Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utch20

Additive Gaussian Process for Computer Models With Qualitative and Quantitative Factors

X. Deng, C. Devon Lin, K.-W. Liu & R. K. Rowe

To cite this article: X. Deng, C. Devon Lin, K.-W. Liu & R. K. Rowe (2017) Additive Gaussian Process for Computer Models With Qualitative and Quantitative Factors, Technometrics, 59:3, 283-292, DOI: 10.1080/00401706.2016.1211554

To link to this article: http://dx.doi.org/10.1080/00401706.2016.1211554

Accepted author version posted online: 27 Sep 2016.
Published online: 27 Sep 2016.

TECHNOMETRICS
2017, VOL. 59, NO. 3, 283–292
http://dx.doi.org/10.1080/00401706.2016.1211554

Taylor & Francis
Taylor & Francis Group

Additive Gaussian Process for Computer Models With Qualitative and Quantitative Factors

X. Deng[a], C. Devon Lin[b], K.-W. Liu[c], and R. K. Rowe[d]

[a]Department of Statistics, Virginia Tech, Blacksburg, VA; [b]Department of Mathematics and Statistics, Queen's University, Kingston, Canada; [c]School of Civil Engineering, Southwest Jiaotong University, Emeishan, Leshan, Sichuan, China; [d]Department of Civil Engineering, Queen's University, Kingston, Canada

ABSTRACT

Computer experiments with qualitative and quantitative factors occur frequently in various applications in science and engineering. Analysis of such experiments is not yet completely resolved. In this work, we propose an additive Gaussian process model for computer experiments with qualitative and quantitative factors. The proposed method considers an additive correlation structure for qualitative factors, and assumes that the correlation function for each qualitative factor and the correlation function of quantitative factors are multiplicative. It inherits the flexibility of unrestrictive correlation structure for qualitative factors by using the hypersphere decomposition, embracing more flexibility in modeling the complex systems of computer experiments. The merits of the proposed method are illustrated by several numerical examples and a real data application. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received May 2015
Revised May 2016

KEYWORDS
Additive model; Computer experiment; Fractional factorial design; Hypersphere decomposition; Kriging

1. Introduction

Computer experiments that use complex computer models to study real systems have attracted wide attentions in many science and engineering applications. Because computer models are typically expensive in terms of computational time, an emulator (surrogate model) is often needed (Sacks et al. 1989). An emulator can be used as a stand-in for a computer model and it is cornerstone for further analysis such as optimization, sensitivity analysis, and uncertainty quantification. The bulk of the work on building emulators assume that the input factors in computer models are quantitative (Santner, Williams, and Notz 2003) and those methods are not directly applicable to computer models that contain both qualitative and quantitative factors.

Computer models with both qualitative and quantitative factors arise in various applications. For example, this work is motivated by a real computer experiment of finite element modeling of full-scale embankment over soft soils considering reinforcement stiffness and column length as qualitative factors (Rowe and Liu 2015; Liu and Rowe 2015). Examples from the literature are: the computational fluid-dynamics simulation for studying data center often contains qualitative factors such as "hot air return vent location" and "power unit type" (Qian, Wu, and Wu 2008); the agent-based network modeling in epidemiology study (Bhuiyan et al. 2014) often constructs the network with shuffle type as a qualitative factor; for investigating wear mechanisms of total knee replacements in bioengineering, Han et al. (2009) presented the knee models with qualitative factors such as "prosthesis design" and "force pattern."

Gaussian process modeling is a common technique for building emulators of computer models. For computer models with both qualitative and quantitative factors, it is not trivial to build emulators using Gaussian process modeling. This is because different level combinations of qualitative factors may not have specific distance measurement. To accommodate both qualitative and quantitative factors into Gaussian process modeling, there are two key challenges: how to construct a proper covariance structure for the qualitative factors, and how to specify the relationship between the correlation function for qualitative factors and the correlation function for quantitative factors.

For constructing correlation function for qualitative factors, one approach is to consider a restrictive correlation function (McMillian et al. 1999; Joseph and Delaney 2007). Such an approach can simplify the computational complexity for the model estimation. However, restrictive correlation functions lack flexibility to quantify general correlation structure of qualitative factors. Another approach is to construct an unrestrictive correlation structure for qualitative factors. Qian, Wu, and Wu (2008) developed a general framework of constructing an unrestrictive correlation structure for qualitative factors in Gaussian process models. Their method adopts semidefinite optimization in estimation to ensure the positive-definiteness of the correlation structure. To overcome the complicated estimation procedure in Qian, Wu, and Wu (2008), Zhou et al. (2011) developed a hypersphere parameterization (Pinheiro and Bates 1996; Rebonato and Jackel 1999) method to model the unrestrictive correlation structure, which can automatically guarantee the positive-definiteness of the correlation structure. Recently, Zhang and Notz (2015) considered an indicator function approach to model the correlation function for the qualitative factors, which is in a similar spirit as that in Zhou

CONTACT C. Devon Lin cdlin@mast.queensu.ca Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada.
Color versions of one or more of the figures in this article are available online at www.tandfonline.com/r/TECH.
Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH.

et al. (2011). The third direction is to adopt the hierarchical Bayesian Gaussian process models (Swiler et al. 2014) to accommodate both qualitative and quantitative factors. For constructing correlation functions with both qualitative and quantitative factors, the majority of existing work assumes the multiplication between the correlation function of qualitative factors and the correlation function of quantitative factors. The potential drawback of using multiplication in correlation function is that multiplication could improperly characterize the effect of the correlation from qualitative factors. For example, when the correlation function of one qualitative factor is small or close to zero, it would result in the overall correlation function to be very small.

In this work, we propose an additive Gaussian process model for computer experiments with qualitative and quantitative factors. The key idea of the proposed method is to consider an additive correlation structure for qualitative factors, and to assume that the correlation function for each qualitative factor and the correlation function of quantitative factors are multiplicative. Such a formulation also allows different covariance structures for quantitative factors with respect to different qualitative factors. It inherits the flexibility of an unrestrictive correlation structure for qualitative factors by using the hypersphere decomposition.

The remainder of this article is organized as follows. Section 2 provides notation and a general background. Section 3 details the proposed additive Gaussian process model, along with the estimation, prediction and interpretation. Section 4 presents numerical examples to illustrate the effectiveness of the proposed method. In Section 5, the proposed method is applied in a real computer experiment of finite element modeling for full-scale embankment over soft soils. We conclude this work with brief summary and discussion in Section 6.

2. Notation and Background

This section provides notation and background for later development. We review Gaussian process models with only quantitative factors, and the hypersphere parameterization to construct the correlation structure for qualitative factors in Zhou et al. (2011). Throughout, we consider computer experiments with p quantitative factors $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}} \in \mathcal{R}^p$ and q qualitative factors $\mathbf{z} = (z_1, \ldots, z_q)^{\mathrm{T}}$ with the jth qualitative factor having m_j levels, $j = 1, \ldots, q$, and the corresponding output is denoted by Y. Suppose that the data obtained from the computer experiment are $(y_i, \mathbf{w}_i)$, $i = 1, \ldots, n$, where $\mathbf{w}_i = (\mathbf{x}_i, \mathbf{z}_i)$.

