1,c_2,c_3 are scale parameters; $\beta_1,\ldots,\beta_4$ are interaction parameters, modeling the two-way and three-way interactions among input, space and time. If we fix the smoothness parameters $\gamma_j = 0.5, j = 1,2,3$, then it is more clearly to see that β_j's determine the interaction effects. Although the model (4) is very flexible, it involves too many unknown model parameters. In this paper, we will focus on the nonseparable auto-correlation model between two components of input, space and time, with similar form to (3).

2.3 FSA-Block approximation

Since the computations of R^{-1} and $|R|$ in (2) become expensive or even infeasible when n is large, we need to employ some computational techniques to overcome the computation bottleneck. There are several existing paradigms to facilitate computations of the Gaussian process model, here we resort to covariance matrix approximation technique which gains its popularity recently. Popular covariance approximation models include the Gaussian predictive

process [16,20], the fixed rank kriging model [21], and the covariance tapering [17,22], to name a few.

In this work we propose to use the Full-Scale Approximation with Block modulating function (FSA-Block) to speed up computations [14,15]. It consists of a summation of a reduced rank covariance and a sparse covariance with the block diagonal structure. This approach combines the merits of both reduced rank and sparse covariances without adding much computational complexity.

In the following we will describe the FSA-Block approach for a Gaussian process with zero mean, unit variance and a correlation function $\rho(\cdot,\cdot)$. The FSA-Block approximation is motivated from the Karhunen-Loéve orthogonal expansion (K-L expansion) of the Gaussian Process, which decomposes a covariance function as:

$$\rho(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{\infty} \lambda_i \psi_i(\mathbf{x})\psi_i(\mathbf{x}'), \tag{5}$$

where λ_i are the eigenvalues of the process, $\psi_i(\mathbf{x})$ are the corresponding orthonormal eigenfunctions; the eigenvalue-eigenfunction pairs are solutions to the integral equation $\int_D \rho(\mathbf{x},\mathbf{t})\psi_i(\mathbf{t})d\mathbf{t} = \lambda_i\psi_i(\mathbf{x})$.

The leading terms in (5) are often assumed to capture the main feature of the covariance and thus the residual terms are typically dropped from the expansion to yield a reduced rank approximation of the covariance. Although increasing rank can preserve more complete information about the fine scale covariance pattern, computations become more expensive. Motivated from the decomposition in (5), we give a more careful treatment of the covariance that can preserve most information present in both the leading reduced rank terms and the residual covariance yet still achieves computational efficiency.

Solving the integral equation for K-L expansion is typically a challenging task. We use the Nyström discretization [23], a numerical method for solving integral equations, to approximate the reduced rank part of the K-L decomposition. Consider a set of knots $\mathscr{X}^* = \{\mathbf{x}_1^*, \ldots, \mathbf{x}_{n^*}^*\}$. Let R_{**} denote the $n^* \times n^*$ correlation matrix whose (i,j) entry is $\rho(\mathbf{x}_i^*, \mathbf{x}_j^*)$. Let $\{\mathbf{u}_i^{(n^*)}\}$ and $\{\lambda_i^{(n^*)}\}$ be the eigenvectors and the eigenvalues for the correlation matrix R_{**}. The Nyström approximation of the leading n^* eigenfunctions and eigenvalues for the correlation kernel $\rho(\mathbf{x},\mathbf{x}')$ are

$$\psi_i(\mathbf{x}) \approx \frac{\sqrt{n^*}}{\lambda_i^{(n^*)}}\rho(\mathbf{x}, \mathscr{X}^*)\mathbf{u}_i^{(n^*)}, \quad \lambda_i \approx \frac{\lambda_i^{(n^*)}}{n^*}, \quad \text{for} \quad i = 1, \cdots, n^*.$$

It can be further proved that the Nyström approximation method leads to a reduced rank correlation

$$\rho_{pp}(\mathbf{x}, \mathbf{x}') = \rho(\mathbf{x}, \mathscr{X}^*) R_{**}^{-1} \rho^T(\mathbf{x}', \mathscr{X}^*),$$

where $\rho(\mathbf{x}, \mathscr{X}^*) = [\rho(\mathbf{x}, \mathbf{x}')]_{\mathbf{x}' \in \mathscr{X}^*}$. Using the reduced rank model derived from the Nyström approximated correlation, $\rho(\mathbf{x}, \mathbf{x}') - \rho_{pp}(\mathbf{x}, \mathbf{x}')$ remains to be positive semi-definite following the Schur complement property in linear algebra. We approximate it by multiplying the residual correlation with a modulating function $\mathcal{K}(\mathbf{x}, \mathbf{x}')$, that is, $\rho_\epsilon(\cdot, \cdot) = (\rho(\cdot, \cdot) - \rho_{pp}(\cdot, \cdot))\mathcal{K}(\cdot, \cdot)$. The modulating function has to be chosen to ensure $\rho_\epsilon(\mathbf{x}, \mathbf{x}')$ is positive semi-definite. We also assume it has the property of having zero entries for a large proportion of possible location pairs $(\mathbf{x}, \mathbf{x}')$ so that $\rho_\epsilon(\cdot, \cdot)$ evaluated on $\mathscr{X}$ is a sparse matrix. One specific choice of $\mathcal{K}(\cdot, \cdot)$ is the block modulating function. Given a partition of observed locations $\cup_{i=1}^K B_i = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} = \mathscr{X}$, it is defined as

$$\mathcal{K}_{block}(\mathbf{x}, \mathbf{x}') = \begin{cases} 1 & \text{if } \mathbf{x}, \mathbf{x}' \in B_i, i = 1, \ldots, K; \\ 0 & \text{Otherwise.} \end{cases}$$

If observations are grouped together within each B_i, the $\rho_\epsilon(\cdot, \cdot)$ on $\mathscr{X}$ yields a block-diagonal matrix whose inverse can be computed easily.

The FSA-Block method approximates the parent correlation function ρ by the sum of $\rho_{pp}(\cdot, \cdot)$ and an approximated residual correlation function $\rho_\epsilon(\cdot, \cdot)$:

$$\rho^\dagger(\mathbf{x}, \mathbf{x}') = \rho_{pp}(\mathbf{x}, \mathbf{x}') + \rho_\epsilon(\mathbf{x}, \mathbf{x}'),$$

which is still a valid correlation function. Under this correlation approximation, R is approximated by $R^\dagger = R_{nn^*} R_{**}^{-1} R_{nn^*}^T + R_\epsilon$, where $R_{nn^*} = \rho(\mathscr{X}, \mathscr{X}^*)$ and $R_\epsilon = \rho_\epsilon(\mathscr{X}, \mathscr{X})$. The Sherman-Woodbury-Morrison inversion formula yields

$$R^{\dagger -1} = R_\epsilon^{-1} - R_\epsilon^{-1} R_{nn^*} (R_{nn^*}^T R_\epsilon^{-1} R_{nn^*} + R_{**})^{-1} R_{nn^*}^T R_\epsilon^{-1}. \tag{6}$$

Thus $R^{\dagger -1}$ involves the calculations of inverses of a block diagonal matrix R_ϵ and a $n_* \times n_*$ matrix $R_{nn^*}^T R_\epsilon^{-1} R_{nn^*} + R_{**}$. If we choose n^* and block size to be small, the computations of R^{-1} using $R^{\dagger -1}$ can be greatly reduced. Indeed, the computational complexity of calculating $R^{\dagger -1}$ is $\mathcal{O}(nn^{*2} + nn_B^2)$, where n_B is the average block size. The determinant of $R^\dagger$ can also be computed efficiently using Sylvester's determinant theorem

$$|R^\dagger| = |R_\epsilon||R_{**}|^{-1}|R_{**} + R_{nn^*}^T R_\epsilon^{-1} R_{nn^*}|. \tag{7}$$

So instead of computing the determinant of a big $n \times n$ matrix, we only need to compute the determinants of a $n^* \times n^*$ matrix and a block-diagonal matrix.

