Quality Engineering

ISSN: 0898-2112 (Print) 1532-4222 (Online) Journal homepage: https://www.tandfonline.com/loi/lqen20

Computer Experiments with Qualitative and Quantitative Variables: A Review and Reexamination

Yulei Zhang & William I. Notz

To cite this article: Yulei Zhang & William I. Notz (2015) Computer Experiments with Qualitative and Quantitative Variables: A Review and Reexamination, Quality Engineering, 27:1, 2-13, DOI: 10.1080/08982112.2015.968039

To link to this article: https://doi.org/10.1080/08982112.2015.968039

Quality Engineering, 27:2–13, 2015

ISSN: 0898-2112 print / 1532-4222 online
DOI: 10.1080/08982112.2015.968039

Taylor & Francis
Taylor & Francis Group

Computer Experiments with Qualitative and Quantitative Variables: A Review and Reexamination

Yulei Zhang,
William I. Notz
Department of Statistics,
The Ohio State University,
Columbus, Ohio

ABSTRACT In this article, we review and reexamine approaches to modeling computer experiments with qualitative and quantitative input variables. For those not familiar with models for computer experiments, we begin by showing, in a simple setting, that a standard model for computer experiments can be viewed as a generalization of regression models. We then review models that include both quantitative and quantitative variables and present some alternative parameterizations. Two are based on indicator functions and allow one to use standard quantitative inputs-only models. Another parameterization provides additional insight into possible underlying factorial structure. Finally, we use two examples to illustrate the benefits of these alternative models

KEYWORDS computer experiments, qualitative input, GaSP model, best linear unbiased predictor, Gaussian correlation function, indicator variables

INTRODUCTION

Experiments are a powerful tool for understanding how factors affect a response generated by a physical process. Unfortunately, some physical processes are difficult, expensive, or impossible to investigate directly. For example, how changes in emission standards for automobiles would affect average global temperatures 20 years in the future is not amenable to traditional experimentation.

An increasingly popular alternative to physical experiments in these situations is to use computer simulations. We describe a physical process by a mathematical model implemented with code on a computer. This code is sometimes referred to as a *simulator.* The inputs to the code are factors that are believed to affect the response and the output is the simulated response. We assume that the output is deterministic. We use the code to explore or experiment with the physical process; that is, we try different inputs in order to assess their effect on the outputs. We call this a *computer experiment.* The code runs slowly. One run may take a day or longer. Thus, we can only observe (experiment with) the code a small number of times. To augment the limited number of runs of the code, we fit a statistical model (predictor) to the runs and use the statistical model to predict the code at unobserved inputs. This statistical predictor is sometimes called an *emulator.*

This article was presented at the Second Stu Hunter Research Conference in Tempe, Arizona, March 2014.

Address correspondence to William I. Notz, Department of Statistics, Ohio State University, 1958 Neil Ave., Columbus, OH 43210-1247.
E-mail: win@stat.osu.edu

More specifically, we assume that the code produces the deterministic output

$$y(\boldsymbol{x},\ t)$$

that depends on a set of quantitative input variables

$$\boldsymbol{x} = (x_1,\ x_2, \ldots,\ x_d)^\top$$

and a qualitative variable having T levels, here indexed by t.

We also assume that the quantitative input variables are restricted to some subset

$$X \subset \mathbf{R}^d$$

and that we observe the code at n points

$$(\boldsymbol{x}_1, t_1), (\boldsymbol{x}_2, t_2), \ldots, (\boldsymbol{x}_n, t_n).$$

When all of the inputs are quantitative, a popular statistical model for the output of the simulator is the so-called Gaussian stochastic process (GaSP) model. The predictor, or emulator, associated with this model has many attractive properties, including the fact that it interpolates the data. Many statisticians are not familiar with the statistical models (kriging or Gaussian stochastic process models) used in computer experiments and the associated predictor. Thus, in the following section we introduce the GaSP model in a simple setting by demonstrating its connection to more familiar regression models. Then we present the general model for quantitative inputs and review extensions that incorporate qualitative variables. Next, we reexamine a popular model for qualitative and quantitative inputs, providing some alternative parameterizations. Finally, we use two examples to demonstrate the alternative parameterizations and show how they can provide additional insights.

THE GaSP MODEL, A REGRESSION PERSPECTIVE

Modeling in a Simple Setting

The standard simple linear regression model is often written as

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ 1 \le i \le n,$$

where a response Y is observed at n values, $x_1, \ldots, x_n$, of the input variable x. Here, Y_i is the response at input x_i, and the relationship between the expected value of Y and x is assumed to be a straight line with intercept β_0 and slope β_1. The ϵ_i are independent and identically distributed (i.i.d.) $N(0, \sigma^2)$ errors.

To display the dependence on x_i we can write this model as

$$Y(x_i) = \beta_0 + \beta_1 x_i + \epsilon(x_i),\ 1 \le i \le n,$$

where Y_i and ϵ_i have been replaced with $Y(x_i)$ and $\epsilon(x_i)$. To further emphasize the dependence on an arbitrary x, we could re-express the model as

$$Y(x) = \beta_0 + \beta_1 x + \epsilon(x),$$

where $\epsilon(x)$ is a random variable with the property that if Y is observed at $x_1, \ldots, x_n$, then $\epsilon(x_1), \ldots, \epsilon(x_n)$ are i.i.d. $N(0, \sigma^2)$.

In some situations we might not wish to model the errors $\epsilon(x_i)$ as independent. For example, if we have repeated measurements on a single subject we might prefer to model the $\epsilon(x_i)$ as $N(0, \sigma^2)$ but with

$$\mathrm{cov}(\epsilon(x_i), \epsilon(x_j)) = \begin{cases} \sigma^2 & \text{if } i = j \\ \sigma^2\theta & \text{if } i \ne j \end{cases},$$

where $-\frac{1}{n} < \theta < 1$. Or if x represents the time at which Y is observed, we might model the ϵ_i as $N(0, \sigma^2)$ but with

$$\mathrm{cov}(\epsilon(x_i), \epsilon(x_j)) = \begin{cases} \sigma^2 & \text{if } i = j \\ \sigma^2\theta^{|x_i - x_j|} & \text{if } i \ne j \end{cases}.$$

This corresponds to a time series model with an exponential autocorrelation function.

An alternative way to express this is to write $\mathrm{cov}(\epsilon(x_i), \epsilon(x_j)) = \sigma^2 R(x_i, x_j|\,\cdot)$ for some known function R that may depend on certain parameters. For example, in our reported measures example, we might define

$$R(x_i, x_j|\theta) = \begin{cases} 1 & \text{if } i = j \\ \theta & \text{if } i \ne j \end{cases}.$$

In our time series example, we might define

$$R(x_i, x_j|\theta) = \begin{cases} 1 & \text{if } i = j \\ \theta^{|x_i - x_j|} & \text{if } i \ne j \end{cases}.$$

To be more precise, let $R(\cdot, \cdot|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of (unknown) parameters, be a function with the

following property: For any integer $n \geq 1$ and any n values of x, say $x_1, \ldots, x_n$, $\boldsymbol{R} = ((R(x_i, x_j|\boldsymbol{\theta})))$, the $n \times n$ matrix with i, jth entry $R(x_i, x_j|\boldsymbol{\theta})$, is a valid correlation matrix (symmetric, positive definite, with unit diagonal entries). Such a function $R(\cdot, \cdot|\boldsymbol{\theta})$ is called a (valid) correlation function.