2.1 Gaussian Process Model With Quantitative Factors

With quantitative factors as inputs in the model, the key idea of Gaussian processes is to consider the outputs from computer experiments as a realization of a Gaussian process. The correlation of responses at two data points is determined by a correlation function of inputs, such as Matern correlation function (Matern 1986). To model the relationship between output Y and inputs $\mathbf{x}$, one stationary Gaussian process model, known as ordinary Kriging model, assumes,

$$Y(\mathbf{x}) = \mu + Z(\mathbf{x}), \tag{1}$$

where μ is the constant mean, $Z(\mathbf{x})$ is a Gaussian process with mean zero and the covariance function $\phi(\cdot) = \sigma^2 R(\cdot|\boldsymbol{\theta})$. Here σ^2 is the variance and $R(\cdot|\boldsymbol{\theta})$ is referred to as the correlation function with the correlation parameter vector $\boldsymbol{\theta}$. A popular choice of the correlation function is the Gaussian correlation function

$$R(\mathbf{x}_i, \mathbf{x}_j|\boldsymbol{\theta}) = \exp\left\{-\sum_{k=1}^{p} \theta_k (x_{ik} - x_{jk})^2\right\}, \tag{2}$$

for any two inputs $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ and $\mathbf{x}_j = (x_{j1}, \ldots, x_{jp})^{\mathrm{T}}$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^{\mathrm{T}}$ and $\theta_k \geq 0, k = 1, \ldots, p$. Consider the inputs of quantitative factors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and the corresponding responses $y_1, \ldots, y_n$. For the ordinary Kriging, the best linear predictor at an input $\mathbf{x}$ is given by

$$\hat{y}(\mathbf{x}) = \hat{\mu} + \mathbf{r}^T(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}), \tag{3}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$, $\mathbf{R}$ is an $n \times n$ correlation matrix with the (i, j)th element $R(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{r}(\mathbf{x}) = (R(\mathbf{x}, \mathbf{x}_1), \ldots, R(\mathbf{x}, \mathbf{x}_n))^{\mathrm{T}}$, $\mathbf{1}$ is an n-dimensional column vector of all 1's, and $\hat{\mu} = (\mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{1})^{-1}\mathbf{1}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{y}$ is the estimate of μ. Note that the predictor in (3) involves with unknown correlation parameters in the correlation function. To estimate these parameters, a common approach is the maximum likelihood estimation method (Santner, Williams, and Notz 2003; Fang, Li, and Sudjianto 2005).

2.2 Unrestrictive Correlation for Qualitative Factors

When computer experiments have qualitative factors as inputs, the usual correlation function in (2) does not work properly because the Euclidian distance cannot be used for different level combinations of qualitative factors. To address this issue, Zhou et al. (2011) introduced a hypersphere parameterization to quantify the correlations of the qualitative factors. Recall that the jth qualitative factor z_j has m_j levels. For $j = 1, \ldots, q$, let $\mathbf{T}_j = (\tau_{r,s}^{(j)})$ be an $m_j \times m_j$ correlation matrix of the m_j levels of the qualitative factor z_j. The key idea of Zhou et al. (2011) is to apply the hypersphere decomposition to model $\mathbf{T}_j$ such that $\mathbf{T}_j$ is a positive-definite matrix with unit diagonal elements. Their approach consists of two steps. Step 1 is to find a lower triangular matrix with strictly positive diagonal entries $\mathbf{L}_j = (l_{r,s}^{(j)})$ through a Cholesky-type decomposition, that is, $\mathbf{T}_j = \mathbf{L}_j\mathbf{L}_j^T$ for $j = 1, \ldots, q$. In Step 2, each row vector $(l_{r,1}^{(j)}, \ldots, l_{r,r}^{(j)})$ in $\mathbf{L}_j$ is specified in the following way: for $r = 1$, $l_{j,1,1} = 1$ and for $r = 2, \ldots, m_j$,

$$\begin{cases} l_{r,1}^{(j)} = \cos(\varphi_{j,r,1}) \\ l_{r,s}^{(j)} = \sin(\varphi_{j,r,1}) \ldots \sin(\varphi_{j,r,s-1}) \\ \qquad\qquad \times \cos(\varphi_{j,r,s}), \text{ for } s = 2, \ldots, r-1 \\ l_{r,r}^{(j)} = \sin(\varphi_{j,r,1}) \ldots \sin(\varphi_{j,r,r-2}) \sin(\varphi_{j,r,r-1}), \end{cases}$$

where $\varphi_{j,r,s} \in (0, \pi)$ and $\tau_{r,r}^{(j)} = \sum_{s=1}^{r} (l_{r,s}^{(j)})^2 = 1$ for $r = 1, \ldots, m_j$.

For computer experiments with both quantitative factors and qualitative factors, Zhou et al. (2011) combined the Gaussian correlation function for quantitative factors and hypersphere decomposition for qualitative factors. Specifically, for any two inputs $\mathbf{w}_1 = (\mathbf{x}_1, \mathbf{z}_1)$ and $\mathbf{w}_2 = (\mathbf{x}_2, \mathbf{z}_2)$, they proposed the

covariance between $Y(\mathbf{w}_1)$ and $Y(\mathbf{w}_2)$ to be

$$\phi(Y(\mathbf{w}_1), Y(\mathbf{w}_2)) = \sigma^2 \text{cor}(Z(\mathbf{w}_1), Z(\mathbf{w}_2)) = \sigma^2 R(\mathbf{x}_1, \mathbf{x}_2|\boldsymbol{\theta}) \prod_{j=1}^{q} \tau^{(j)}_{z_{1j}, z_{2j}}, \quad (4)$$

where $R(\mathbf{x}_1, \mathbf{x}_2|\boldsymbol{\theta})$ quantifies the correlation between inputs $\mathbf{x}_1$ and $\mathbf{x}_2$, and $\tau^{(j)}_{z_{1j}, z_{2j}}$ represents the correlation between level z_{1j} and level z_{2j} of the jth qualitative factor.

The correlation function in (4) considers the multiplicity of correlations between quantitative factors and qualitative factors. Such a formulation may not be very flexible to accommodate the complex effects of qualitative factors on the outputs of computer experiments. For example, a zero value of any $\tau^{(j)}_{z_{1j}, z_{2j}}$ in (4) would result in the overall correlation $\phi(Y(\mathbf{w}_1), Y(\mathbf{w}_2))$ being zero. Therefore, it calls for novel, flexible correlation functions for modeling computer experiments with both quantitative factors and qualitative factors.

3. Additive Gaussian Process Model

To enhance the flexibility in capturing the correlations between qualitative and quantitative factors for accurate prediction, we propose a novel Gaussian process model to analyze data for computer experiments with both qualitative and quantitative factors. For the p quantitative factors $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}}$ and the q qualitative factors $\mathbf{z} = (z_1, \ldots, z_q)^{\mathrm{T}}$, we model the corresponding response Y as

$$Y(\mathbf{x}, z_1, \ldots, z_q) = \mu + G_1(z_1, \mathbf{x}) + \cdots + G_q(z_q, \mathbf{x}), \quad (5)$$

where μ is the overall mean, G_j's are independent Gaussian processes with mean zero and the covariance function ϕ_j, for $j = 1, \ldots, q$. Here we adopt the additive form to quantify the contributions of q qualitative input factors to the output. It is in a similar spirit to the additive model in machine learning literature (Hastie and Tibshirani 1990). As the effects of qualitative factors on responses are complicated, the motivation of the additive form is to emphasize the effect of each qualitative factor coupled with quantitative factors. Moreover, the additive formulation enables us to infer the significance of each individual qualitative factor in the model. The proposed model in (5) incorporates interactions between qualitative factors and quantitative factors, and interactions among quantitative factors. However, its current form does not take into account the interactions among qualitative factors.

To construct the covariance function ϕ_j in each Gaussian process G_j, we adopt the approach in Zhou et al. (2011) for the qualitative factors. Recall that the correlation matrix $\mathbf{T}_j = (\tau^{(j)}_{r,s})$ of the m_j levels of the qualitative factor z_j, $j = 1, \ldots, q$. The covariance function ϕ_j for two inputs $\mathbf{w}_1 = (\mathbf{x}_1, \mathbf{z}_1)$ and $\mathbf{w}_2 = (\mathbf{x}_2, \mathbf{z}_2)$ is given by

$$\phi_j(G_j(\mathbf{w}_1), G_j(\mathbf{w}_2)) = \sigma_j^2 \text{cor}(G_j(\mathbf{w}_1), G_j(\mathbf{w}_2)) = \sigma_j^2 \tau^{(j)}_{z_{1j}, z_{2j}} R(\mathbf{x}_1, \mathbf{x}_2|\boldsymbol{\theta}^{(j)}), \quad (6)$$

where σ_j^2 is the variance component associated with G_j, and $R(\mathbf{x}_1, \mathbf{x}_2|\boldsymbol{\theta}^{(j)})$ represents the correlation induced by the quantitative parts $\mathbf{x}_1$ and $\mathbf{x}_2$ with the correlation parameter vector $\boldsymbol{\theta}^{(j)}$. Here we adopt the commonly used Gaussian correlation function in (2) for $R(\mathbf{x}_1, \mathbf{x}_2|\boldsymbol{\theta}^{(j)})$.