As described above, fast computations can be achieved using the FSA-Block approach. The correlation function of the FSA-Block approach is

$$\rho^{\dagger}(\mathbf{x}, \mathbf{x}') = \begin{cases} \rho(\mathbf{x}, \mathbf{x}') & \text{if } \mathbf{x}, \mathbf{x}' \in B_i, i = 1, \ldots, K; \\ \rho_{pp}(\mathbf{x}, \mathbf{x}') & \text{otherwise.} \end{cases}$$

Therefore, the correlation within blocks are preserved exactly and the correlation across blocks are approximated by that of the predictive process part. Since the FSA-Block approach provides a general way of approximating any given covariance functions without further restrictions on the parent covariance structures, it also applies to the separable auto-covariance model in which computations in certain dimensions are infeasible due to large sample sizes in those dimensions. For example, if learning the response on a highly fine-resolution spatial grid is desirable in certain studies, the FSA-Block approach can be applied only to a spatial correlation function to facilitate computations.

3 Bayesian inference of model parameters and prediction

3.1 Prior specifications

For the multivariate Gaussian process regression model, the unknown parameter set is $\{B, \Sigma, \boldsymbol{\theta}\}$. We assume the prior distributions of $\{B, \Sigma\}$ and $\boldsymbol{\theta}$ are independent, namely $\pi(B, \Sigma, \boldsymbol{\theta}) = \pi(B, \Sigma)\pi(\boldsymbol{\theta})$. For $\pi(B, \Sigma)$, we assign a non-informative conjugate prior

$$\pi(B, \Sigma) \propto |\Sigma|^{-\frac{q+1}{2}}. \tag{8}$$

The prior specification of $\pi(\boldsymbol{\theta})$ depends on the specific form of the covariance function. Customarily, the inverse-gamma prior can be assigned on the nugget τ^2; the input (spatial/temporal) range parameter can be assigned with a reasonably informative prior, e.g. an uniform prior with its support specified according to the belief of the practical input (spatial/temporal) dependence range of the computer model outputs; for the smoothness parameter, a uniform prior with a reasonable support reflecting prior information about the smoothness of the process can be assigned.

If a model parameter has a closed-form full conditional distribution, we draw its posterior samples using the Gibbs sampler; otherwise we draw its posterior samples using the Metropolis-Hasting (M-H) algorithm [24]. A log-normal

distribution or a truncated normal distribution centered at the current value can be used as the proposal distribution in the M-H algorithms.

3.2 Posterior distributions of model parameters

By using the conjugate prior in (8), it is easy to see the posterior conditional distributions of B and Σ have the following closed-forms:

$$B|Y, \Sigma, \boldsymbol{\theta} \sim \mathcal{N}_{m\times q}(\hat{B}_{gls}, (H^T R^{-1} H)^{-1}, \Sigma), \tag{9}$$
$$\Sigma|Y, B, \boldsymbol{\theta} \sim \mathcal{IW}(n, (Y - HB)^T R^{-1}(Y - HB)), \tag{10}$$

where $\mathcal{N}_{m\times q}$ stands for the matrix-normal distribution, $\mathcal{IW}$ stands for the Inverse Wishart distribution, and $\hat{B}_{gls}$ is the generalized least squares estimator of B, i.e., $\hat{B}_{gls} = (H^T R^{-1} H)^{-1} H^T R^{-1} Y$. We can integrate out B or Σ in (9) and (10) respectively to improve the mixing of posterior samples [4,25]. Specifically, by integrating out Σ, we obtain

$$B|Y, \boldsymbol{\theta} \sim \mathcal{T}_{m\times q}(\hat{B}_{gls}, (H^T R^{-1} H)^{-1}/\nu, (Y^T R^{-1} Y - \hat{B}_{gls}^T (H^T R^{-1} H)\hat{B}_{gls}), \nu),$$

where $\mathcal{T}$ stands for a matrix-t distribution with degrees of freedom $n-m-q+1$; similarly by integrating out B,

$$\Sigma|Y, \boldsymbol{\theta} \sim \mathcal{IW}(n - m, Y^T R^{-1} Y - \hat{B}_{gls}^T (H^T R^{-1} H)\hat{B}_{gls}).$$

Since $p(\boldsymbol{\theta}|Y)$ does not have a closed-form, we need the Metropolis-Hasting algorithm to obtain posterior samples of $\boldsymbol{\theta}$. By integrating out B and Σ

$$p(\boldsymbol{\theta}|Y) \propto \pi(\theta)|R|^{-\frac{q}{2}}|H^T R^{-1} H|^{-\frac{q}{2}}|(Y - H\hat{B}_{gls})^T R^{-1}(Y - H\hat{B}_{gls})|^{-\frac{n-m}{2}}, \tag{11}$$

When sample size n is large, we can replace R in (9), (10), and (11) with the FSA-Block approximated correlation matrix $R^\dagger$ introduced in Section 2.3, and then apply the inversion formula (6) to efficiently calculate the inverse of $R^\dagger$.

3.3 Prediction

The Bayesian predictive distribution provides a natural measure on the function space of surrogate models. At a new point $\mathbf{x}_p$, the predictive distribution has its mean

$$E(\mathbf{f}(\mathbf{x}_p)|B, \boldsymbol{\theta}, Y) = \mathbf{h}^T(\mathbf{x}_p)B + R_{p,n}R^{-1}(Y - HB) \tag{12}$$

and variance

$$\mathrm{Var}(\mathbf{f}(\mathbf{x}_p)|B, \Sigma, \boldsymbol{\theta}, Y) = (1 - R_{p,n}R^{-1}R_{p,n}^T) \otimes \Sigma, \tag{13}$$

where $R_{p,n} = [\rho(\mathbf{x}_p, \mathbf{x})]_{\mathbf{x}\in\mathscr{X}}$ is the $1 \times n$ correlation vector. When n is too large so that R^{-1} is computationally prohibitive, we can again apply the FSA-Block approximation method to approximating R with $R^\dagger$ whose inversion can be done efficiently using (6).

In uncertainty quantification, the mean response surface of each output averaged over input space as well as its corresponding error bar are of key interest. We can obtain a sample of the mean response of the ith output by integrating out the input parameters $\boldsymbol{\xi}$ in $\mathbf{x}_p$ in (12),

$$M_p^i = \bar{\mathbf{h}}_p^T B_i + \bar{R}_p R^{-1}(Y_i - HB_i), \tag{14}$$

where $Y_i = (f_i(\mathbf{x}_1), \ldots, f_i(\mathbf{x}_n))^T$, $B_i = (B_{1i}, \ldots, B_{mi})^T$, $\bar{h}_p = \int \mathbf{h}(\mathbf{x}_p)p(\boldsymbol{\xi})d\boldsymbol{\xi}$, $\bar{R}_p = \int R_{p,n}p(\boldsymbol{\xi})d\boldsymbol{\xi}$, and $p(\boldsymbol{\xi})$ is the joint density of input variables. The covariance of M_p^i is

$$V_p^i = \int E(f_i(\mathbf{x}_p)|B_i, \boldsymbol{\theta}, Y_i)E(f_i(\mathbf{x}_p)|B_i, \boldsymbol{\theta}, Y_i)^T p(\boldsymbol{\xi})d\boldsymbol{\xi} - M_p^i M_p^{iT}. \tag{15}$$

When the above integrations don't have close-forms, the Monte Carlo method on a dense grid of the input space can be used to approximate them.

4 Numerical results

4.1 2-input and 1-output example

We use a simulation example to show that the nonseparable covariance model can outperform the separable model in some cases. We consider the following function to generate output data,

$$f(x_1, x_2) = x_1 \exp\left(-\sqrt{x_1^2 + x_2^2}\right),$$

where the inputs $x_1, x_2 \in [-6, 6]$. In this case $f(x_1, x_2) \neq f_1(x_1)f_2(x_2)$, and hence the separable covariance model may not be adequate according to the theory in [10].