We can then characterize $\epsilon(x)$ as a random variable with the property that for any integer $n \geq 1$ and any n values of x, say $x_1, \ldots, x_n$, $\epsilon(x_1), \ldots, \epsilon(x_n)$ have a joint multivariate normal distribution with mean $\mathbf{0}_{n \times 1}$ and $n \times n$ covariance matrix whose i, jth entry is $\sigma^2(R(x_i, x_j|\boldsymbol{\theta}))$ for some correlation function $R(\cdot, \cdot|\boldsymbol{\theta})$. Here $\mathbf{0}_{n \times 1}$ is the $n \times 1$ vector of 0s.

To summarize, we have now written the simple linear regression model with possibly correlated observations as

$$Y(x) = \beta_0 + \beta_1 x + \epsilon(x),$$

where $\epsilon(x)$ is a random variable with the property that if Y is observed at $x_1, \ldots, x_n$, then the joint distribution of $\epsilon(x_1), \ldots, \epsilon(x_n)$ is multivariate normal with mean $\mathbf{0}_{n \times 1}$ and $n \times n$ covariance matrix $\sigma^2((R(x_i, x_j|\boldsymbol{\theta})))$ for some correlation function $R(\cdot, \cdot|\boldsymbol{\theta})$. This is still the standard regression model but with additional, seemingly clumsy, notation.

In the stochastic process literature, a random quantity such as $\epsilon(x)$ is referred to as a Gaussian stochastic process with mean 0, variance σ^2, and correlation function $R(\cdot, \cdot|\boldsymbol{\theta})$. If this correlation function has the property that $R(x_i, x_j|\boldsymbol{\theta}) = R^*(x_i - x_j|\boldsymbol{\theta})$, in other words is a function of x_i, x_j only through $x_i - x_j$, then $\epsilon(x)$ is said to be a second-order stationary Gaussian process. Instead of referring to our simple linear regression model as a linear model with possibly correlated errors, researchers in computer experiments refer to the model as a GaSP model with mean function $\beta_0 + \beta_1 x$.

We could interpret this model as

$$Y(x) = \underbrace{\beta_0 + \beta_1 x}_{\text{regression or global trend}} + \underbrace{\epsilon(x)}_{\text{local trend or residual variation}}.$$

The global trend is a straight line and, if the errors are independent, the local trend is (white) noise.

In a first regression course one plots the residuals versus x to assess the validity of the model. Nonrandom patterns in the local trend may indicate that a straight line is not a good fit or the presence of a correlation. For example, unusually long runs of positive residuals or negative residuals may indicate the presence of autocorrelation. To put this another way, if the errors are correlated, this correlation may produce local variation that does not look like white noise. What is surprising is that certain types of correlation produce a local trend that is actually a smooth curve. For example, the correlation function

$$R(x_i, x_j|\theta) = e^{-\theta(x_i - x_j)^2}$$

produces a local trend that is infinitely differentiable. This correlation function is called the Gaussian correlation function. Figure 1 shows four examples of the local trend produced by the Gaussian correlation for values of the correlation parameter $\theta = 2$, 4, and 10, respectively.

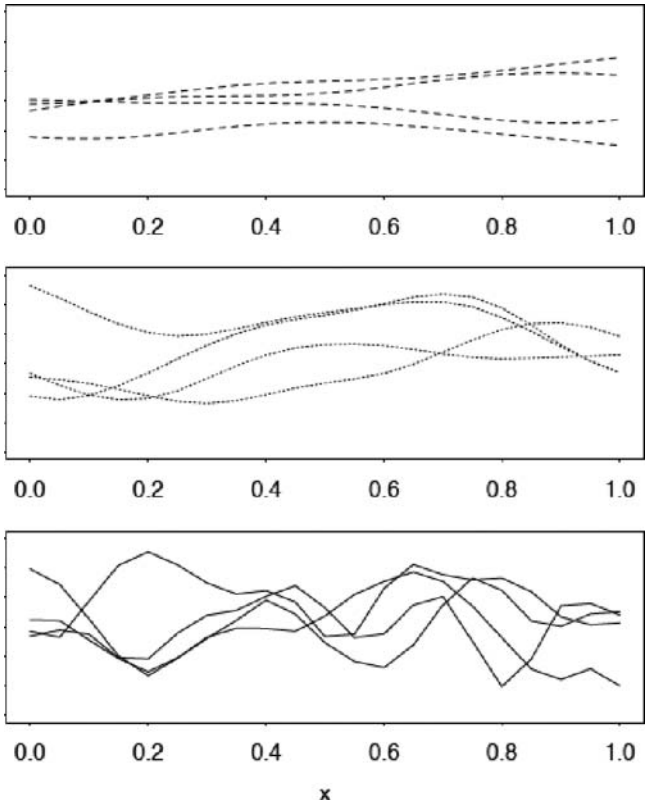

FIGURE 1 **The local trend produced by the Gaussian correlation with θ = 2, 4, and 10.**

In the context of a deterministic computer experiment we expect $Y(x)$ to be a smooth function of x and would model the correlation function in such a way as to produce a smooth local trend. Because the Gaussian correlation does this, it is a popular choice in computer experiments.

Prediction

In simple linear regression with independent errors, one typically estimates β_0 and β_1 by least squares. If we denote these estimates by $\widehat{\beta}_0$ and $\widehat{\beta}_1$ one usually predicts $Y(x)$ by $\widehat{Y}(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x$. If the errors are correlated one estimates β_0 and β_1 by generalized least squares and $\widehat{\beta}_0$ and $\widehat{\beta}_1$ in the above are the generalized least-squares estimates.

However, suppose we wish to predict $Y(x_0)$, given n observations taken at $x = x_1, x_2, \ldots, x_n$. According to our GaSP model with correlation function $R(\cdot, \cdot|\boldsymbol{\theta})$, the joint distribution of $Y(x_0), Y(x_1), \ldots, Y(x_n)$ is multivariate normal with mean $(\beta_0 + \beta_1 x_0, \beta_0 + \beta_1 x_1, \ldots, \beta_0 + \beta_1 x_n)^\top$ and covariance matrix

$$\sigma^2 \begin{pmatrix} 1 & \boldsymbol{r_0}^\top \\ \boldsymbol{r_0} & \boldsymbol{R} \end{pmatrix},$$

where $\boldsymbol{R}$ is the $n \times n$ matrix with i,jth entry $R(x_i, x_j|\boldsymbol{\theta}), 1 \le i,j \le n$, and $\boldsymbol{r_0}$ is the $n \times 1$ vector whose ith coordinate is $R(x_0, x_i|\boldsymbol{\theta}), 1 \le i \le n$. The conditional distribution of $Y(x_0)$ given $Y(x_1), \ldots, Y(x_n)$ (assuming $\beta_0, \beta_1, \sigma^2$, and $\boldsymbol{\theta}$ are known) is normal with mean

$$\beta_0 + \beta_1 x_0 + \boldsymbol{r_0}^\top \boldsymbol{R}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta}) \qquad [1]$$

where $\boldsymbol{Y} = (Y(x_1), \ldots, Y(x_n))^\top$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$, and

$$\boldsymbol{F} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

This follows from well-known properties of the multivariate normal distribution.