By defining $G(\mathbf{x}, \mathbf{z}) = G_1(z_1, \mathbf{x}) + \cdots + G_q(z_q, \mathbf{x})$, it is straightforward to see that $G(\mathbf{x}, \mathbf{z})$ is an additive Gaussian process. That is, the response Y in (5) follows a Gaussian process with mean zero and the covariance function ϕ specified by

$$\begin{aligned}\phi(Y(\mathbf{w}_1), Y(\mathbf{w}_2)) &= \text{cov}(Y(\mathbf{x}_1, \mathbf{z}_1), Y(\mathbf{x}_2, \mathbf{z}_2)) \\ &= \sum_{j=1}^{q} \sigma_j^2 \text{cor}(G_j(z_{1j}, \mathbf{x}_1), G_j(z_{2j}, \mathbf{x}_2)) \\ &= \sum_{j=1}^{q} \sigma_j^2 \tau^{(j)}_{z_{1j}, z_{2j}} R(\mathbf{x}_1, \mathbf{x}_2|\boldsymbol{\theta}^{(j)}), \end{aligned} \quad (7)$$

where $\mathbf{T}_j = (\tau^{(j)}_{r,s})$ is defined as in Section 2.2. The correlation function in (7) is different from that in (4) in two aspects. First, the correlation function for qualitative factors in (4) takes multiplicative form while the counterpart in (7) takes additive form. It implies that if there is one $\tau^{(j)}_{z_{1j}, z_{2j}} = 0$ in (4), then the correlation between $Y(\mathbf{w}_1)$ and $Y(\mathbf{w}_2)$ becomes 0. However, the responses $Y(\mathbf{w}_1)$ and $Y(\mathbf{w}_2)$ can still be highly correlated because of the correlations due to quantitative variables and other qualitative factors. In contrast, the correlation function in (7) does not have this problem. It will not result in zero correlation between $Y(\mathbf{w}_1)$ and $Y(\mathbf{w}_2)$ even if there exist some qualitative factors z_j's having $\tau^{(j)}_{z_{1j}, z_{2j}} = 0$.

Second, the formulation in (7) allows G_j's having different covariance structures for quantitative factors with respect to different qualitative factors. It provides more flexibility for modeling the complex systems of computer experiments than the one in (4) which only has a fixed covariance structure for all quantitative factors. It is worth pointing out that the proposed method is not restricted to the approach in Zhou et al. (2011) for modeling the correlation function of the qualitative factor in each G_j. One can also consider the use of other correlation functions, such as the exchangeable correlation function in Joseph and Delaney (2007) and the multiplicative correlation function in McMillian et al. (1999), for the qualitative factor in each G_j in the proposed method.

Recall that the data are $(y_i, \mathbf{w}_i)$, $i = 1, \ldots, n$, where $\mathbf{w}_i = (\mathbf{x}_i, \mathbf{z}_i)$. Denote by $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ the resulting outputs with the inputs $\mathbf{w}_1, \ldots, \mathbf{w}_n$. Under the proposed model in (5), the log-likelihood function can be written as

$$l(\mu, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T}) = -\frac{1}{2}\big[\log|\boldsymbol{\Phi}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T})| + (\mathbf{y} - \mu\mathbf{1})^{\mathrm{T}} \boldsymbol{\Phi}^{-1}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T})(\mathbf{y} - \mu\mathbf{1})\big], \quad (8)$$

up to some constant. Here $\boldsymbol{\Phi}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T})$ is the covariance matrix of $\mathbf{y}$. Under the proposed model in (5), one can easily see that the covariance matrix $\boldsymbol{\Phi}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T})$ is guaranteed to be a positive definite matrix.

Proposition 1. Let $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$ be n outputs from the input $(\mathbf{w}_1, \ldots, \mathbf{w}_n)$. Under the proposed model in (5), the covariance matrix in (7) of $\mathbf{y}$ is a positive definite matrix.

Proposition 1 can be readily verified by noting that the covariance matrix for $\mathbf{y}$ is

$$\begin{aligned}\boldsymbol{\Phi}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T}) &= \left(\phi(y_i(\mathbf{w}_i), y_{i'}(\mathbf{w}_{i'}))\right)_{n\times n} \\ &= \sum_{j=1}^{q} \sigma_j^2 \mathbf{R}_j \circ \mathbf{H}_j,\end{aligned}$$

where $\circ$ is a Schur product, and $\mathbf{R}_j = (R(\mathbf{x}_i, \mathbf{x}_{i'}|\boldsymbol{\theta}^{(j)}))_{n\times n}$ and $\mathbf{H}_j = (\tau^{(j)}_{z_{ij}, z_{i'j}})_{n\times n}$. Note that $\mathbf{R}_j$ is positive definite and $\mathbf{H}_j$ is a positive semidefinite matrix with the diagonal entries all equal to 1's. Proposition 1 now follows by Lemma 1 (Horn and Johnson 2012).

Lemma 1 (Schur Product Theorem). If $\mathbf{A}$ is an $n \times n$ positive semidefinite matrix with no diagonal entry equal to zero and $\mathbf{B}$ is an $n \times n$ positive definite matrix, then $\mathbf{A} \circ \mathbf{B}$ is positive definite. If both $\mathbf{A}$ and $\mathbf{B}$ are positive definite, then so is $\mathbf{A} \circ \mathbf{B}$.

3.1 Estimation

The model in (5) contains the parameters μ, σ_j^2, $\varphi_{j,r,s}$ ($r = 2, \ldots, m_j, s < r$), and $\boldsymbol{\theta}^{(j)}$, for $j = 1, \ldots, q$. Thus, there are totally $1 + q + \sum_{j=1}^{q} m_j(m_j - 1)/2 + pq$ parameters to be estimated. Let $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_q^2)$, $\mathbf{T} = (\mathbf{T}_1, \ldots, \mathbf{T}_q)$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(q)})$. Furthermore, denote the covariance matrix by $\boldsymbol{\Phi} \equiv \boldsymbol{\Phi}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T}) = \left(\phi(y_i(\mathbf{w}_i), y_{i'}(\mathbf{w}_{i'}))\right)_{n\times n}$. It is easy to see $\phi(y_i, y_i) = \sum_{j=1}^{q} \sigma_j^2$, and $\phi(y_i, y_{i'})$ is specified by (7) if $i \neq i'$.

Recall that the proposed additive Gaussian process model has an explicit expression of the log-likelihood function in (8). Thus, it is natural to consider estimating parameters via the likelihood approach, that is, maximizing the log-likelihood function for parameter estimation. Specifically, given $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T})$, the maximum likelihood estimator of μ is easy to obtained,

$$\hat{\mu} = (\mathbf{1}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{1})^{-1}\mathbf{1}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{y}. \tag{9}$$

Substituting (9) into (8), we obtain that the maximum of (8) is

$$\begin{aligned}l(\hat{\mu}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T}) = -\frac{1}{2}\big[&\log|\boldsymbol{\Phi}| + (\mathbf{y}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{y}) \\ &- (\mathbf{1}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{1})^{-1}(\mathbf{1}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{y})^2\big].\end{aligned}$$

The estimators of $\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T}$ can be obtained as

$$\begin{aligned}[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma}}^2, \hat{\mathbf{T}}] = \operatorname{argmin} &\log|\boldsymbol{\Phi}| + (\mathbf{y}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{y}) \\ &- (\mathbf{1}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{1})^{-1}(\mathbf{1}^{\mathrm{T}}\boldsymbol{\Phi}^{-1}\mathbf{y})^2.\end{aligned} \tag{10}$$

The minimization problem in (10) requires $\varphi_{j,r,s} \in (0, \pi)$ and $\sigma_j \geq 0$ for $j = 1, \ldots, q$. It can be solved using standard nonlinear optimization algorithms in Matlab or R.