We then use a Gaussian process model (1) with constant means as a surrogate to fit the simulated function values. We consider the following nonseparable covariance model [11],

$$\mathcal{C}(f(x_1, x_2), f(x_1', x_2')) = \frac{\sigma^2}{\left(\frac{v^{2\alpha}}{a} + 1\right)} \exp\left(-\frac{u^2}{c\left(\frac{v^{2\alpha}}{a} + 1\right)^{\beta}}\right), \tag{16}$$

where $u = |x_1 - x_1'|$ and $v = |x_2 - x_2'|$. We also consider two separable models for comparisons; (a). the covariance model as in (16) but with the interaction parameter $\beta = 0$ (denoted by "Sep"), and (b). the commonly used squared exponential covariance model (denoted by "Sqexp"). A fixed small nugget effect $\tau^2 = 10^{-6}$ was added to the covariance function for numerical stability. We experimented with different sample sizes n for the training set. We also fixed a prediction set and evaluate the mean squared prediction errors (MSPE) to compare the prediction performance of different covariance models. Specifically, the training sets were $n = 200, 500$ function values evaluated at locations selected by Latin Hypercube Sampling (LHS). The prediction set was fixed to be 100 function values evaluated at hold-out locations selected by LHS. Uniform priors with a reasonable support were assigned to the dependence range parameters a and c; the uniform prior on $[0, 1]$ was assigned to the smoothness parameter α and interaction parameter β in (16). We collected 6000 posterior samples after a burn-in period of 1000 iterations. Then the posterior means of model parameters were plugged in (12) to obtain prediction results. The parameter estimation and prediction results are summarized in Table 1.

Table 1
Posterior means and MSPEs.

		a	c	α	β	B_0	σ^2	MSPE
$n = 200$	Nonsep	3.728	4.633	1.000	0.641	-0.001	0.0061	$1.67 \cdot 10^{-5}$
	Sep	2.978	3.711	1.000	0	$-6.59 \cdot 10^{-5}$	0.0053	$3.67 \cdot 10^{-5}$
	Sqexp	1.105	1.984	–	–	$-4.28 \cdot 10^{-4}$	0.0033	$3.88 \cdot 10^{-5}$
$n = 500$	Nonsep	2.693	2.314	0.990	0.721	$2.30 \cdot 10^{-4}$	0.0039	$4.28 \cdot 10^{-7}$
	Sep	2.246	1.987	0.979	0	$3.28 \cdot 10^{-5}$	0.0033	$1.57 \cdot 10^{-6}$
	Sqexp	0.933	1.016	–	–	$3.31 \cdot 10^{-5}$	0.0023	$9.58 \cdot 10^{-6}$

First we observed that in both experimental cases, the posterior mean estimates of β of the nonseparable model are far from zero, indicating the existence of the interaction effect. Also for both cases, the nonseparable model outperforms the separable models in terms of the prediction. For relatively large sample size $n = 500$, the estimation of β becomes more accurate (posterior variance reduces from 0.021 to 0.008) and the prediction results of the nonsep-

arable model are obviously better than those of the two separable covariance models.

4.2 Krainchnan-Orszag three-mode problem

In this example, we consider the system of ordinary differential equations with respect to t as in [26],

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

subject to stochastic initial conditions $y_1(0) = 1, y_2(0) = 0.1\xi_1, y_3(0) = \xi_2$, where $\xi_i \sim U(-1, 1), i = 1, 2$. This problem has 2 input variables and 3 outputs. It is of interest because the response has a discontinuity line at $\xi_1 = 0$, inducing a nonstationary response surface in input space. Here we applied the stationary nonseparable covariance model to this problem with a relatively large sample size n to obtain reasonable prediction results. However the Bayesian inference is computationally intensive due to large sample size, hence the FSA-Block approach was applied to the nonseparable model for computational efficiency.

The training set was obtained at 600 input points selected by Latin Hypercube Sampling on a time grid $T = 1, 2, \ldots, 10$. We considered the validation set of a 31×31 input grid on time points 11 and 12, which allows us to assess prediction performance in both input and time scenario. The multivariate Gaussian process model in (1) with constant means were used to fit for each output $y_i(t), i = 1, 2, 3$. The nonseparable auto-correlation function considered was the model in (3) with $k_\xi = 2$, and the same two separable models as in the previous simulation study were used for comparison purpose. A fixed nugget $\tau^2 = 10^{-6}$ was added to the covariance model to improve numerical stability. When implementing the nonseparable model, we applied the FSA-Block approach with 20 input knots selected by LHS at each time grid point (in total 200 knots) and 30 blocks created by K-means clustering algorithm. The prior specifications of model parameters were similar to those in the previous example. After a burn-in period of 1000 iterations, we collected 6000 posterior samples for inference. The posterior means of model parameters were plugged in (12) to obtain the predictive response surface.

Table 2 shows the parameter estimations and prediction results of the three

Table 2
Posterior means of model parameters and MSPEs. MSPE results for each output are displayed as the format of $(y_1(t), y_2(t), y_3(t))$

$n_\xi = 600$	Nonsep	Sep	Sqexp
a	3.306	1.315	1.265
c_1	5.748	6.540	4.608
c_2	21.648	22.796	16.351
α	0.999	0.999	–
β	0.996	0	–
B_{10}	0.478	0.461	0.474
B_{20}	−0.106	0.002	−0.006
B_{30}	−0.044	−0.209	−0.197
Σ_{11}	0.296	0.505	0.312
Σ_{12}	0.007	−0.001	0.0003
Σ_{13}	−0.017	−0.047	−0.033
Σ_{22}	1.257	1.187	0.768
Σ_{23}	0.108	0.114	0.072
Σ_{33}	1.567	1.729	1.107
$\text{MSPE}_{sp\&t}$	$(0.078, 0.233, 0.610)$	(0.090, 0.247, 0.674)	$(0.092, 0.290, 0.594)$

covariance models. The posterior mean of β by the nonseparable model is very close to 1, suggesting that modeling the interaction between input and time may be beneficial. The separable model (3) with $\beta = 0$ produced parameter estimates close to those from the nonseparable model, except for a much smaller estimate of the time dependence parameter. In terms of the prediction, the nonseparable model obviously outperforms the separable models for $y_1(t), y_2(t)$, with slightly inferior performance to the squared exponential model for $y_3(t)$.

We then checked the predictive input surface of each output by the nonseparable model at selected time points, and the results are shown in Figure 1, Figure 2 and Figure 3, respectively. We randomly chose the results at time points 5 and 10 for illustrations. For $y_1(t)$ and $y_2(t)$, the predictive input surfaces by the nonseparable model are very close to the true response surfaces in general, but the prediction errors are high around $\xi_1 = 0$ for $y_3(t)$ at time point 10. Figure 4 shows the MSPE surfaces of 3 outputs in input space by averaging MSPEs over time, and it is more clear that the prediction errors of $y_3(t)$ peaked at $\xi_1 = 0$. Recall that here the computer model outputs have a discontinuous point at $\xi_1 = 0$. Therefore, the stationary covariance function may not be adequate to model this nonstationary feature. In order to have better prediction results, we need more observations sampled around $\xi_1 = 0$ for the stationary covariance models or consider a nonstationary covariance model. We also checked the predictive input surfaces of 3 outputs by the 2 separable models, and the results are similar to the nonseparable model results.

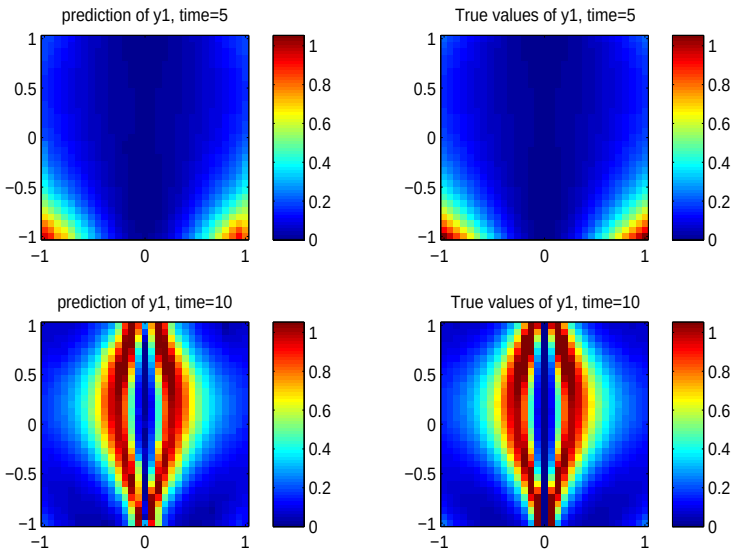


Fig. 1. Predictive surface of input space of y_1 at time points 5 and 10 using the nonseparable model.