Given $\boldsymbol{\theta}$, if we replace β_0, β_1 in Eq. [1] with their generalized least-squares estimates, the conditional mean, as a predictor of $Y(x_0)$, turns out to be the best linear unbiased predictor (BLUP). Thus, in this simple setting, we have seen that the GaSP model and BLUP can be viewed as arising from regression models with correlated observations.

We now turn to the general case.

THE GENERAL CASE

The Model for Quantitative Variables Only

If the output of the simulator depends only on quantitative inputs $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top$, we model $y(\boldsymbol{x})$ as a realization of the random function

$$Y(\boldsymbol{x}) = \sum_{j=1}^{J} \beta_j f_j(\boldsymbol{x}) + Z(\boldsymbol{x}),$$

where the β_j are unknown (regression) parameters, the f_j are known (regression) functions, and $Z(\boldsymbol{x})$ is a mean zero, second-order stationary Gaussian process with variance σ_Z^2.

In addition, we assume

$$\mathrm{cov}(Z(\boldsymbol{x}_i), Z(\boldsymbol{x}_j)) = \sigma_Z^2 R(\boldsymbol{x}_i, \boldsymbol{x}_j),$$

where $R(\cdot, \cdot)$ is a valid correlation function, namely, $R(\boldsymbol{x}_i, \boldsymbol{x}_i) = 1$ and for any finite set of inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ the $n \times n$ matrix $\boldsymbol{R}$ whose $i,\ j$th entry is $R(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is positive definite.

There are many possible choices for the correlation function $R(\cdot, \cdot)$. A very popular choice is the Gaussian correlation function

$$R(\boldsymbol{x}_i, \boldsymbol{x}_j|\boldsymbol{\theta}) = \prod_{k=1}^{d} e^{-\theta_k (x_{i,k} - x_{j,k})^2}, \qquad [2]$$

where $\boldsymbol{x}_l = (x_{l,1}, x_{l,2}, \ldots, x_{l,d})^\top$ for $l = i, j$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)^\top$ are unknown parameters (often referred to as the correlation parameters) with $\theta_i \ge 0$ for all i. For purposes of this article, we assume that the correlation function is the Gaussian correlation function.

To predict $Y(\boldsymbol{x}_0)$ in this model, one choice is the BLUP, $\widehat{Y}(\boldsymbol{x}_0)$ (see Santner et al. 2003). As in the prediction subsection, the BLUP is also the conditional mean of $Y(\boldsymbol{x}_0)$ given all previously observed output from the simulator, with the β_j replaced by their generalized least-squares estimates. The BLUP requires that σ_Z^2 and $\boldsymbol{\theta}$ are known. If one substitutes estimates for σ_Z^2 and $\boldsymbol{\theta}$ into the BLUP, the resulting predictor is sometimes referred to as the empirical best linear unbiased estimator (EBLUP). Estimates might be based on maximum likelihood, restricted maximum likelihood, Bayes, or even cross-validation.

To be specific, the EBLUP is

$$\widehat{Y}(\boldsymbol{x}_0) = \boldsymbol{f}_0^\top \widehat{\boldsymbol{\beta}} + \boldsymbol{r}_0^\top \widehat{\boldsymbol{R}}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}), \qquad [3]$$

where

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{F}^\top \widehat{\boldsymbol{R}}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^\top \widehat{\boldsymbol{R}}^{-1}\boldsymbol{Y}.$$

$\boldsymbol{f}_0 = \left(f_1(\boldsymbol{x}_0), \ldots, f_J(\boldsymbol{x}_0)\right)^\top$, $\boldsymbol{r}_0$ is the $n \times 1$ vector of (estimated) correlations between $\boldsymbol{x}_0$ and each of the n inputs for which we have data , $\boldsymbol{F}$ is the $n \times J$ matrix with i, jth entry $f_j(\boldsymbol{x}_i)$, and $\widehat{\boldsymbol{R}}$ is the estimate of $\boldsymbol{R}$ obtained from estimates of the correlation parameters $\boldsymbol{\theta}$.

An estimator of the prediction variance of the EBLUP is (see Santner et al. 2003)

$$s^2(\boldsymbol{x}_0) = \widehat{\sigma}_Z^2[(1 - \boldsymbol{r}_0^\top \widehat{\boldsymbol{R}}^{-1}\boldsymbol{r}_0 + \boldsymbol{f}_0^\top - \boldsymbol{r}_0^\top \widehat{\boldsymbol{R}}^{-1}\boldsymbol{F})(\boldsymbol{F}^\top \widehat{\boldsymbol{R}}^{-1}\boldsymbol{F})^{-1}(\boldsymbol{f}_0^\top - \boldsymbol{r}_0^\top \widehat{\boldsymbol{R}}^{-1}\boldsymbol{F})^\top], \qquad [4]$$

where $\widehat{\sigma}_Z^2$ is an estimate of σ_Z^2.

Because the EBLUP is an interpolator (regardless of the form of the regression part of the model), and because realizations of $Z(\boldsymbol{x})$ can take on a wide variety of shapes, it is not uncommon to assume that the regression portion of the model is very simple. Many papers assume that the regression part is constant and use the constant mean model

$$Y(\boldsymbol{x}) = \beta + Z(\boldsymbol{x}).$$

Another option is to assume that the regression portion is just a linear trend.

The GaSP model and EBLUP have become a standard method for developing a statistical predictor or emulator. Unfortunately, for this model typical correlation functions (such as the Gaussian) assume that all of the inputs are quantitative. How can one incorporate qualitative variables into the GaSP model?

Models for Qualitative and Quantitative Variables: A Review

We now assume that the simulator produces deterministic output $y(\boldsymbol{x}, t)$ that depends on the quantitative input variables $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^\top$ and a qualitative variable having T levels, here indexed by t. One can interpret $y(\boldsymbol{x}, t)$ as determining t different curves or response surfaces, indexed by t.

If, in fact, we have $Q > 1$ qualitative variables, with the qth qualitative variable having T_q levels, assume that the $T = \prod_{q=1}^{Q} T_q$ possible combinations of levels are indexed by a single symbol taking on values from 1 to T (lexicographically ordered). This suppresses the inherent factorial structure, and we will return to this later.

We model $y(\boldsymbol{x}, t)$ as

$$Y(\boldsymbol{x}, t) = \sum_{j=1}^{p} \beta_j f_j(\boldsymbol{x}, t) + Z_t(\boldsymbol{x}), \qquad [5]$$

where the β_j are unknown (regression) parameters, the f_j are known (regression) functions, and $Z_t(\boldsymbol{x})$ is a mean zero, second-order stationary Gaussian process with variance $\sigma_{Z,t}^2$.