3.2 Prediction and Interpolation

The prediction of the proposed additive Gaussian process is similar to the prediction procedure in (3) for the ordinary kriging. Specifically, the prediction of y at a new location $\mathbf{w}_0 = (\mathbf{x}_0, \mathbf{z}_0)$ is the condition mean, that is,

$$E(y(\mathbf{w}_0)|y_1, \ldots, y_n) = \mu + \boldsymbol{\phi}(\mathbf{w}_0)^{\mathrm{T}}\boldsymbol{\Phi}^{-1}(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \mathbf{T})(\mathbf{y} - \mu\mathbf{1}), \tag{11}$$

where $\boldsymbol{\phi}(\mathbf{w}_0) = (\phi(\mathbf{w}_0, \mathbf{w}_1), \ldots, \phi(\mathbf{w}_0, \mathbf{w}_n))^{\mathrm{T}}$. Thus given the estimates $\hat{\mu}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma}}, \hat{\mathbf{T}}$, we have the prediction of y at a new location $\mathbf{w}_0$ is

$$\hat{y}(\mathbf{w}_0) = \hat{\mu} + \boldsymbol{\phi}(\mathbf{w}_0)^{\mathrm{T}}\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma}}^2, \hat{\mathbf{T}})(\mathbf{y} - \hat{\mu}\mathbf{1}). \tag{12}$$

For interpolation, it is straightforward to show that when $\mathbf{w}_0 = \mathbf{w}_i$, the coefficient $\boldsymbol{\phi}(\mathbf{w}_0)^{\mathrm{T}}\boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma}}^2, \hat{\mathbf{T}})$ in (12) is an n-dimensional vector with the ith entry being 1 and otherwise 0. Thus, $\hat{y}(\mathbf{w}_0) = y_i$, achieving the property of interpolation.

Moreover, the single-stage predictor in (12) can be viewed as a sequential predictor. To see this, let $\boldsymbol{\Phi}_j = \sigma_j^2 \mathbf{R}_j \circ \mathbf{H}_j$, we have $\boldsymbol{\Phi} = \sum_{j=1}^{q} \boldsymbol{\Phi}_j$ by Proposition 1. Correspondingly, let $\boldsymbol{\phi}_j(\mathbf{w}_0) = \sigma_j^2 \mathbf{R}(\mathbf{w}_0, \mathbf{w}|\boldsymbol{\theta}^{(j)}) \circ \mathbf{H}_{0j}$, where $\mathbf{R}(\mathbf{w}_0, \mathbf{w}|\boldsymbol{\theta}^{(j)}) = (R(\mathbf{w}_0, \mathbf{w}_1|\boldsymbol{\theta}^{(j)}), \ldots, R(\mathbf{w}_0, \mathbf{w}_n|\boldsymbol{\theta}^{(j)})^{\mathrm{T}}$ and $\mathbf{H}_{0j} = (\tau^{(j)}_{z_{0j}, z_{1j}}, \ldots, \tau^{(j)}_{z_{0j}, z_{nj}})^{\mathrm{T}}$. We have $\boldsymbol{\phi}(\mathbf{w}_0) = \boldsymbol{\phi}_1(\mathbf{w}_0) + \cdots + \boldsymbol{\phi}_q(\mathbf{w}_0)$. Therefore, the prediction at $\mathbf{w}_0$ can be rewritten as

$$\begin{aligned}\hat{y}(\mathbf{w}_0) = \hat{\mu} + (\boldsymbol{\phi}_1(\mathbf{w}_0) + \cdots + \boldsymbol{\phi}_q(\mathbf{w}_0))^{\mathrm{T}} \\ \left(\sum_{j=1}^{q} \boldsymbol{\Phi}_j\right)^{-1} (\mathbf{y} - \hat{\mu}\mathbf{1}),\end{aligned} \tag{13}$$

and referred to as the "joint predictor."

The sequential predictor works as follows. First, we build a predictive model $\hat{y}_1$ by using $\boldsymbol{\phi}_1$. Then obtain the residual vector $\mathbf{e}_1 = \mathbf{y} - \hat{\mathbf{y}}_1$ where $\hat{\mathbf{y}}_1 = (\hat{y}_1(\mathbf{w}_1), \ldots, \hat{y}_1(\mathbf{w}_n))^{\mathrm{T}}$. Then the vector $\mathbf{e}_1$ is treated as a new response vector for building a predictive model $\hat{y}_2$ using $\boldsymbol{\phi}_2$, and one can similarly define the residual vector $\mathbf{e}_2 = \mathbf{e}_1 - \hat{\mathbf{y}}_2$. We continue this procedure till the last stage. More specifically, the sequential prediction provides

$$\hat{y}_1(\mathbf{w}_0) = \hat{\mu} + \boldsymbol{\phi}_1^{\mathrm{T}}(\mathbf{w}_0)\left(\sum_{j=1}^{q} \boldsymbol{\Phi}_j\right)^{-1}(\mathbf{y} - \hat{\mu}\mathbf{1}), \quad \mathbf{e}_1 = \mathbf{y} - \hat{\mathbf{y}}_1;$$

$$\hat{y}_2(\mathbf{w}_0) = \boldsymbol{\phi}_2^{\mathrm{T}}(\mathbf{w}_0)\left(\sum_{j=2}^{q} \boldsymbol{\Phi}_j\right)^{-1}\mathbf{e}_1, \quad \mathbf{e}_2 = \mathbf{y} - \hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_2;$$

$$\vdots$$

$$\hat{y}_k(\mathbf{w}_0) = \boldsymbol{\phi}_k^{\mathrm{T}}(\mathbf{w}_0)\left(\sum_{j=k}^{q} \boldsymbol{\Phi}_j\right)^{-1}\mathbf{e}_{k-1}, \quad \mathbf{e}_k = \mathbf{y} - \sum_{j=1}^{k} \hat{\mathbf{y}}_j;$$

$$\vdots$$

$$\hat{y}_q(\mathbf{w}_0) = \boldsymbol{\phi}_q^{\mathrm{T}}(\mathbf{w}_0)\boldsymbol{\Phi}_q^{-1}\mathbf{e}_{q-1}.$$

We then form a sequential predictor as

$$\hat{y}_{\mathrm{seq}}(\mathbf{w}_0) = \sum_{j=1}^{q} \hat{y}_j(\mathbf{w}_0). \tag{14}$$

The following theorem shows that the "joint predictor" and "sequential predictor" are equivalent. The proof of the theorem is given in the supplementary materials.

Theorem 1. With the same parameter values, the "joint predictor" in (13) and "sequential predictor" in (14) are equivalent.

3.3 Connection to Composite Gaussian Process

The proposed model in (5) can be viewed as a generalization of a composite Gaussian process model proposed by Ba and Joseph (2012) for computer experiments with quantitative factors. A composite Gaussian process model assumes

$$Y = \mu + G_1(\mathbf{x}) + G_2(\mathbf{x}), \tag{15}$$

where G_1 and G_2 are two independent Gaussian process models with the covariance functions $\sigma_1^2 R(\cdot|\boldsymbol{\theta}^{(1)})$ and $\sigma_2^2 R(\cdot|\boldsymbol{\theta}^{(2)})$, respectively. By imposing the constraints $\sigma_1^2 \geq \sigma_2^2$ and $\boldsymbol{\theta}^{(1)} \leq \boldsymbol{\theta}^{(2)}$, the model in (15) aims that the global component $G_1(\mathbf{x})$ can capture more variation in the response than the local process $G_2(\mathbf{x})$. The composite Gaussian process model in Ba and Joseph (2012) can also accommodate the nonstationary variance components.