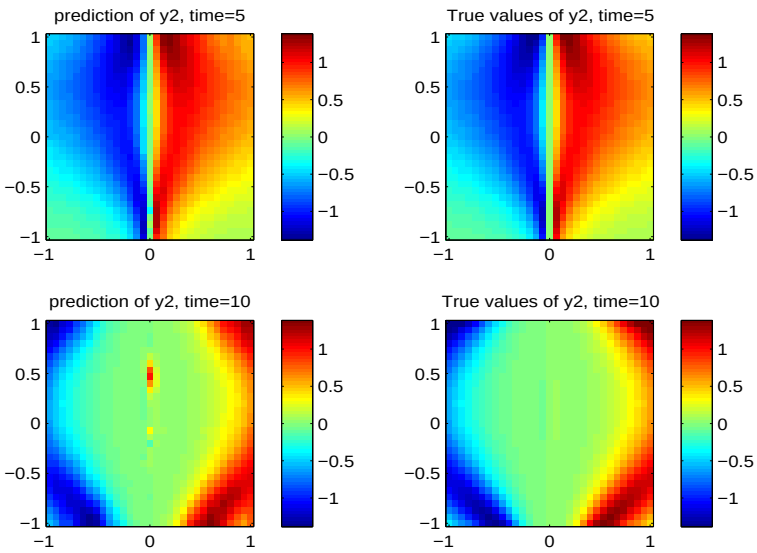


Fig. 2. Predictive surface of input space of y_2 at time points 5 and 10 using the nonseparable model.

The predictive mean time curve is also of crucial interest, which shows the shape of the mean response curve averaged over input space. We obtained the predictive mean response curve of t for each output over 100 LHS selected input points. Then predictions were done at 50 equidistant time steps in $[0, 10]$ and it was repeated for 100 times to obtain the error bars of the predictive mean time curve. To calculate the mean of a sampled response curve, we used the method in (14). The results of computer code outputs at integer time grid

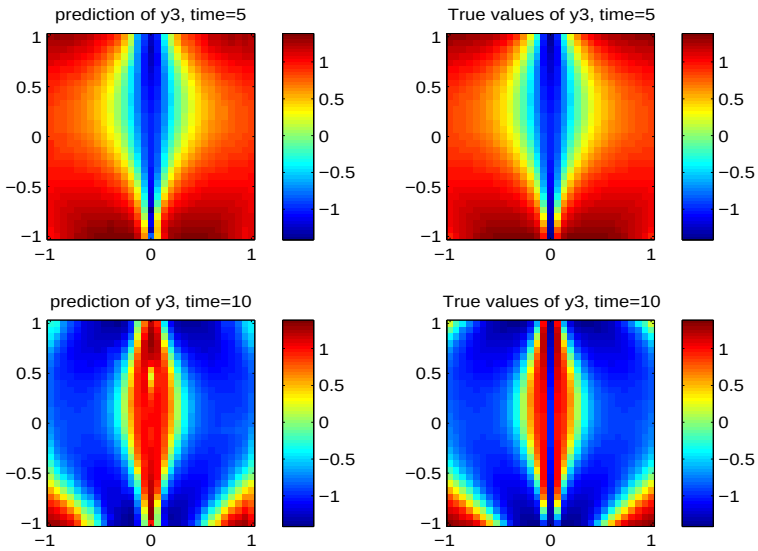


Fig. 3. Predictive surface of input space of y_3 at time points 5 and 10 using the nonseparable model.

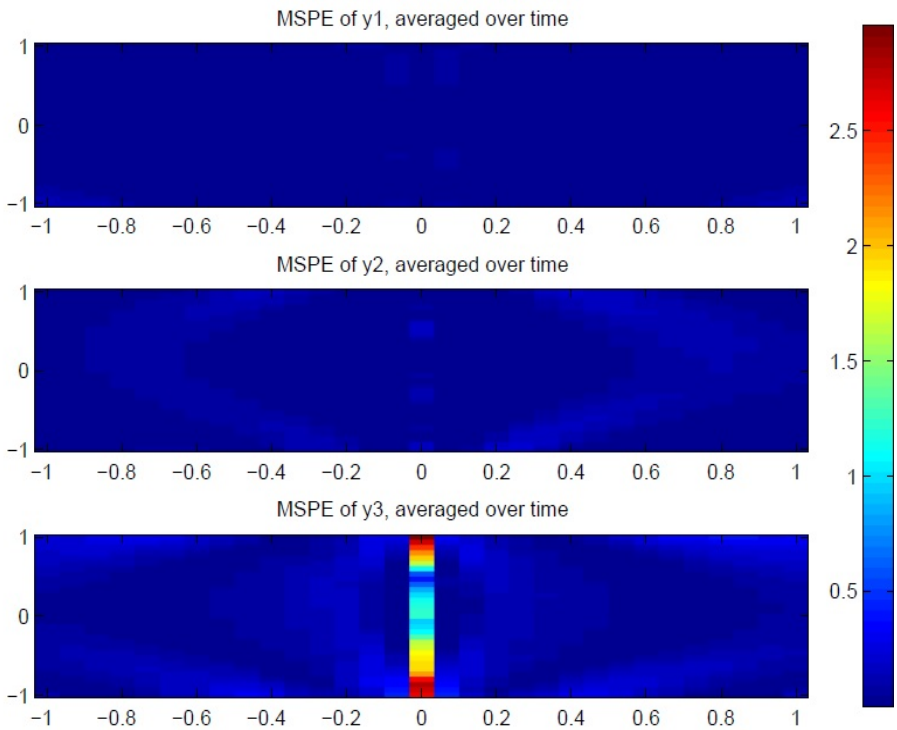


Fig. 4. The MSPEs averaged over time by the nonseparable model.

were used as baselines. Figure 5 shows the predictive mean curves for $y_1(t)$ and $y_3(t)$ as a function of time, as well as their corresponding 95% confidence intervals using the nonseparable model. We can see that the error bars of the predictive mean curves are tight and can cover the true means of the computer

code outputs, indicating the effectiveness of the nonseparable model.

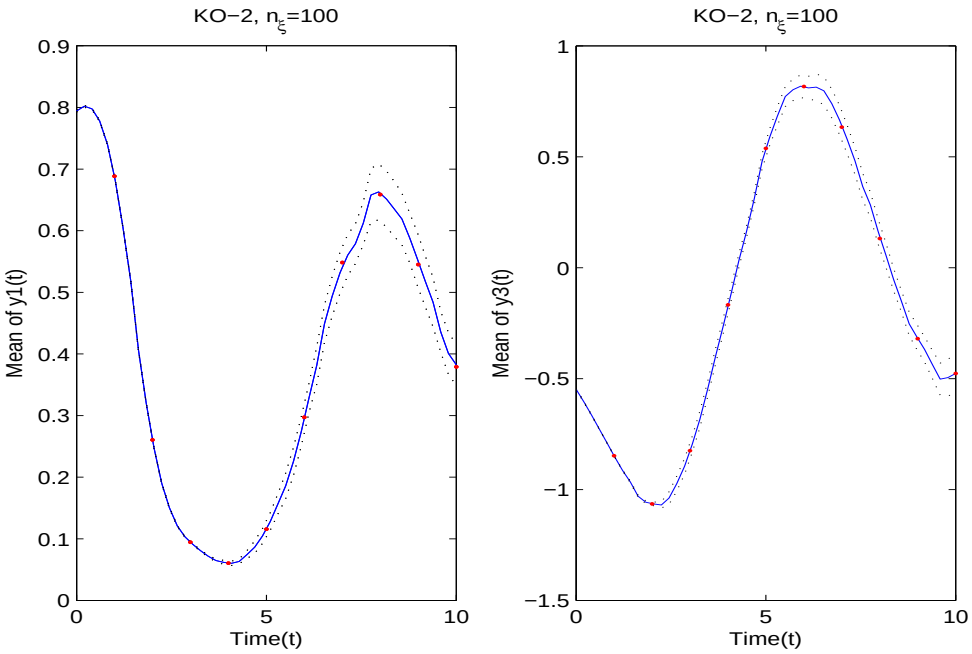


Fig. 5. The predictive mean curve in time by the nonseparable model. The blue line is the predictive mean curve of time; the dash lines are corresponding 95% confidence intervals; the red dots are the means of the computer code outputs.