If we treat each value of t as determining a separate response surface, we can fit separate GaSP models to each of these response surfaces. However, if the response surfaces are similar, perhaps we can do better by developing predictors for each surface that "borrow" information from the other surfaces. This is similar to what one does in multiple regression by using indicator variables to represent different response surfaces. For example, one can use indicator variables to write a single multiple regression model representing several lines. This single model has more degrees of freedom for error than fitting separate lines. This comes at the expense of having to assume that the error variance is the same for each line.

In multiple regression, we use indicator variables to incorporate qualitative predictor variables into our models. Can we do this here? What happens if we simply add indicator variables to our model and act as though they are quantitative? The use of indicator variables in the regression portion of Eq. [5] (in particular, allowing some of the $f_j(\boldsymbol{x}, t)$ to involve indicator variables) presents no difficulties. Unfortunately, using indicator variables in the correlation function in a naïve manner does create problems. To see this, consider the following. For $1 \le t \le \mathrm{T}$ define

$$I_t(i) = \begin{cases} 1 & \text{if } i = t \\ 0 & \text{otherwise} \end{cases}. \qquad [6]$$

Treating these as quantitative, the Gaussian correlation function in Eq. [2] becomes

$$\begin{aligned} &R((\boldsymbol{x}_1,t_1),(\boldsymbol{x}_2,t_2)) \\ &= \prod_{l=1}^{T} e^{-\xi_l^*(I_l(t_1)-I_l(t_2))^2} \times \prod_{k=1}^{d} e^{-\xi_k(x_{1,k}-x_{2,k})^2} \\ &= e^{-\xi_{t_1}^*} \times e^{-\xi_{t_2}^*} \times \prod_{k=1}^{d} e^{-\xi_k(x_{1,k}-x_{2,k})^2} \\ &= \tau_{t_1}\tau_{t_2} \times \prod_{k=1}^{d} e^{-\xi_k(x_{1,k}-x_{2,k})^2}, \end{aligned}$$

where $\tau_{t_j} = e^{-\xi_{t_j}^*}$. Notice $0 < \tau_{t_j} < 1$.

Suppose $T = 4$ with response surfaces 1 and 2 very similar, response surfaces 3 and 4 very similar, but response surfaces 1 and 3 very different. In particular, suppose 1 and 2 are very similar in the sense that they are highly correlated implying that $\tau_1\tau_2$ is close to 1. In this case, both τ_1 and τ_2 must be close to 1. Similarly, if 3 and 4 are highly correlated, $\tau_3\tau_4$ should be close to 1 and hence both τ_3 and τ_4 must be close to 1. However, if 1 and 3 are essentially uncorrelated, we would expect $\tau_1\tau_3$ to be close to 0. But this is impossible if both τ_1 and τ_3 are close to 1. (Of course, this assumes that we can interpret the τ_i as correlations between response surfaces.)

This shows that if we naïvely use indicator variables to represent qualitative variables (as we would do in standard regression) we impose a certain structure on the "between response surfaces" correlations. Thus, a naïve approach using indicator variables to represent qualitative variables (at least in the correlation function) does not work well in general. So one has to model the qualitative variables more carefully, at least in terms of the correlation structure.

For simplicity, in what follows we assume a constant means model of the form

$$Y(\boldsymbol{x},t) = \beta_t + Z_t(\boldsymbol{x}).$$

One approach to incorporating indicator variables into the correlation function was suggested by Kennedy and O'Hagan (2000), although in a very specific context. They assumed that one has a collection of multifidelity computer simulations (simulations of differing degrees of accuracy), each involving the same quantitative factors, and that these simulations can be modeled collectively by a single computer model with a common set of quantitative factors and a qualitative factor to describe the different accuracy of the simulations. This approach implicitly assumes that as the level of the qualitative factor changes (increases) the fidelity increases. Thus, it may not be appropriate for the general case of incorporating a qualitative variable.

An approach that is popular in the literature was introduced in Qian et al. (2008) and further developed in Zhou et al. (2011). They assumed

$$\mathrm{Corr}(Z_{t_1}(\boldsymbol{x}_1), Z_{t_2}(\boldsymbol{x}_2)) = \tau_{t_1,t_2} \prod_{i=1}^{d} e^{-\xi_i(x_{1,i}-x_{2,i})^2}, \quad [7]$$

where τ_{t_1,t_2} is the cross-correlation between the response surfaces corresponding to categories t_1 and t_2 of the qualitative variable. The $T \times T$ matrix τ with r, sth entry $\tau_{r,s}$ must be a positive definite matrix with unit diagonal elements to guarantee that the correlation matrix whose i, jth entry $\mathrm{Corr}(Z_{t_i}(\boldsymbol{x}_i), Z_{t_j}(\boldsymbol{x}_j))$ is a valid correlation matrix. In addition, they assume $\sigma_t^2 = \sigma_Z^2$ for all t.

A hypersphere decomposition, originally introduced by Rebonato and Jackel (1999), was used to model τ in Zhou et al. (2011). This decomposition is quite useful for fitting the model and is constructed as follows. In Step 1, a Cholesky-type decomposition is applied to express $\boldsymbol{\tau}$ as $\boldsymbol{A}\boldsymbol{A}^\top$, where $\boldsymbol{A}$ is a lower triangular matrix with strictly positive diagonal entries and whose i, jth entry is $a_{i,j}$. In Step 2, each row vector $(a_{i,1}, a_{i,2}, \ldots, a_{i,i})$ is represented as the coordinates of a point on the surface of an i-dimensional unit hypershpere. In particular, $a_{1,1} = 1$, and for $i = 1, \ldots, T$

$$\begin{aligned} &a_{i,1} = \cos(\theta_{i,1}) \\ &a_{i,j} = \sin(\theta_{i,1})\sin(\theta_{i,2})\cdots\sin(\theta_{i,j-1})\cos(\theta_{i,j}) \text{ for } s = 2, \ldots, i-1 \\ &a_{i,i} = \sin(\theta_{i,1})\sin(\theta_{i,2})\cdots\sin(\theta_{i,i-2})\cos(\theta_{i,i-1}), \end{aligned}$$

where $\theta_{i,j} \in (0,\pi)$. Because the $\theta_{i,j}$ are restricted to $(0,\pi)$, the diagonal entries $a_{i,i}$ of $\boldsymbol{A}$ are strictly positive and $\boldsymbol{\tau} = \boldsymbol{A}\boldsymbol{A}^\top$ will be a valid correlation matrix (positive definite with unit diagonal entries). One can then estimate τ by estimating the $T(T-1)/2$ $\theta_{i,j}$ under the constraint that $\theta_{i,j} \in (0,\pi)$. Because $\theta_{i,j} \in (0,\pi)$, the off-diagonal entries in τ can be either positive or negative.

This approach appears to work well in practice. See Swiler et al. (2012) in which this method is compared to two others.

