Journal of Computational Physics 300 (2015) 623–642

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

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

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

[a] Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA
[b] Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45221-0025, USA
[c] Department of Mathematics, Purdue University, West Lafayete, IN 47907-2067, USA
[d] Department of Mathematics & School of Mechanical Engineering, Purdue University, West Lafayete, IN 47907-2067, USA

ARTICLE INFO

Article history:
Received 17 January 2015
Received in revised form 11 June 2015
Accepted 2 August 2015
Available online 10 August 2015

Keywords:
Gaussian process emulator
Uncertainty quantification
Nonseparable covariance function
Full-scale approximation
Carbon capture unit

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 is an effective and accurate covariance approximation method to reduce computations for Gaussian process models, which applies to both nonseparable and partially separable covariance models. We illustrate the effectiveness of our method through simulation studies and compare it with emulators with separable covariances. We also apply our method to a real computer code of the carbon capture system.

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. 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. Oakley and O'Hagan [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 used auto-covariance function is usually stationary and

* Corresponding authors.
E-mail addresses: huiyan@stat.tamu.edu (H. Sang), guanglin@purdue.edu (G. Lin).

http://dx.doi.org/10.1016/j.jcp.2015.08.006
0021-9991/

separable in each input dimension; the cross-covariance among outputs is also assumed to be separable from dependence in other dimensions 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 [4] and [6] to a Bayesian Treed multivariate Gaussian process model, accounting for both the nonstationarity and the multivariate features of the data. Based on separable covariance models and an adaptive algorithm for partitioning the stochastic space, the multivariate local Gaussian process emulator [7] also provides an effective tool to model large nonstationary datasets.

The assumption of separability allows fitting Gaussian process model in each input dimension separately. It leads to a separable covariance structure of covariance function and hence alleviates the computational demand by reducing the dimension of the covariance matrices to be inverted. One such example is in [8], where the authors 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. [9] pointed out that if a stationary spatio-temporal covariance function is separable, then the temporal dependence structure cannot vary spatially and the spatial dependence structure cannot 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 [10]. 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 [11].

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 [9,12,13]. 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 [14], where the authors extended methods in [12] to propose a nonseparable cross-covariance model for multivariate random fields. Motivated by this work in spatial statistics, we develop a flexible class of nonseparable auto-covariances for uncertainty quantification of computer models. In particular, this class of models includes separable models as special cases.

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. Large sample size n typically makes the computations for the Gaussian process emulator prohibitive, unless some particular structures of the covariance functions are assumed, e.g. the separability. To overcome the computational bottleneck, we introduce the Full-Scale approximation (FSA) approach to reduce computations [15,16], which applies to both separable and nonseparable covariance structure. The FSA approach combines the merits of several popular approaches such as a reduced rank Gaussian process [17] and sparse covariance approximation [18] 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 computational cost significantly. [15] showed the effectiveness of using this method for model fitting and prediction in the spatial context. In this paper, we tailor this state-of-the-art computational tool and investigate its performance for the purpose of quantifying uncertainties of computer code outputs.

The major contributions of this paper have two folds: first we propose a flexible new class of nonseparable auto-covariance functions for each computer output to model the interaction effect among input, space and time. This class of models relaxes the separability assumption that is typically made for Gaussian process emulators and provides a more flexible and general tool to describe dependence for computer model outputs. Second, we introduce the FSA approach in the uncertainty quantification context 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.