4.3 Flow through porous media example

We use this example to show the effectiveness of our method in modeling large computer code outputs. The proposed Gaussian process surrogate model was applied to a petroleum reservoir simulation of a much larger data size. The object is a two-dimensional, single phase, steady flow through a random permeability field. The spatial domain $\mathscr{X}_s = [0,1]^2$, representing an idealized oil reservoir. The pressure p and the velocity fields of the flow $\mathbf{u}$ are of key interests and they are connected via Darcy law: $\mathbf{u} = -\mathbf{K}\nabla p$ in $\mathscr{X}_s$, where $\mathbf{K}$ is the permeability tensor. The pressure p satisfies $-\nabla \cdot (\mathbf{K}\nabla p) = f$ in $\mathscr{X}_s$, where f may be used to model the injection/production wells. One specific choice of f is

$$f(\mathbf{s}) = \begin{cases} -r, \text{ if} |s_i - w/2| < w/2, \text{ for } i = 1,2, \\ r, \quad \text{if} |s_i - 1 + w/2| < w/2, \text{ for } i = 1,2, \\ 0, \quad \text{otherwise.} \end{cases}$$

There are two wells at the lower-left and upper-right corners of the spatial domain, with r and w specifying well rates and well sizes respectively ($r = 10$ and w=1/8 in this case). After specifying the permeability tensor $\mathbf{K}$ and imposing certain boundary conditions, the pressure p and the velocities $\mathbf{u}$ can be solved numerically.

Specifically, $\mathbf{K}$ is assumed to be isotropic $K_{ij} = K\delta_{ij}$ and modeled by a log-Gaussian process: $K(\mathbf{s}) = \exp(G(\mathbf{s}))$, where δ is the Kronecker delta function and $G(\cdot) \sim \mathcal{GP}(m_G, \mathcal{C}_G(\cdot,\cdot))$. The covariance function $\mathcal{C}_G(\cdot,\cdot)$ is the separable exponential covariance. The truncated Karhunen-Loève expansion on $G(\cdot)$ gives its finite dimension representation

$$K(\boldsymbol{\xi}; \mathbf{s}) = \exp\left(m_G + \sum_{k=1}^{p} w_k \psi_k(\mathbf{s})\right),$$

where w_k are uncorrelated standard Gaussian random variables and $\psi_k(\mathbf{s})$ are eigenfunctions of the exponential covariance function $\mathcal{C}_G(\cdot,\cdot)$. Then uniform variables $\xi_k = \Phi(w_k) \sim U([0,1])$ are treated as input variables. More details of this example can be found in [7]. In this example, we truncate $G(\cdot)$ after $p = 50$ terms.

The training set is on a $24 \times 32 \times 32$ input-spatial grid, so the total training data has 24576×3 observations. The nonseparable model considered here is

$$\rho(\mathbf{x}, \mathbf{x}') = \left(|d_u|^{2\alpha} + 1\right)^{-1} \exp\left(-\frac{\sqrt{|h_1|^2/c_1^2 + |h_2|^2/c_2^2}}{(|d_u|^{2\alpha} + 1)^{\beta/2}}\right), \tag{17}$$

where $d_u = \sqrt{\sum_{i=1}^{k_\xi} \frac{u_i^2}{\phi_i^2}}, u_i = |\xi_i - \xi_i'|, h_j = |s_j - s_j'|$ for $i = 1, \ldots, k_\xi$ and $j = 1, 2$. A small nugget effect $\tau^2 = 10^{-6}$ was fixed during the parameter estimation step. We applied the FSA-Block approach with 200 knots selected by LHS and 100 blocks created by K-means algorithm to the nonseparable model, making computations of model implementation feasible. After the parameter estimation step, the posterior means of model parameters were plugged in (12) to make predictions at a finer $100 \times 64 \times 64$ input-spatial grid. We considered the separable model in (17) with $\beta = 0$ and the squared exponential model for comparisons. Although this problem has a high-dimensional input space (50 input variables), we experimented using the first $k_\xi = 3$ input variables (corresponding to first 3 leading terms in the K-L expansion of $G(\cdot)$). We also experimented using a larger number of input variables ($k_\xi = 5$), but there are only slight differences for the prediction performances. So we focused on the smaller dimension $k_\xi = 3$ case.

Table 3 gives the parameter estimation results of different models as well as

MSPEs. For the nonseparable model, the posterior mean estimates of β is close to 1, implying modeling the interaction effect may be beneficial. Figure 6 shows the predictive mean response surfaces in spatial domain by the nonseparable model, using the first 3 input variables. We can see the spatial patterns of the predictive mean response surfaces are very similar to the Monte Carlo estimates using the computer model outputs, which can be viewed as the true values. Figure 7 gives the error bars of the predictive response surfaces in spatial domain.

Table 3
Posterior means of model parameters and prediction results of each output component.

$n_\xi = 24$	Nonsep	Sep ($\tau^2 = 0.01$)	Sqexp ($\tau^2 = 0.01$)
ϕ_1	5.564	9.689	25.722
ϕ_2	95.443	11.111	34.041
ϕ_3	142.046	47.540	76.473
c_1	1.265	0.686	0.724
c_2	1.258	0.222	0.420
α	0.083	0.0938	–
β	0.999	0	–
B_{10}	−0.478	0.1038	−0.177
B_{20}	−0.550	−0.0033	−0.498
B_{30}	−0.026	0.0002	0.003
Σ_{11}	0.0086	0.0099	10.758
Σ_{12}	0.0015	0.0016	2.946
Σ_{13}	2.8×10^{-5}	2.9×10^{-5}	0.0164
Σ_{22}	0.0084	0.0036	8.491
Σ_{23}	3.9×10^{-5}	2.6×10^{-5}	0.056
Σ_{33}	0.0001	0.0001	0.139
MSPE_{sp}	$(0.0035, 0.0036, 0.1030)$	$(0.0029, 0.0028, 0.1030)$	$(0.0033, 0.0035, 0.1030)$

For the separable models, we also used $k_\xi = 3$ in order to do fair comparisons with the nonseparable model. When we fixed a small nugget $\tau^2 = 10^{-6}$ for the input and spatial covariance parts, the estimates of parameters of the two separable models had relatively large variances, due to the numerical instability. Especially for the squared exponential model assuming infinite smoothness in input space, it yielded estimates of mean parameters with very large variances (over 10^6) and can not obtain reasonable prediction results. Then we increased $\tau^2 = 0.01$ suggested in [7] for these separable models and the results are shown in Table 3. The prediction results of the separable model with $\beta = 0$ are better than those of the nonseparable model, this may be because a covariance function separable in input and space was used in the log-normal process to model the permeability field K, which might lead to certain level of separability in input and space for the outputs. Besides, the usage of full correlation model without any approximations for the separable model may also attribute to better prediction results. However, we remark that although the squared