Several special cases of this model are mentioned by Qian et al. (2008). Each reduces the number of parameters to estimate to fit the model.

The simplest case assumes all $\tau_{t_i,t_j} = \tau$ for $i \neq j$. This is sometimes referred to as the exchangeable model.

Another case (see McMillan et al. 1999) assumes

$$\tau_{t_i,t_j} = e^{-(\xi_i^* + \xi_j^*)} I_i(j),$$

where ξ_i^* and ξ_j^* are positive and $I_i(j)$is the indicator function as defined in Eq. [6]. This model is almost equivalent to the Qian et al. (2008) model for $T \leq 3$, differing only in that the τ_{t_i,t_j} are restricted to be ≥ 0. For $T \geq 4$ the McMillan et al. (1999) model is only a special case of Qian et al. (2008). Note that we encountered the McMillan et al. (1999) structure previously when we simply used indicator variables to represent the qualitative variable. We saw that this had undesirable properties in general ($T \geq 4$).

A third special case is a Kronecker product structure. Suppose we have J qualitative variables, and the jth qualitative variable has T_j levels. Let $\boldsymbol{t} = (t_1, t_2, \ldots, t_J)^\top$ denote the vector of qualitative variable levels for an input that has qualitative variable j at level t_j for $j = 1, \ldots, J$. A legitimate correlation function is

$$\mathrm{Corr}(Z_{\boldsymbol{t}_1}(\boldsymbol{x}_1), Z_{\boldsymbol{t}_2}(\boldsymbol{x}_2)) = \prod_{j=1}^{J} \tau_{j,t_{1,j},t_{2,j}} \prod_{i=1}^{d} e^{-\xi_i \left(x_{1,i} - x_{2,i}\right)^2},$$

where $\boldsymbol{\tau}_j$, the $T_j \times T_j$ matrix with r,sth entry $\tau_{j,r,s}$, is positive definite with unit diagonal entries. This corresponds to taking $\boldsymbol{\tau} = \boldsymbol{\tau}_1 \otimes \cdots \otimes \boldsymbol{\tau}_J$, where $\otimes$ is the Kronecker product. This case reduces the number of τ_{t_i,t_j} parameters in the model. It also imposes a sort of multiplicative main effects structure on the τ_{t_i,t_j} and in this way takes into account the factorial structure.

Qian et al. (2008) consider additional forms for $\boldsymbol{\tau}$ and the τ_{t_i,t_j} that assume the levels of the qualitative factor can be organized into similar groups and that allow for ordinal qualitative factors.

The flexibility of the formulation in Qian et al. (2008) makes their model attractive. However, this model contains some implicit assumptions about the different response surfaces determined by t. In particular, in the correlation structure given in Eq. [7], the correlation parameters ξ_k and the process variance σ_Z^2 are the same for all values of t. This implies that the "shape" of the local variation as a function of the quantitative variables is the same for all t. If this is not the case, this model may perform worse than simply fitting separate GaSP models to curve (see Example 2 later in the article).

For the model of Qian et al. (2008), the EBLUP of Eq. [3], with $\widehat{\boldsymbol{R}}$ an estimate of the correlation matrix in Eq. [7], is used as an emulator. An estimate of the prediction variance is given by Eq. [4], again with $\widehat{\boldsymbol{R}}$ an estimate of the correlation matrix in Eq. [7].

A REEXAMINATION: ALTERNATIVE REPRESENTATIONS OF THE CORRELATION IN MODELS WITH QUANTITATIVE AND QUALITATIVE VARIABLES

Representations with Indicator Variables

Previously we argued that a naïve use of indicator variables does not work well in general. Thus, it may be a bit surprising that, in fact, it is possible to use indicator variables in the Gaussian correlation function to generate the correlation structure in Eq. [7] . One way is as follows. Let $I_p(i)$ be an indicator variable as in Eq. [6], and for $1 \leq p,\ q \leq T-1$ define

$$W_{p,q}(i) = \begin{cases} I_p(i) + I_q(i) & \text{if } p \neq q \\ I_p(i) & \text{if } p = q \end{cases}$$

and assume

$$\mathrm{Corr}(Z_{t_1}(\boldsymbol{x}_1), Z_{t_2}(\boldsymbol{x}_2)) = \prod_{p,q=1}^{T-1} e^{-\xi_{p,q}^* \left(W_{p,q}(t_1) - W_{p,q}(t_2)\right)^2} \times \prod_{k=1}^{d} e^{-\xi_k \left(x_{1,k} - x_{2,k}\right)^2}.$$

One can show with some algebra, assuming $\xi_{p,q}^* = \xi_{q,p}^*$ and $\tau_{i,j} > 0$, that for $i \neq j,\ i < T,\ j < T$

$$-\ln(\tau_{i,j}) = \xi_{i,i}^* + \xi_{j,j}^* - 4\xi_{i,j}^* + 2 \sum_{q=1, q \neq i}^{T-1} \xi_{i,q}^* + 2 \sum_{q=1, q \neq j}^{T-1} \xi_{j,q}^* \quad [8]$$

and for $i \neq j,\ i = T,\ j < T$

$$-\ln(\tau_{T,j}) = \xi^*_{j,j} + 2\sum_{q=1, q\neq j}^{T-1} \xi^*_{j,q} \qquad [9]$$

and for $i \neq j,\ i < T,\ j = T$

$$-\ln(\tau_{i,T}) = \xi^*_{i,i} + 2\sum_{q=1, q\neq i}^{T-1} \xi^*_{i,q}. \qquad [10]$$

Also, for $i \neq j,\ i < T,\ j < T$

$$\xi^*_{i,j} = \frac{1}{4}(\ln(\tau_{i,j}) - \ln(\tau_{T,j}) - \ln(\tau_{i,T})$$

and for $i < T$

$$\xi^*_{i,i} = -\frac{1}{2}\sum_{q=1, q\neq i}^{T} \ln(\tau_{i,q}) + \frac{1}{2}\sum_{q=1, q\neq i}^{T-1} \ln(\tau_{T,q}).$$

Thus for $\tau_{i,j} > 0$ there is a one-to-one correspondence between the $\tau_{i,j},\ i \neq j$ and the $\xi^*_{p,q},\ p < m,\ q < m$ in the sense that given the $\tau_{i,j}$ we can determine the corresponding $\gamma_{p,q}$ and vice versa.

This formulation with the variables $W_{p,q}(\cdot)$ allows one to use standard software for fitting the Gaussian correlation function to estimate the $\xi^*_{p,q}$ and ξ_k and then obtain the $\tau_{i,j}$ in the Qian et al. (2008) model, assuming all $\tau_{i,j} > 0$, using Eqs. [8], [9], and [10]. For example, this formulation can be used in JMP software package (version 11, SAS Institute Inc., Cary, NC) to fit models with quantitative and qualitative variables.