A computer experiment with quantitative factors $\mathbf{x}$ can be viewed as having two additional qualitative factors z_1 and z_2 each having only one level. The correlation matrices for z_1 and z_2 are $\mathbf{T} = \mathbf{1}\mathbf{1}^{\mathrm{T}}$. By imposing $\sigma_1^2 \geq \sigma_2^2$ and $\boldsymbol{\theta}^{(1)} \leq \boldsymbol{\theta}^{(2)}$, we can see that the proposed method $Y(\mathbf{x}, z_1, z_2) = \mu + G_1(z_1, \mathbf{x}) + G_2(z_2, \mathbf{x})$ is equivalent to a composite Gaussian process model.

4. Examples

In this section, we conduct a simulation study to demonstrate the effectiveness of the proposed model in (5) with the covariance function in (7). Specifically, we compare the proposed model with three methods investigated in Zhou et al. (2011). These three methods consider a Gaussian process model in (1) with the correlation function in (4). In particular, the Gaussian correlation function is used for quantitative factors and the following three correlation functions are adopted for qualitative factors for the three methods, respectively:

(a) the exchangeable correlation function $\tau_{r,s} = c$ $(0 < c < 1)$ for $r \neq s$ (Joseph and Delaney 2007; Qian, Wu, and Wu 2008);
(b) the multiplicative correlation function $\tau_{r,s} = \exp\{-(\theta_r + \theta_s)\}$ $(\theta_r > 0, \theta_s > 0)$ for $r \neq s$ (McMillian et al. 1999; Qian, Wu, and Wu 2008);
(c) the unrestrictive correlation function $\tau_{r,s}$ in Section 2.2 (Zhou et al. 2011).

Following the notation in Zhou et al. (2011), the three methods are denoted by "EC," "MC," and "UC," respectively, while the proposed method is denoted by "AD_UC" as it adopts "UC" for each qualitative factors. Here we have not included Zhang and Notz (2015) for comparison, since their method also adopts the multiplicity of correlations between quantitative and qualitative factors, in a very similar spirit as that in Zhou et al. (2011). For the comparison purpose, we also include the proposed additive Gaussian process model with EC and MC correlation structures for each qualitative factor, denoted as "AD_EC" and "AD_MC," respectively.

Several criteria can be used to evaluate the performance of different methods in comparison. Following Zhou et al. (2011), we adopt the root mean square error (RMSE) given by

$$\mathrm{RMSE} = \sqrt{\frac{1}{|\mathcal{W}_{\mathrm{pred}}|} \sum_{\mathbf{w} \in \mathcal{W}_{\mathrm{pred}}} (\hat{y}(\mathbf{w}) - y(\mathbf{w}))^2}, \tag{16}$$

where $\hat{y}(\mathbf{w})$ and $y(\mathbf{w})$ are the predicted response and the true response at the new input $\mathbf{w}$ in the hold-out set $\mathcal{W}_{\mathrm{pred}}$. A relevant criterion is Nash-Sutcliffe efficiency (Kaufman et al. 2011) given by

$$\mathrm{NSE} = 1 - \frac{\sum_{\mathbf{w} \in \mathcal{W}_{\mathrm{pred}}} (\hat{y}(\mathbf{w}) - y(\mathbf{w}))^2}{\sum_{\mathbf{w} \in \mathcal{W}_{\mathrm{pred}}} (\hat{y}(\mathbf{w}) - \bar{y})^2}, \tag{17}$$

where $\hat{y}(\mathbf{w})$ and $y(\mathbf{w})$ are defined as in (16) and $\bar{y}$ is the average of the predictions. The second term in (17) is the ratio of an estimated predicted mean square error to the unstandardized variance of $Y(\mathbf{w})$. The NSE represents an estimate of the proportion of the variability in Y that is explained by the model. Thus, the NSE can be interpreted as a performance measure in analogy to R^2 in linear regression. For the relationship between RMSE and NSE, methods with lower RMSE typically yield higher NSE.

Examples 1–3 consider three computer models and evaluate the performance of the six methods, "EC," "MC," "UC," "AD_EC," "AD_MC," "AD_UC." These examples are chosen under the consideration of having strong interactions between qualitative factors and quantitative factors. Specifically, we choose the three examples to evaluate the performance of the different methods under three scenarios. Example 1 considers that the significant two-factor interactions in the model are those between qualitative variables and quantitative variables. Example 2 considers that the significant two-factor interactions in the model are those between qualitative factors and quantitative variables as well as those between quantitative variables. Example 3 considers that the significant two-factor interactions include interactions between quantitative variables, between quantitative variable and qualitative variables, and between qualitative factors. In each example, we implement the six methods over 100 simulations and report the associated RMSEs.

Example 1. Consider a computer experiment with $p = 6$ quantitative factors and $q = 5$ qualitative factors each having three levels. Data are generated from the computer model

$$y = \sum_{i=1}^{5} \frac{x_i z_{6-i}}{4000} + \prod_{i=1}^{5} \cos\left(\frac{x_i}{\sqrt{i}}\right) \sin\left(\frac{z_{6-i}}{\sqrt{i}}\right), \tag{18}$$

where $-100 < x_i < 100$ for $i = 1, \ldots, p$ and $z_j = \{-50, 0, 50\}$ for $j = 1, \ldots, q$. In each simulation, an 81-run design is adopted, where a three-level fractional factorial design (Wu and Hamada 2009) is used for qualitative factors and a random Latin hypercube design (McKay, Beckman, and Conover 1979) is used for quantitative factors. The RMSE in (16) is computed based on the hold-out set $\mathcal{W}_{\mathrm{pred}}$ with 2430 points consisting of four replicates of a full factorial three-level design for qualitative factors and a random Latin hypercube design for quantitative factors. Figure 1 displays the boxplots of the RMSEs associated with "EC," "MC," "UC," "AD_EC," "AD_MC," and "AD_UC" for $p = 6$ over 100 simulations. From Figure 1, it is clearly seen that the

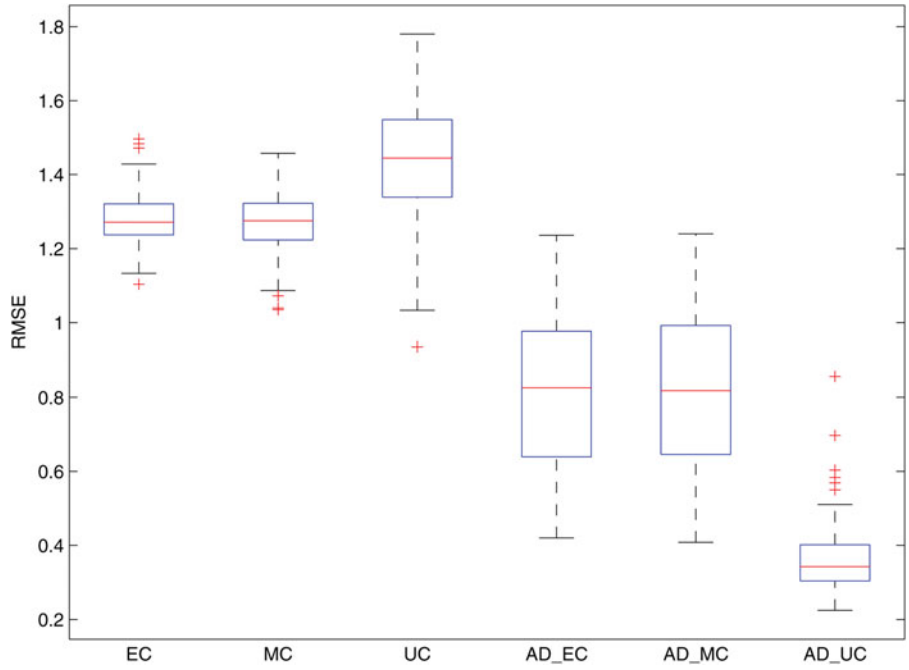

Figure 1. Boxplots of RMSEs associated with "EC," "MC," "UC," "AD_EC," "AD_MC," and "AD_UC" for the computer model in (18) with $p = 6$ over 100 simulations.

proposed method "AD_UC" outperforms the other three methods since its corresponding RMSE is significantly lower than that of the alternatives. Also the additive Gaussian process methods "AD_EC," "AD_MC," and "AD_UC" perform better than "EC," "MC," and "UC." It is noting that although the multiplicative correlation model UC itself has relatively larger RMSE than EC and MC in this example, the "AD_UC" gets a smaller RMSE than "AD_EC" and "AD_MC."