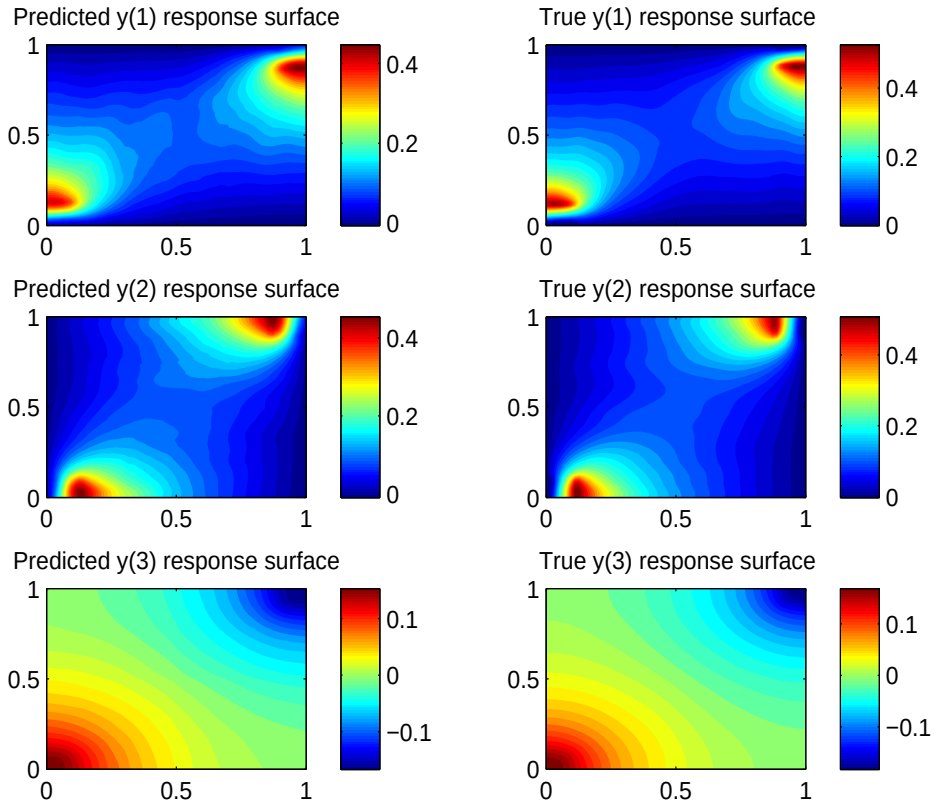
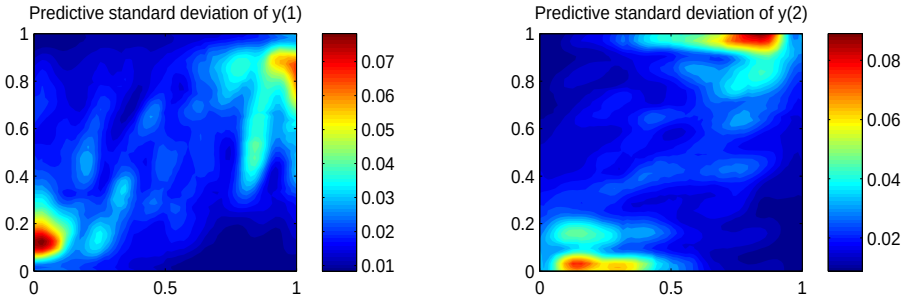


Fig. 6. Predictive mean surfaces by the nonseparable model versus the Monte Carlo estimates based on 24 input points. Upper panels: velocity in y-direction u_y; middle panels: velocity in x-direction u_x; lower panels: pressure p.

exponential separable model produced reasonable prediction results, the estimates of variance are in general fairly large compared with the data scales. In contrast, the nonseparable model with the FSA-Block approximation does not suffer much with the numerical stability problem, this may be because the correlations cross data blocks are approximated well by that of the predictive process of reduced rank.

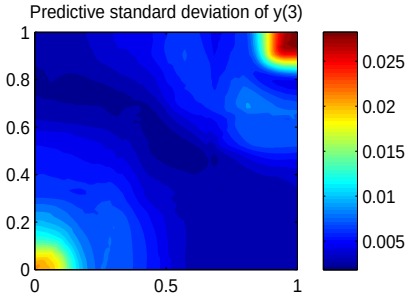
5 The regenerator of a carbon capture unit

In this Section, we applied the nonseparable Gaussian process surrogate model to a real example from the regenerator of a carbon capture unit. A carbon capture unit provides an alternative solution for limiting the carbon dioxide (CO_2) emissions. All carbon capture units contain an absorber device and a regenerator device. The solid sorbent particles capable of reacting with the CO_2 gas are looped through these two devices. In the absorber, the exhaust



(a) Error bar of u_y

(b) Error bar of u_x



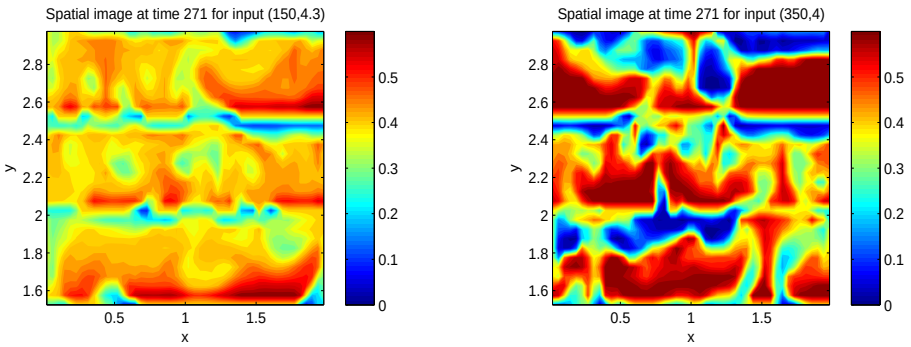
(c) Error bar of the pressure p

Fig. 7. Predictive standard deviations by the nonseparable model.

flue gas from a power plant reacts with the solid sorbent particles and its CO_2 component is trapped. Then after further processing steps, the cleaned exhaust flue gas is released into the atmosphere and the depleted sorbent particles are transferred to the regenerator. In the regenerator, the reverse chemical reaction is done to release CO_2 from the depleted sorbent particles for further processing (i.e. liquefaction and sequestration for long-term storage) and the regenerated sorbent particles are recycled back to the absorber. Since the bulk of the energy penalty is related to the regenerator, its efficiency is of crucial interests.

Recently, [27] developed a computational fluid dynamics (CFD)-based model for the fluid dynamics of the regenerator. The flow of sorbent particles is characterized by the density of solid volume fraction, which is sensitive to the system operating conditions such as the particle diameter d_p (with unit micro meters, μm) and the scaled velocity v_g/u_{mf} of gas injected at the bottom inlet (dimensionless; details see [28]), denoted by $(d_p, v_g/u_{mf})$. Figure 8 shows the solid volume fractions for 2 input points at a given time. It is clear that both the value and the spatial pattern of the solid volume fractions can change drastically for different input points. If the intermediate solid volume fraction values in $[0.2, 0.4]$ are more likely to result in better efficiency of the regenerator device, then the input point $(150, 4.3)$ is superior to $(350, 4)$ according to figure 8, since it has a larger proportion of intermediate solid volume fractions. Specifying the operation conditions in favor of a certain range of solid volume fractions needs a number of computer simulations. The CFD-based simula-

tions are very time-consuming, taking days to complete one simulation under parallelled computing system. So it is challenging to run a large number of simulations to study the behaviors of the sorbent distribution under different operating conditions. Therefore, the Gaussian process model is used instead as an effective tool for the uncertainty quantification purpose [5].



(a) $d_p = 150$ and $v_g/u_{mf} = 4.3$ (b) $d_p = 350$ and $v_g/u_{mf} = 4$

Fig. 8. Images of the solid volume fractions at time 271 for 2 input points.

The computer model outputs are the solid volume fractions over a time period, ranging from 0 to 0.6083. We focused on solid volume fraction values in $(0.1, 0.6]$, since without much knowledge of the reaction kinetics, we expect that the intermediate values are more likely to result in better performance of the regenerator [5]. We focused on the discrete distribution of the solid volume fractions created using 5 equal length bins $(0.1, 0.2], \cdots, (0.5, 0.6]$, aiming to check effects of the distribution of the solid volume fractions on the reaction kinetics. Denote the response vector by $\mathbf{f}(\boldsymbol{\xi}, t) = (\pi_1, \pi_2, \cdots, \pi_5)^T$, it was treated as a function of 2 input variables (d_p and v_g/u_{mf}), as well as time t.

The training data set was on a 46×101 input-time grid, and we randomly held out 4 input points on the same time grid for evaluating model performances. The Gaussian process regression model (1) with constant means was fitted to this data and the nonseparable correlation function was the same as in the Section 4.2. The FSA approach with 15 blocks and 300 knots were applied to the full model to speed up computations. The results of the same two separable models as in previous studies were again included for comparisons.