Another way to reformulate the Qian et al. (2008) model so that the correlation structure looks like one determined by the Gaussian correlation function is the following. In Kennedy and O'Hagan (2001), the Gaussian correlation function is expressed in the more general form

$$R(\boldsymbol{x}_i, \boldsymbol{x}_j | \Xi^*) = e^{-(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top \Xi^* (\boldsymbol{x}_i - \boldsymbol{x}_j)},$$

where Ξ^* is an unknown $d \times d$ positive definite symmetric matrix whose i,jth entry is $\xi^*_{i,j}$. This reduces to the form in Eq. [2] if Ξ^* is a diagonal matrix.

As before, let $I_p(i)$ be an indicator variable as in Eq. [6], define

$$\boldsymbol{I}(i) = (I_1(i), I_2(i), \ldots, I_T(i))^\top$$

and assume

$$\mathrm{Corr}(Z_{t_1}(\boldsymbol{x}_1), Z_{t_2}(\boldsymbol{x}_2)) = e^{-(\boldsymbol{I}(t_1) - \boldsymbol{I}(t_2))^\top \Xi^* (\boldsymbol{I}(t_1) - \boldsymbol{I}(t_2))} \times e^{-(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top \mathrm{diag}(\xi_1, \ldots, \xi_d)(\boldsymbol{x}_i - \boldsymbol{x}_j)}. \qquad [11]$$

In this formulation, one can show for $i \neq j$ and assuming all $\tau_{i,j} > 0$,

$$-\ln(\tau_{i,j}) = \xi^*_{i,i} + \xi^*_{j,j} - 2\xi^*_{i,j}.$$

Notice that this reduces to the case $\tau_{i,j} = \tau$ when Ξ^* is a multiple of the identity matrix, and it reduces to the McMillan et al. (1999) model when Ξ^* is a diagonal matrix.

One can define separate $T_j \times T_j$ matrices Ξ^*_j and obtain the Kronecker product formulation of Qian et al. (2008), with each Ξ^*_j corresponding to the $\boldsymbol{\tau}_j$.

This representation with indicator variables is less complicated than the first one we presented, but it requires one to write code to fit the more general form of the Gaussian correlation function in Eq. [11]. The first representation is convenient in that it can be used with software that fits the Gaussian correlation function in Eq. [2], which includes some commercial packages as well as existing software for fitting GaSP models. Notice that both of the representations with indicator variables are less general that the model of Qian et al. (2008) in that they force $\tau_{i,j} > 0$.

A Representation Inspired by a Characterization of the Multivariate Normal Distribution

A third way to incorporate qualitative variables, inspired by how one can characterize the multivariate normal distribution, is as follows. Let $N_1(\boldsymbol{x})$, $N_2(\boldsymbol{x})$, $\ldots$, $N_S(\boldsymbol{x})$ be S i.i.d. mean zero, second-order stationary Gaussian processes with variance σ^2_Z. Assume that each satisfies

$$\mathrm{Corr}(N_i(\boldsymbol{x}_1), N_i(\boldsymbol{x}_2)) = \prod_{k=1}^{d} e^{-\xi_i (x_{1,i} - x_{2,i})^2}$$

and for $1 \leq t \leq T$

$$Z_t(\boldsymbol{x}) = \sum_{i=1}^{S} a_{i,t} N_i(\boldsymbol{x}).$$

Then

$$(Z_1(\boldsymbol{x}),\ldots,Z_T(\boldsymbol{x}))^\top = \boldsymbol{A}(N_1(\boldsymbol{x},\ldots,N_S(\boldsymbol{x}))^\top,$$

where $\boldsymbol{A}$ is the $T \times S$ matrix with $a_{i,j}$ as its i,jth entry.

This yields the Qian et al. (2008) model provided that $\boldsymbol{\tau} = \boldsymbol{A}\boldsymbol{A}^\top$. Qian et al. (2008) use this representation to prove that τ must be a positive definite symmetric matrix with unit diagonal entries for the correlation structure in their model to be valid. But there is much more that can be done with this representation. Much of what we now say is inspired by multivariate normal methods.

The exchangeable model can be represented by

$$Z_i(\boldsymbol{x}) = \sqrt{\tau}N_1(\boldsymbol{x}) + \sqrt{1-\tau}N_{i+1}(\boldsymbol{x}).$$

This implies that the $Y(\boldsymbol{x},i)$ are composed of a common overall trend (the $N_1(\boldsymbol{x})$ term) and independent realizations of a "treatment effect" trend (the $N_{i+1}(\boldsymbol{x})$ terms). Both the overall trend and treatment effect trends are of the same magnitude ($\sqrt{\tau}$ and $\sqrt{1-\tau}$, respectively) for each $Y(\boldsymbol{x},i)$.

The McMillan (1999) model can be represented by

$$Y(x,i) = \tau_i N_1(x) + \sqrt{1-\tau_i^2}N_{i+1}(x),$$

This is similar to the exchangeable model, except the overall trend and treatment effect trends can be of different magnitudes for each $Y(\boldsymbol{x},i)$. Both the exchangeable model and the McMillan et al. (1999) model could be interpreted as implying a one-way analysis of variance structure or a constant mean (trend) structure plus noise.

To represent the Kronecker product structure, let

$$\boldsymbol{N}^j(\boldsymbol{x}) = (N_1^j(\boldsymbol{x}),\ldots,N_{T_j}^j(\boldsymbol{x}))^\top,$$

where the $N_i^j(\boldsymbol{x})$ are i.i.d. mean zero, second order stationary Gaussian processes with variance σ_Z^2. Let $\boldsymbol{A}_j$ be the $T_j \times T_j$ matrix satsfying $\boldsymbol{\tau}_j = \boldsymbol{A}_j\boldsymbol{A}_j^\top$. Then in our general formulation, $\boldsymbol{A}(N_1(\boldsymbol{x}),\ldots,N_S(\boldsymbol{x}))^\top$ becomes

$$\left(\otimes_{j=1}^J \boldsymbol{A}_j\right)\left(\otimes_{j=1}^J \boldsymbol{N}^j(\boldsymbol{x})\right).$$

One can also use this approach to model a factorial structure for the $Y(\boldsymbol{x},i)$. For example, suppose we believe the $Y(\boldsymbol{x},i)$ are determined by two factors F and G with f and g levels, respectively. Suppose $Y(\boldsymbol{x},i)$ corresponds to F at level ϕ and G at level γ, and thus

$$Y(\boldsymbol{x},i) \propto a_i^\mu N^\mu(\boldsymbol{x}) + a_i^F N_\phi^F(\boldsymbol{x}) + a_i^G N_\gamma^G(\boldsymbol{x}),$$

where $N^\mu(\boldsymbol{x})$ is an overall mean effect (trend), $N_\phi^F(\boldsymbol{x})$ is the effect of level ϕ of F, and $N_\gamma^G(\boldsymbol{x})$ the effect of level γ of G. This looks like a two-factor main effects only model.

An alternative is to include a small "error" effect $N_i^\epsilon(\boldsymbol{x})$, which gives

$$Y(\boldsymbol{x},i) \propto a_i^\mu N^\mu(\boldsymbol{x}) + a_i^F N_\phi^F(\boldsymbol{x}) + a_i^G N_\gamma^G(\boldsymbol{x}) + a_i^\epsilon N_i^\epsilon(\boldsymbol{x}).$$

One might think of a_i^ϵ as small relative to a_i^μ, a_i^F, and a_i^G.