Example 2. Consider a computer experiment with $p = 7$ quantitative factors and $q = 6$ qualitative factors $z_1, \ldots, z_6$, where z_1 and z_2 have two levels, z_3, z_4, and z_5 have three levels, and z_6 has four levels. Consider the computer model,

$$\begin{aligned} y = &\sum_{i=1}^{5} \exp\{-x_i\}\cos(4z_{7-i})\sin(4x_{i+2}) \\ &+ \exp\{-x_6\}\cos(4z_1)\sin(4z_1) \\ &+ \exp\{-x_7\}\cos(4x_7)\sin(4z_2), \end{aligned} \tag{19}$$

where $0 < x_i < 1$ for $i = 1, \ldots, p$, $z_j = \{0.3, 0.8\}$ for $j = 1, 2$, $z_j = \{0.1, 0.5, 0.9\}$ for $j = 3, 4, 5$ and $z_{p+6} = \{0.05, 0.35, 0.65, 0.95\}$. In each simulation, a 142-run design is adopted, where two replicates of a 72-run mixed-level fractional factorial design are used for qualitative factors and a random Latin hypercube design is used for quantitative factors. The RMSE in (16) is computed based on the hold-out set $\mathcal{W}_{\text{pred}}$ with 2160 points consisting of five replicates of a full factorial design for qualitative factors and a random Latin hypercube design for quantitative factors. Figure 2 displays the boxplots of the RMSEs associated with "EC," "MC," "UC," "AD," "AD_EC," "AD_MC," and "AD_UC" over 100 simulations. The results in Figure 2 clearly indicate that the proposed method "AD_UC" provides much lower RMSEs. One can also see that the prediction performances of "EC," "MC," and "UC" are very comparable in this example, while "AD_UC" is still relatively better than "AD_EC" and "AD_MC" in terms of RMSE values.

Example 3. Consider the computer model,

$$\begin{aligned} y = &\, 5g_1(z_5) + 3g_2(z_4) + 1.2g_1(x_4) + 1.5g_2(x_8) + 2.3g_3(x_2) \\ &+ 7g_2(x_2) + 4g_3(1.5z_3 + x_1) + 7g_4(1.2z_2 + x_3) \\ &+ 4.5g_3(x_9 + x_6) + 3g_3(x_4 + x_5) \\ &+ 1.1g_2(z_2 + z_3) + 1.5g_2(z_1 + z_5), \end{aligned} \tag{20}$$

where $0 < x_i < 1$ for $i = 1, \ldots, p$, $x_j \in \{0.1, 0.5, 0.9\}$ for $j = p+1, \ldots, k$, $g_1(x) = x$, $g_2(x) = (2x-1)^2$, $g_3(x) = \sin(2\pi x)/[2 - \sin(2\pi x)]$, and $g_4(x) = 0.1\sin(2\pi x) + 0.2\cos(2\pi x) + 0.3[\sin(2\pi x)]^2 + 0.4[\cos(2\pi x)]^3 + 0.5[\sin(2\pi x)]^3$. The functions g_1, g_2, g_3, g_4 are defined in Reich, Storlie, and Bondell (2009). In each simulation, a three-level fractional factorial design of 81 runs is used for qualitative factors and a random Latin hypercube design of 81 runs is used for quantitative factors. The RMSE in (16) is computed based on the hold-out set $\mathcal{W}_{\text{pred}}$ with 1215 points consisting of five replicates of a full factorial design for qualitative factors and a random Latin hypercube design for quantitative factors. Figure 3 displays the boxplots of the RMSEs associated with "EC," "MC," "UC," "AD_EC," "AD_MC," and "AD_UC" for $q = 5$ and $p = 9$. The results show that the additive Gaussian process methods "AD_EC," "AD_MC," and "AD_UC" still perform better than the other three methods in comparison. However, the advantage of the additive Gaussian process methods over the other three methods is not as significant as those in Examples 1 and 2. It is also worth noting that in this example, the advantage of the proposed method "AD_UC" is comparable to "AD_EC" and "AD_MC." One possible explanation is that the computer model in (20) does not have as strong interactions between qualitative factors and quantitative factors as those in (18) and (19).

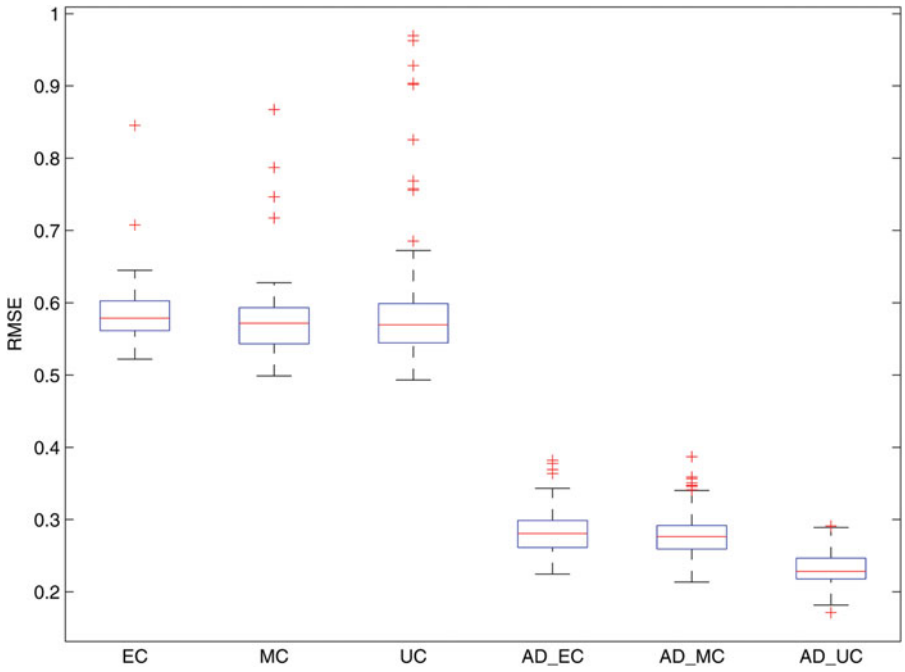

Figure 2. Boxplots of the RMSEs associated with "EC," "MC," "UC," "AD_EC," "AD_MC," and "AD_UC" for the computer model in (19) over 100 simulations.

5. Real Data Analysis

In this section, we apply the proposed method to a real application in which the computer experiment has both qualitative and quantitative factors. A fully 3D coupled finite element model has been calibrated and verified by successfully capturing both the deformations and stresses of full scale embankments involving unreinforced, piled, and two different reinforced and piled sections (Rowe and Liu 2015). Given the cost of building and monitoring full-scale reinforced and column-supported embankment in field, a validated numerical modeling is usually regarded as a cost-effective tool to advance the knowledge of complex issues in such a system involving geosynthetic reinforced platform, embankment fill, columns, and geosynthetic reinforcement. In this study, the aforementioned validated numerical model was used to investigate the influence of three qualitative factors and one quantitative factor for improving the performance of reinforced embankments with floating columns over soft clay. Figure 4 illustrates the structure of this full scale embankment. A 7 meter (m) thick reinforced embankment was constructed over a 15 m soft clay deposit improved with 1-m-diameter and 9-m-long columns at 2 m center-to-center spacing. The finite element discretization for the case examined had