The parameter estimation and prediction results are summarized in Table 4. The posterior mean estimate of the input-time interaction parameter β of the nonseparable model is very close to 1, indicating the existence of the interaction effect. The nonseparable model and the separable model (3) with $\beta = 0$ have close estimates of the smoothness parameter α, which is not surprising since it is related to the process properties in time dimension. Besides, all three correlation models produced a large estimate of the range parameter in d_p dimension and a small estimate of the range parameter in v_g/u_{mf}

dimension, indicating that the computer code outputs are more sensitive to the v_g/u_{mf} variable. In terms of prediction, the nonseparable model with the

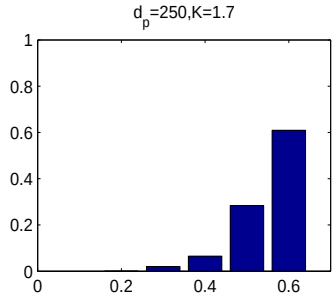
Table 4
Posterior means of model parameters and the overall MSPEs.

$n_\xi * n_t = 4646$	a	c_1	c_2	α	β	B_{10}	B_{20}	B_{30}	B_{40}	B_{50}
Nonsep	15.131	92.993	0.943	0.638	0.993	0.095	0.120	0.138	0.178	0.236
Sep	6.263	485.10	6.791	0.586	0	0.104	0.134	0.087	0.110	0.252
Sqexp	1.147	499.280	7.349	–	–	0.105	0.136	0.086	0.106	0.257
	Σ_{11}	Σ_{12}	Σ_{13}	Σ_{14}	Σ_{15}	Σ_{22}	Σ_{23}	Σ_{24}	Σ_{25}	Σ_{33}
Nonsep	0.0020	−0.0003	−0.0001	0.0048	−0.0054	0.0038	−0.0001	0.0199	−0.0224	0.0037
Sep	0.0088	−0.0037	−0.0010	−0.0002	0	0.0119	−0.0061	−0.0008	0.0001	0.0172
Sqexp	0.0209	−0.0081	−0.0022	−0.0009	−0.0001	0.0277	−0.0132	−0.0021	−0.0004	0.0392
	Σ_{34}	Σ_{35}	Σ_{44}	Σ_{45}	Σ_{55}	MSPE				
Nonsep	0.0120	−0.0151	0.3009	−0.3310	0.3677	0.0011				
Sep	−0.0086	−0.0011	0.0233	−0.0140	0.0191	0.0017				
Sqexp	−0.0194	−0.0030	0.0615	−0.0401	0.0548	0.0018				

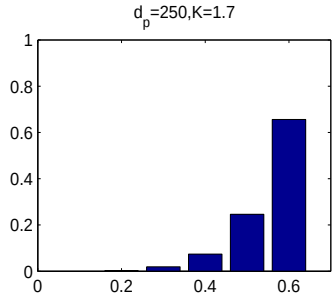
FSA approximation approach outperforms the two separable models. Specifically, the nonseparable model has the same prediction performance as the separable models for π_1, π_2, π_3, but it has smaller MSPEs $(0.0022, 0.0013)$ for π_4 and π_5, compared with $(0.0032, 0.0033)$ by the separable model (3) with $\beta = 0$ and $(0.0034, 0.0036)$ by the squared exponential model. Figure 9 shows the predictive probabilities by the nonseparable model and the real computer code results, for the hold-out set of 4 input points. We can observe that the predictive probabilities are close to the real data results in general. Figure 10 shows the predictive mean input surfaces of 5 probabilities by the nonseparable model, where the predictions were made on a dense 70×70 input grid at time points $t = 270, 271, \ldots, 280$. Based on these predictive mean surfaces of the input space, particular input regions can be found to improve the efficiency of the regenerator unit. For example, if the intermediate solid volume fractions in $(0.3, 0.4]$ would result in better efficiency of the regenerator unit, then we may pay special attention to the high probability area of the predictive mean surface of π_3. So in this case, from figure 10, $d_p \in (150\mu m, 250\mu m)$ and $v_g/u_{mf} \in (4, 8)$ would be good choices of operating conditions. Other predictive mean surfaces of probabilities may also be useful for specifying the values of input variables that can result in good efficiency of the regenerator.

6 Concluding remarks

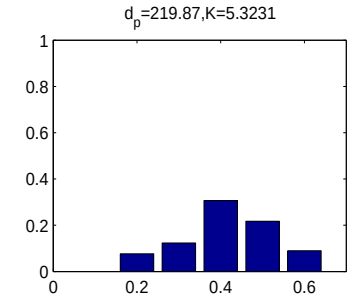
In this paper, we extended the commonly used separable covariance Gaussian process surrogate models by using a more flexible nonseparable autocovariance function, which includes separable model as a special case. The



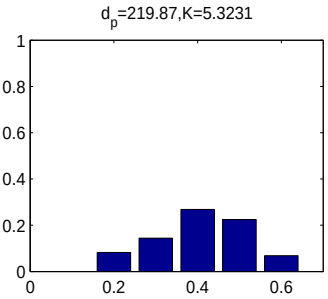
(a) Solid fraction distribution



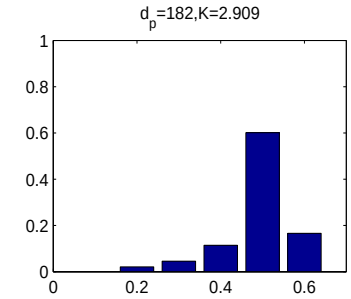
(b) Solid fraction distribution



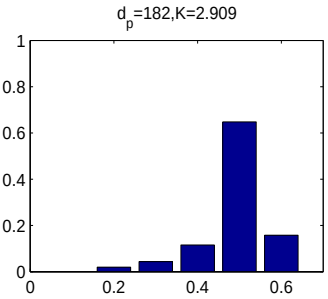
(c) Solid fraction distribution



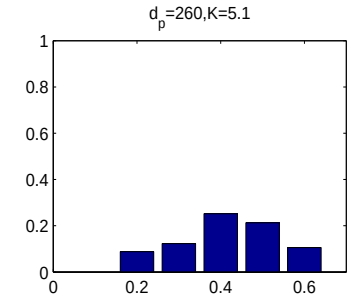
(d) Solid fraction distribution



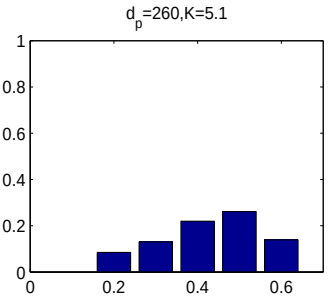
(e) Solid fraction distribution



(f) Solid fraction distribution



(g) Solid fraction distribution



(h) Solid fraction distribution

Fig. 9. The left panels are predictive distributions of solid volume fractions under different combinations of d_p and $K = v_g/u_{mf}$; the right panels are corresponding computer code results.

nonseparable model has the advantage of not only modeling dependence within each dimension in input, space and time but also interactions in dependence among them. This model can relax the restrictions imposed by separable mod-