If one does not require $\boldsymbol{A}$ to satisfy $\boldsymbol{\tau} = \boldsymbol{A}\boldsymbol{A}^\top$, then the formulation

$$(Y(\boldsymbol{x},1),\ldots,Y(\boldsymbol{x},T))^\top = \boldsymbol{A}(N_1(\boldsymbol{x},\ldots,N_S(\boldsymbol{x}))^\top$$

allows the $Y(\boldsymbol{x},i)$ to have different variances and hence some degree of nonstationarity.

Another application of this representation of the Qian et al. (2008) model is analogous to factor analysis. Estimate the matrix $\boldsymbol{\tau}$ and find a parsimonious matrix $\boldsymbol{A}$ so that $\boldsymbol{\tau} = \boldsymbol{A}\boldsymbol{A}^\top$. The form of $\boldsymbol{A}$ may suggest some sort of factorial structure or perhaps that the $Y(\boldsymbol{x},i)$ depend mostly on a relatively small number of the $N_j(\boldsymbol{x})$. Notice that in this representation of the Qian et al. (2008) model the matrix $\boldsymbol{A}$ is only determined up to multiplication by an orthogonal matrix, and one can use some of the algorithms in factor analysis to search for parsimonious forms for $\boldsymbol{A}$. We will do this in the examples in the next section.

We note that in the hypersphere parameterization of $\boldsymbol{\tau}$ in Zhou et al. (2011), they use $\boldsymbol{\tau} = \boldsymbol{A}\boldsymbol{A}^\top$ with $\boldsymbol{A}$ being lower triangular. This might be an initial estimate for $\boldsymbol{A}$ to which various rotations can be applied.

EXAMPLES

Example 1

Zhou et al. (2011) consider a simple example involving a single quantitative variable and a single qualitative

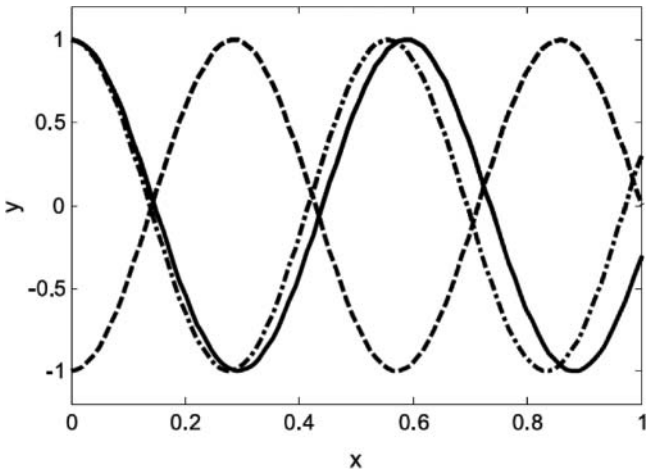

FIGURE 2 **Function curves at $t=1(-)$, $t=2(--)$, and $t=3(-\cdot)$.**

variable with three levels, defining the three curves,

$$y(x,t)=\begin{cases}\cos\left(\dfrac{6.8\pi x}{2}\right) & \text{if } t=1\\ -\cos\left(\dfrac{7\pi x}{2}\right) & \text{if } t=2\,.\\ \cos\left(\dfrac{7.2\pi x}{2}\right) & \text{if } t=3\end{cases}$$

These curves are displayed in Figure 2.

Zhou et al. (2011) use a Latin hypercube design to observe (the same) eight points on each curve. Using the general model to estimate the correlation parameters, Zhou et al. (2011) find $\tau_{1,2}=-0.95$, $\tau_{1,3}=0.81$, and $\tau_{2,3}=-0.93$.

Using the McMillan et al. (1999) model to estimate the correlation parameters they find $\tau_{1,2}=0.01$, $\tau_{1,3}=0.88$, and $\tau_{2,3}=0.01$. Notice that with only three curves ($T<4$), the McMillan et al. formulation only differs from the general model in that the τ must be nonnegative.

Using our reparameterization based on indicator variables, we define

$$\begin{aligned}&W_{1,1}(1)=1,\ W_{1,1}(2)=0,\ W_{1,1}(3)=0\\ &W_{1,2}(1)=1,\ W_{1,2}(2)=1,\ W_{1,2}(3)=0\\ &W_{2,1}(1)=1,\ W_{2,1}(2)=1,\ W_{2,1}(3)=0\\ &W_{2,2}(1)=0,\ W_{2,2}(2)=1,\ W_{2,2}(3)=0.\end{aligned}\quad [12]$$

We fit the model in JMP, and solving for the $\tau_{i,j}$, we find $\tau_{1,2}=0$, $\tau_{1,3}=0.93$, and $\tau_{2,3}=0$. This should be equivalent to the McMillan et al. (1999) model (because $T<4$) and indeed the results are almost identical.

For the general model in Zhou et al. (2011), a simple factor analysis (using Minitab) applied to the correlation matrix yields

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Var 1	0.954	0.295	0.056	1.000
Var 2	−0.995	−0.022	0.097	1.000
Var 3	0.946	−0.321	0.045	1.000
Variance	2.7947	0.1907	0.0146	3.0000
%Var	0.932	0.064	0.005	1.000

Here the rows labeled Var1, Var2, and Var3 correspond to the three curves. The columns labeled Factor 1, Factor 2, and Factor 3 correspond to the underlying $N_i(x)$. This analysis suggests that there is a single dominant underlying curve. This underlying curve determines curves 1 and 3. The negative of this underlying curve determines curve 2. This agrees nicely with what we see in the plot.

For the McMillan et al. (1999) model, a simple factor analysis applied to the correlation matrix yields

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Var 1	0.969	0.011	−0.245	1.000
Var 2	0.022	−1.000	0.000	1.000
Var 3	0.969	0.011	0.245	1.000
Variance	1.8802	0.9998	0.1200	3.0000
%Var	0.627	0.333	0.040	1.000

For our reparameterization a simple factor analysis applied to the correlation matrix yields

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Var 1	0.982	0.000	0.187	1.000
Var 2	0.000	1.000	0.000	1.000
Var 3	0.982	0.000	−0.187	1.000
Variance	1.9300	1.0000	0.0700	3.0000
%Var	0.643	0.333	0.023	1.000

The McMillan et al. (1999) and our reparameterization approaches suggest there are two dominant underlying curves. One of these essentially determines curves 1 and 3. The other essentially determines curve 2. A third underlying curve accounts for differences in curves 1 and 3.