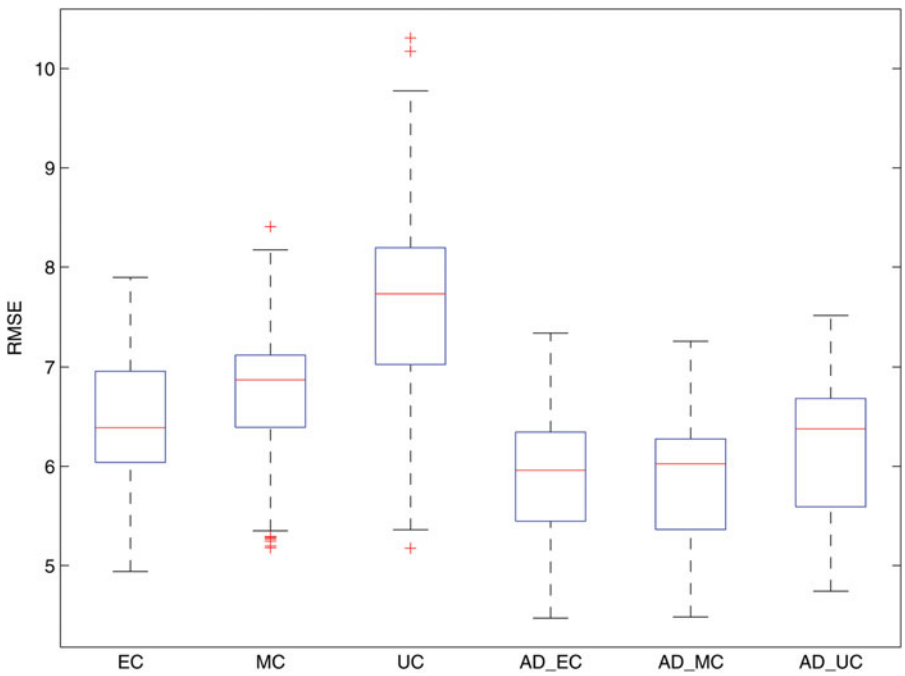

Figure 3. Boxplots of the RMSEs associated with "EC," "MC," "UC," "AD_EC," "AD_MC," and "AD_UC" for the computer model in (20) over 100 simulations.

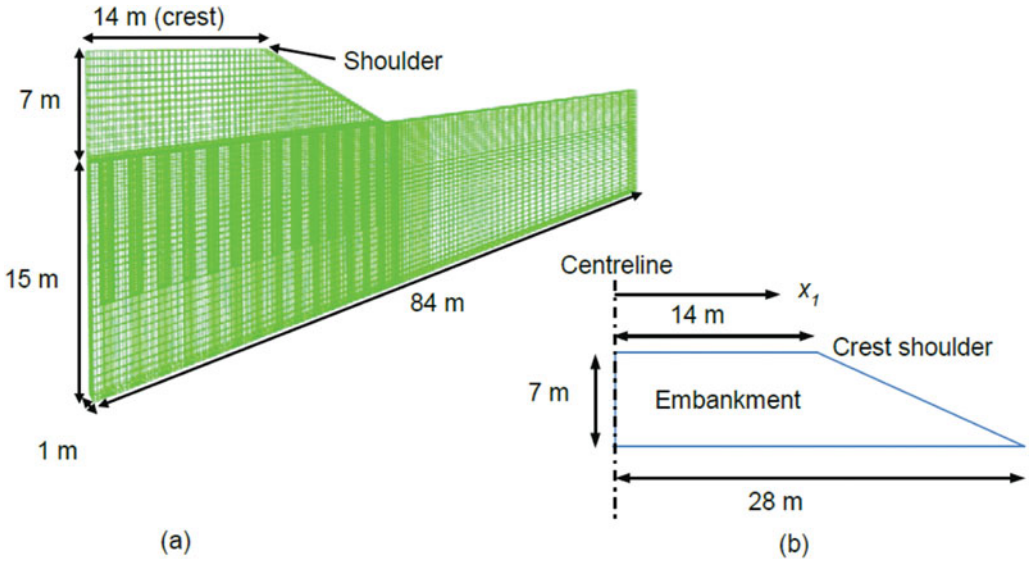

Figure 4. The embankment examined: (a) finite element mesh; (b) the schematic view of embankment constructed on foundation soil.

36,802 elements and 69,667 nodes (Figure 4(a)). The three qualitative factors are embankment construction rate (z_1), Young's modulus of columns (z_2), and reinforcement stiffness (z_3), and the quantitative factor x_1 is the distance from the embankment centerline to the embankment shoulder (Figure 4(b)). An average run for one case of this size took roughly 9 hr on a 12-noded parallel super-computer at the High Performance Computing Virtual Laboratory (HPCVL). The response variable considered herein is the final embankment crest settlement U_3, which is a crucial embankment working indicator.

For the computer experiment, each of the three qualitative factors z_1, z_2, z_3 has three levels: the levels of z_1 are 1, 5, 10 m/month; the levels of z_2 are 50, 100, 200 MPa; and the levels of z_3 are 1578, 4800, 8000 kN/m. The quantitative factor x_1 takes the 29 values uniformly from the interval [0, 14]. For each value of the quantitative factor, a three-level fractional factorial design of nine runs is used for the qualitative factors. Thus, there are 261 design points, which are used for model estimation for the four methods "EC," "MC," "UC," and "AD," respectively.

To compare the prediction performance of these four methods, we evaluate their prediction performance on the test data. Specifically, the test dataset contains 29 input settings in which the values of quantitative factor x_1 are taken uniformly from the interval [0, 14], and the setting of the qualitative factors is $(z_1, z_2, z_3) = (5, 100, 4800)$. Note that such a setting of qualitative factors is not used in the nine-run three-level fractional factorial design.

Figure 5 displays the boxplots of the RMSEs in (16) associated with "EC," "MC," "UC," and "AD_UC" for the fully 3D coupled finite element model over 100 replications. From the boxplots in the figure, one can clearly see that the proposed method performs much better than the other three methods in terms of prediction accuracy and precision. It is worth pointing out that the RMSEs associated with "EC," "MC," "UC" have large variations with a significant portion of outliers in their boxplots. A possible explanation is that those methods have not full captured the underlying correlation structures of qualitative and qualitative factors, resulting in large bias in the estimation of the mean response. In addition, the proposed method gives nearly identical RMSEs over 100 replications because in this particular example the maximum likelihood estimates are found nearly identical over those replications regardless the initial values of parameters in the maximum likelihood optimization. In summary for the case examined, the proposed "AD_UC" method presents a very promising and time-saving tool to achieve excellent agreement with the computations from complex computer experiments (i.e., fully 3D coupled finite element modeling).

6. Discussion

In this work, we propose an additive Gaussian process for modeling computer experiments with both quantitative and qualitative factors. Several illustrative examples and a real application have demonstrated that the proposed method can build a more accurate emulator comparing with the existing methods. The reason that might explain the success of the proposed method is that, the proposed model employs a more flexible covariance structure that is capable of accommodating the complex interaction effects between qualitative factors and quantitative factors.

A few remarks are worth mentioning here. First, the current work only considers the additive Gaussian process with respect to the qualitative factors. There are several existing work considering additive kernels for the quantitative factors (Durrande et al. 2011; Duvenaud, Nickisch, and Rasmussen 2011). How to incorporate additive kernels for the quantitative factors in our method can be an interesting topic for future research. To extend the proposed method for accommodating the interactions between qualitative factors, one possibility is to include those interactions in the mean part. By conducting variable selection to include significant interactions in the mean part, the proposed method deserves thorough investigations in the future work.