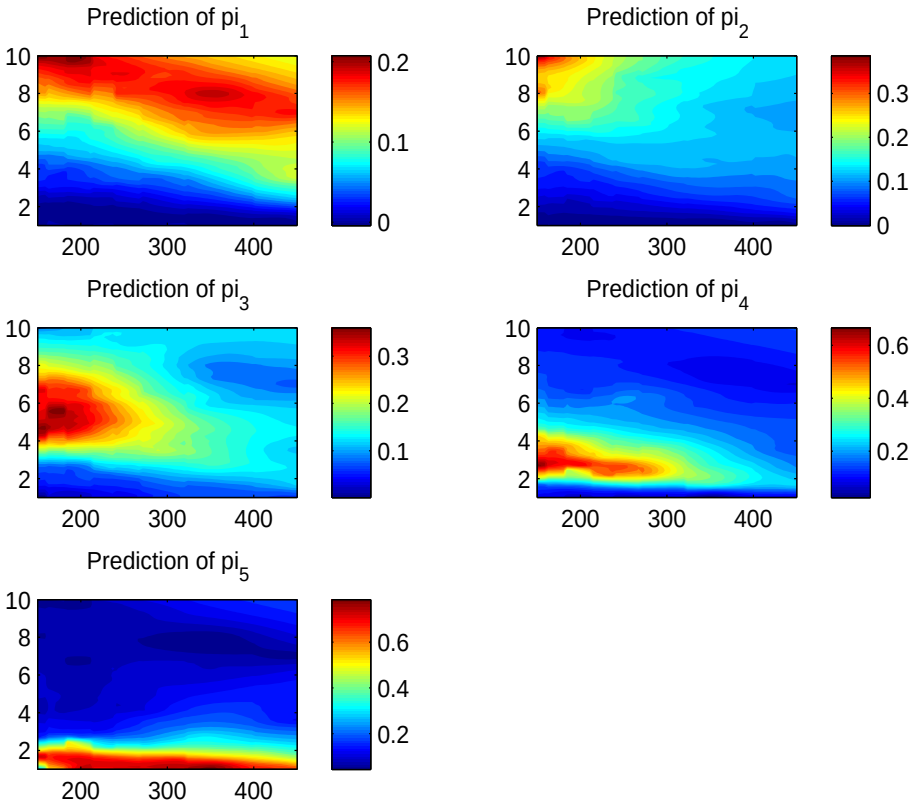


Fig. 10. Predictive mean input surfaces of 5 probabilities by the nonseparable model.

els on the conditional and marginal properties of a Gaussian process [10], and hence has broader applications especially in cases where the assumption of separability is problematic.

We also introduced a new computational method, referred to as the full-scale approximation with block modulating function approach (FSA-Block), to achieve fast computations when applying the nonseparable model to large computer code outputs. We illustrated the effectiveness of the nonseparable model with the FSA-Block approach through various simulation examples and a real data set from the computer code of the regenerator device of a carbon capture system.

The FSA-Block approach introduced in this paper does not depend on the separable structure of the covariance matrix and hence can be used flexibly in various ways. For example, when the covariance function is partially separable and there are lots of observations in certain nonseparable dimensions, we can apply the FSA-Block approach only to the nonseparable part to facilitate computations. Besides, for large computer code outputs, it can enjoy better numerical stability than using the separable full covariance models, since the

cross-correlations within each output are approximated by the results of the predictive process part of reduced dimension. Although we focused on the stationary auto-covariance functions in this paper, this computational method can also apply to nonstationary covariance functions such as the nonstationary model in [29], where spatial regions have different dependence structures.

In this paper, we considered a separable cross-covariance structure among different outputs. A nature extension is to build a totally nonseparable emulator with both a nonseparable auto-covariance function and a nonseparable cross-covariance. The Linear Model of Coregionalization (LMC) [30,31,19] and the cross-covariance functions based on latent dimensions [13] may be applied to relax the separable cross-covariance assumption. Investigations on new computational methods are also problems of interest to facilitate the more demanding computational needs of the totally nonseparable model.

Acknowledgements: G. Lin would like to acknowledge the support of 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 is also partially supported by NSF Grant DMS-1115887. The research of H. Sang is partially supported by NSF Grant CNS-1343155.

References

[1] C. Currin, T. Mitchell, M. Morris, D. Ylvisaker, Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, Journal of the American Statistical Association 86 (416) (1991) 953–963.

[2] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, et al., Design and analysis of computer experiments, Statistical science 4 (4) (1989) 409–423.

[3] J. Oakley, A. O'Hagan, Bayesian inference for the uncertainty distribution of computer model outputs, Biometrika 89 (4) (2002) 769–784.

[4] S. Conti, A. OHagan, Bayesian emulation of complex multi-output and dynamic computer models, Journal of statistical planning and inference 140 (3) (2010) 640–651.

[5] B. Konomi, G. Karagiannis, A. Sarkar, X. Sun, G. Lin, Bayesian treed multivariate gaussian process with adaptive design: Application to a carbon capture unit, Technometrics (just-accepted).

[6] R. B. Gramacy, H. K. Lee, Bayesian treed gaussian process models with an application to computer modeling, Journal of the American Statistical Association 103 (483).

[7] I. Bilionis, N. Zabaras, B. A. Konomi, G. Lin, Multi-output separable gaussian process: Towards an efficient, fully bayesian paradigm for uncertainty quantification., J. Comput. Physics 241 (2013) 212–239.

[8] N. Cressie, H.-C. Huang, Classes of nonseparable, spatio-temporal stationary covariance functions, J. Amer. Statist. Assoc. 94 (1999) 1330–1340.

[9] M. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63 (3) (2001) 425–464.

[10] J. Rougier, A representation theorem for stochastic processes with separable covariance functions.

[11] T. Gneiting, Nonseparable, stationary covariance functions for space-time data, J. Amer. Statist. Assoc. 97 (2002) 590–600.

[12] M. Stein, Space-time covariance functions, J. Amer. Statist. Assoc. 100 (2005) 310–321.

[13] T. V. Apanasovich, M. G. Genton, Cross-covariance functions for multivariate random fields based on latent dimensions, Biometrika 97 (1) (2010) 15–30.

[14] H. Sang, J. Huang, A full scale approximation of covariance functions for large spatial data sets, Journal of the Royal Statistical Society Series B 74 (1) (2012) 111–132.

[15] H. Sang, M. Jun, J. Huang, Covariance approximation for large multivariate spatial data sets with an application to multiple climate model errors, The Annals of Applied Statistics 5 (4) (2011) 2519–2548.

[16] S. Banerjee, A. Gelfand, A. Finley, H. Sang, Gaussian predictive process models for large spatial data sets, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70 (4) (2008) 825–848.

[17] C. Kaufman, M. Schervish, D. Nychka, Covariance tapering for likelihood-based estimation in large spatial data sets, J. Amer. Statist. Assoc. 103 (484) (2008) 1545–1555.

[18] K. V. Mardia, C. R. Goodall, Spatial-temporal analysis of multivariate environmental monitoring data, Multivariate Environmental Statistics 6 (347-385) (1993) 76.

[19] T. E. Fricker, J. E. Oakley, N. M. Urban, Multivariate gaussian process emulators with nonseparable covariance structures, Technometrics 55 (1) (2013) 47–56.

[20] A. Finley, H. Sang, S. Banerjee, A. Gelfand, Improving the performance of predictive process modeling for large datasets, Computational statistics & data analysis 53 (8) (2009) 2873–2884.

[21] N. Cressie, G. Johannesson, Fixed rank kriging for very large spatial data sets, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70 (1) (2008) 209–226.

[22] R. Furrer, M. Genton, D. Nychka, Covariance tapering for interpolation of large spatial datasets, Journal of Computational and Graphical Statistics 15 (3) (2006) 502–523.

[23] K. Atkinson, The numerical solution of integral equations of the second kind, Vol. 4, Cambridge university press, 1997.

[24] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian data analysis, CRC press, 2013.

[25] I. Bilionis, N. Zabaras, Multi-output local gaussian process regression: Applications to uncertainty quantification, Journal of Computational Physics 231 (17) (2012) 5718–5746.

[26] X. Wan, G. E. Karniadakis, An adaptive multi-element generalized polynomial chaos method for stochastic differential equations, Journal of Computational Physics 209 (2) (2005) 617–642.

[27] A. Sarkar, W. Pan, D. Suh, E. D. Huckaby, X. Sun, Multiphase flow simulations of a moving fluidized bed regenerator in a carbon capture unit, Powder Technology.

[28] D. Gidaspow, Multiphase flow and fluidization: continuum and kinetic theory descriptions, Academic press, 1994.

[29] C. J. Paciorek, M. J. Schervish, Spatial modelling using a new class of nonstationary covariance functions, Environmetrics 17 (5) (2006) 483–506.

[30] A. E. Gelfand, A. M. Schmidt, S. Banerjee, C. Sirmans, Nonstationary multivariate process modeling through spatially varying coregionalization, Test 13 (2) (2004) 263–312.

[31] A. E. Gelfand, S. Banerjee, D. Gamerman, Spatial process modelling for univariate and multivariate dynamic spatial data, Environmetrics 16 (5) (2005) 465–479.