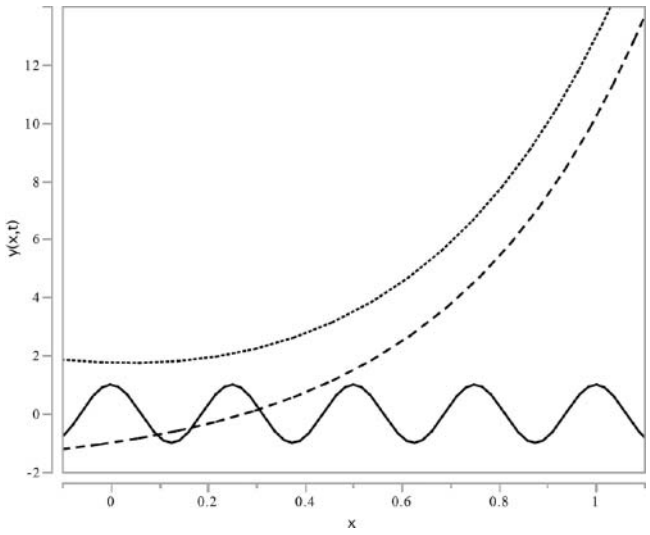

FIGURE 3 **The dotted line is $y(x,1)$, the dashed line is $y(x,2)$, and the solid line is $y(x,3)$.**

The overall fit in all three cases is quite good, but the general method in Qian et al. (2008), allowing the curves to be negatively correlated, is more informative than the parameterizations restricting the correlations to be nonnegative.

Example 2

Here is another simple example involving three curves (a single qualitative variable with three levels). See Figure 3 for a plot of these three curves.

$$y(x,t)=\begin{cases}\exp(2.5x)+3(x-0.5)^2 & \text{if } t=1\\ \exp(2.5x)-2 & \text{if } t=2\\ 2\cos(8\pi x) & \text{if } t=3\end{cases}.$$

All three methods give very similar estimates for the correlation parameters and the results from our first reparameterization using indicator variables (the same as in Eq. [12]) give $\tau_{1,2}=0.89$, $\tau_{1,3}=0.06$, and $\tau_{2,3}=0.07$.

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
Var 1	0.969	0.077	−0.234	1.000
Var 2	0.970	0.066	0.235	1.000
Var 3	0.140	−0.990	−0.003	1.000
Variance	1.8994	0.9907	0.1099	3.0000
%Var	0.633	0.330	0.037	1.000

This analysis suggests that there are two dominant underlying curves. The first determines curves 1 and 2. The second determines curve 3. A third underlying curve accounts for differences between curves 1 and 2. Again, this is what one would probably expect from the plot of the three curves.

One of the assumptions implicit in the Qian et al. (2008) model is that the correlation parameter for the quantitative variable is the same for all three curves. This may be true for the first two curves, but the correlation parameter for the third curve should be substantially larger than the other two. This is confirmed if one fits independent models to the three curves. Fitting separate kriging models to each curve, the fits for the first two are excellent, but the third is poor due to insufficient data. If we fit the Qian et al. (2008) model, the fit of the first two curves is worse (the general trend is captured but there is small oscillation around the trend), but the third is much improved (it looks very much like a cosine curve but the amplitudes at each peak vary a bit). The differences in fits for the first two curves are probably due to the fact that assuming the same correlation parameter (in the Gaussian correlation function) for the quantitative input first two curves is reasonable, but the correlation parameter for quantitative input for the third curve is larger than for the first two. The Qian et al. (2008) model mistakenly "borrows" information from the third curve in fitting the first two; hence the poorer fit for the first two.

CONCLUSIONS

We have reviewed a popular method for incorporating qualitative and quantitative input variables into GaSP models. We have introduced two alternative representations that use indicator variables. One of these has the attractive property that it allows one to fit the model using existing software for fitting the quantitative inputs-only model. We also proposed a third representation inspired by one way of characterizing the multivariate normal distribution. This latter representation can allow one to incorporate factorial structure into the model. It also lends itself to a kind of factor analysis, enabling one to determine if there is a simple underlying structure in the general model. We used two simple examples to illustrate the proposed representations.

There are two issues that we continue to pursue in ongoing research. First, for the third representation,

code for fitting special forms and code for fitting forms of $\boldsymbol{A}$ that allow for nonstationarity is being investigated. Second, what constitutes a good experimental design for fitting models with qualitative and quantitative input variables. Recent papers by Qian (2012) propose special types of Latin hypercube designs, but we believe more work can be done on this question.

Before ending this section, we should mention a Bayesian approach for incorporating qualitative variables. Han et al. (2009) assume the $Z_t(\cdot)$ are mutually independent

$$\begin{aligned}\text{Corr}(Z_t(\boldsymbol{x}_1),\, Z_t(\boldsymbol{x}_2)) &= \prod_{k=1}^{d} e^{-\xi_{t,k}\left(x_{i,k}-x_{j,k}\right)^2} \\ &= \prod_{k=1}^{d} \rho_{t,k}^{\left(x_{1,k}-x_{2,k}\right)^2}\end{aligned}$$

with certain priors on the β_t, the σ_t^2, and the $\rho_{t,k} = e^{-\xi_{t,k}}$ and through the priors's attempts to capture similarities between between the response surfaces $y(\boldsymbol{x}, t)$.

ACKNOWLEDGMENTS

We thank the organizers of the 2014 Stu Hunter Research Conference for inviting us to speak on this topic and the journal *Quality Engineering* for the opportunity to publish this talk. We also thank our discussants for encouragement and very valuable feedback.

ABOUT THE AUTHORS

Yulei Zhang received her Ph.D. in statistics from the Ohio State University in 2014. Dr. Zhang's research interests are the design and analysis of computer experiments. She has recently taken a position with Nationwide Insurance in Columbus, Ohio.

William I. Notz is Professor of Statistics at the Ohio State University. He received his Ph.D. in mathematics from Cornell University. Professor Notz's research interests have focused on experimental design and computer experiments. He is the author of several research papers and of a book on the design and analysis of computer experiments. He is an elected fellow of the American Statistical Association. He has served as the editor of the journal *Technometrics* and as editor of the *Journal of Statistics Education*.

REFERENCES

Han, G., Santner, T. J., Notz, W. I., Bartel, D. L. (2009). Prediction for computer experiments having quantitative and qualitative input variables. *Technometrics*, 51(3): 278–288.

Kennedy, M. C., O'Hagan, A. (2000). Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87: 1–13.

Kennedy, M. C., O'Hagan, A. (2001). Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society B*, 63: 425–464.

McMillan, N. J., Sacks, J., Welch, W. J., Gao, F. (1999). Analysis of protein activity data by Gaussian stochastic process models. *Journal of Biopharmeceutical Statistics*, 9: 145–160.

Qian, P. Z. G. (2012). Sliced latin hypercube designs. *Journal of the American Statistical Association*, 107(497): 393–399.

Qian, P. Z. G., Wu, H., Wu, C. F. J. (2008). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics*, 50: 383–396.

Rebonato, R., 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.

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

Swiler, L. P., Hough, P. D., Qian, P., Xu, X., Storlie, C. B., Lee, H. (2014). Surrogate models for mixed discrete-continuous variables. *Constraint Programming and Decision Making: Studies in Computational Intelligence*, 539: 181–202.

Zhou, Q., Qian, P. Z. G., Wu, H., Zhou, S. (2011). A simple approach to emulation for computer models with qualitative and quantitative factors. *Technometrics*, 53: 266–273.