Second, the empirical study shows that the proposed method is particularly useful in building Gaussian process modeling when the number of qualitative factors is relatively large, and the interaction effects between qualitative factors and quantitative

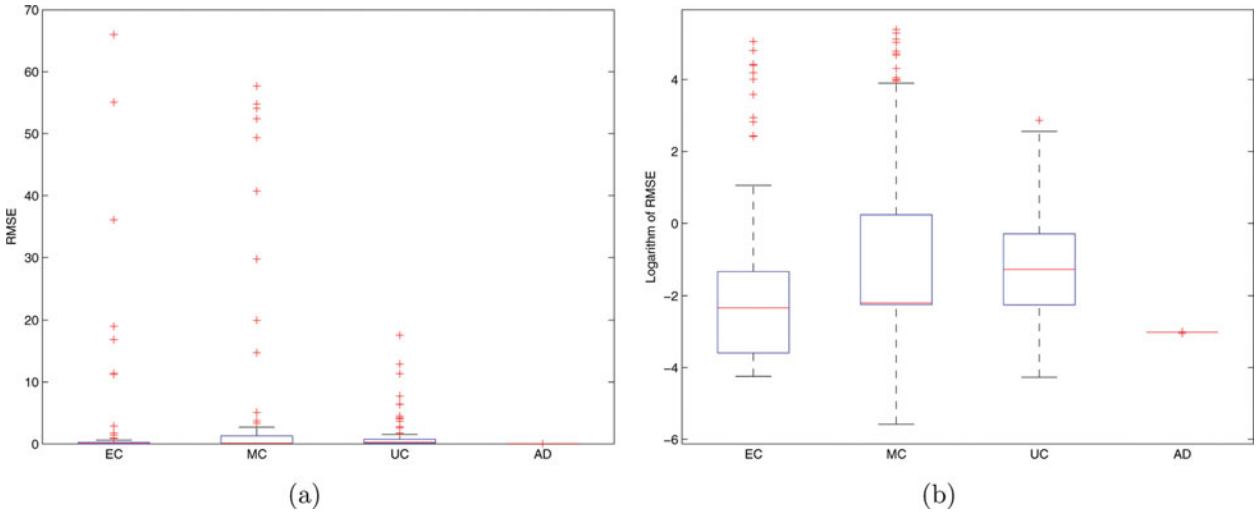

Figure 5. Boxplots of the RMSEs associated with "EC," "MC," "UC," and "AD_UC" for the fully 3D coupled finite element model over 100 replications: (a) RMSE; (b) logarithm of RMSE.

factors are strong and complex. If instead the number of qualitative factors is small and such interaction effects are not strong or complex, the performance of the proposed method and other existing methods are likely to be very comparable. It means that existing methods are probably sufficient in providing an accurate emulator. Third, recall that the number of parameters in the proposed model is $1 + q + \sum_{j=1}^{q} m_j(m_j - 1)/2 + pq$, and thus when the number of input variables and/or the levels of qualitative factors are large, parameter estimation can be computationally cumbersome. There is a need for a computationally efficient estimation procedure. Finally, in a recent article Deng, Hung, and Lin (2015) introduced marginally coupled designs for computer experiments with both qualitative and quantitative factors. Ba, Myers, and Brenneman (2015) proposed optimal sliced Latin hypercube designs paired with fractional factorial designs to accommodate both qualitative and quantitative factors. Note that the space-filling property of a design can have a significant impact on the model performance. It would be interesting to explore the possibility of coupling the proposed model with those designs for more accurate emulators.

Supplementary Materials

The supplementary material for this article contains the following: (1) proof of Theorem 1, (2) Matlab codes for implementation of the proposed method, and (3) data from the real application in Section 5.

Acknowledgments

The authors thank the editor, the associate editor, and two referees for their insightful comments. Lin is supported by the Natural Sciences and Engineering Research Council of Canada. Deng is supported by the National Science of Foundation of United States.

References

Ba, S., and Joseph, V. R. (2012), "Composite Gaussian Process Models for Emulating Expensive Functions," *The Annals of Applied Statistics*, 6, 1838–1860. [287]

Ba, S., Myers, W. R., and Brenneman, W. A. (2015), "Optimal Sliced Latin Hypercube Designs," *Technometrics*, 57, 479–487. [291]

Bhuiyan, H., Chen, J., Khan, M., and Marathe, M. V. (2014), "Fast Parallel Algorithms for Edge-Switching to Achieve a Target Visit Rate in Heterogeneous Graphs," *Parallel Processing (ICPP), 2014 43rd International Conference on IEEE*, pp. 60–69. [283]

Deng, X., Hung, Y., and Lin, C. D. (2015), "Design for Computer Experiments With Qualitative and Quantitative Factors," *Statistica Sinica*, 25, 1567–1581. [291]

Durrande, N., Ginsbourger, D., Roustant, O., and Carraro, L. (2011), " Additive Covariance Kernels for High-Dimensional Gaussian Process Modeling," arXiv preprint arXiv:1111.6233. [290]

Duvenaud, D. K., Nickisch, H., and Rasmussen, C. E. (2011), "Additive Gaussian Processes," *Advances in Neural Information Processing Systems*, pp. 226–234. [290]

Fang, K. T., Li, R., and Sudjianto, A. (2005), *Design and Modeling for Computer Experiments*, New York: Chapman & Hall/CRC Press. [284]

Han, G., Santner, T. J., Notz, W. I., and Bartel, D. L. (2009), "Prediction for Computer Experiments Having Quantitative and Qualitative Input Variables," *Technometrics*, 51, 278–288. [283]

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, New York: CRC Press. [285]

Horn, R. A., and Johnson, C. R. (2012), *Matrix Analysis* (2nd ed.), New York: Cambridge University Press. [286]

Joseph, V. R., and Delaney, J. D. (2007), "Functionally Induced Priors for the Analysis of Experiments," *Technometrics*, 49, 1–11. [283,285]

Kaufman, C. G., Bingham, D., Habib, S., Heitmann, K., and Frieman, J. A. (2011), "Efficient Emulators of Computer Experiments Using Compactly Supported Correlation Functions, With an Application to Cosmology," *The Annals of Applied Statistics*, 5, 2470–2492. [287]

Liu, K.-W., and Rowe, R. K. (2015), "Numerical Study of the Effects of Geosynthetic Reinforcement Viscosity on Behaviour of Embankments Supported by Deep-Mixing-Method (DMM) Columns," *Geotextiles and Geomembranes*, 43, 567–578. [283]

Matern, B. (1986), *Spatial Variation* (2nd ed.), New York: Springer-Verlag. [284]

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, 21, 239–245. [287]

McMillian, N. J., Sacks, J., Welch, W. J., and Gao, F. (1999), "Analysis of Protein Activity Data by Gaussian Stochastic Process Models," *Journal of Biopharmaceutical Statistics*, 9, 145–160. [283,285]

Pinheiro, J. C., and Bates, D. M. (1996), "Unconstrained Parametrizations for Variance-Covariance Matrices," *Statistics and Computing*, 6, 289–296. [283]

Qian, P. Z. G., Wu, H., and Wu, C. F. J. (2008), "Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors," *Technometrics*, 50, 283–396. [283]

Rebonato, R., and Jackel, P. (1999), "The Most General Methodology for Creating a Valid Correlation Matrix for Risk Management and Option Pricing Purposes," *The Journal of Risk*, 2, 17–27. [283]

Reich, B. J., Storlie, C. B., and Bondell, H. D. (2009), "Variable Selection in Bayesian Smoothing Spline ANOVA Models: Application to Deterministic Computer Codes," *Technometrics*, 51, 110–120. [288]

Rowe, R. K., and Liu, K.-W. (2015), "3D Finite Element Modeling of a Full-Scale Geosynthetic-Reinforced, Pile-supported Embankment," *Canadian Geotechnical Journal*, 52, 2041–2054. [283,289]

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409–423. [283]

Santner, T. J., Williams, B. J., and Notz, W. I. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer. [283,284]

Swiler, L. P., Hough, P. D., Qian, P., Xu, X., Storlie, C., and Lee, H. (2014), "Surrogate Models for Mixed Discrete-Continuous Variables," in *Constraint Programming and Decision Making*, M. Ceberio and V. Kreinovich (eds.), New York: Springer, pp. 181–202. [284]

Wu, C. F. J., and Hamada, M. (2009), *Experiments: Planning, Analysis, and Optimization*, New York: Wiley. [287]

Zhang, Y., and Notz, W. I. (2015), "Computer Experiments With Qualitative and Quantitative Variables: A Review and Reexamination," *Quality Engineering*, 27, 2–13. [283,287]

Zhou, Q., Qian, P. Z. G., and Zhou, S. (2011), "A Simple Approach to Emulation for Computer Models With Qualitative and Quantitative Factors," *Technometrics*, 53, 266–273. [283,284,285,287]