Received: 10 October 2019 | Revised: 30 December 2019 | Accepted: 2 January 2020

DOI: 10.1002/widm.1358

ADVANCED REVIEW

WIREs DATA MINING AND KNOWLEDGE DISCOVERY

WILEY

Design and analysis of computer experiments with quantitative and qualitative inputs: A selective review

Xiaoning Kang[1] | Xinwei Deng[2]

[1]International Business College and Institute of Supply Chain Analytics, Dongbei University of Finance and Economics, Dalian, China

[2]Department of Statistics, Virginia Tech, Blacksburg, Virginia

Correspondence
Xinwei Deng, Department of Statistics, Virginia Tech, Blacksburg, VA.
Email: xdeng@vt.edu

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: 71531004, 71903024

Abstract

Computer experiment refers to the study of complex systems by using computer simulations to emulate the physical system. Design and analysis of computer experiments have attracted great attention in past decades. However, many computer experiments involve not only quantitative inputs, but also qualitative inputs, which make the design and analysis more challenging. The Latin hypercube design and its variants are widely used in computer experiments, but mainly for the quantitative inputs. Constructing desirable emulators for computer experiments with qualitative inputs also remains a challenging problem due to the discrete nature of qualitative inputs. In this review, we describe a set of statistical approaches for design and analysis of computer experiments with both quantitative and qualitative factors.

This article is categorized under:

Fundamental Concepts of Data and Knowledge > Key Design Issues in Data Mining

Algorithmic Development > Statistics

KEYWORDS

computer experiment, Gaussian process, uncertainty quantification

1 | INTRODUCTION

In many scientific and engineering applications, physical experiments are not feasible or can be very difficult to perform in terms of material, time, and cost. For example of global warming experiments, it is almost impossible to conduct experiments on an earth sized object. In epidemiology study, the number of input variables is too large to implement physical experiments in practice. In high-speed train crash-prevention tests, the experiments required to gather sufficient information are economically prohibit to run because of the expensive trains. For these situations, computer experiments (Fang, Li, & Sudjianto, 2005; Santner, Williams, & Notz, 2018) often replace physical experiments to study the relationship between a set of input variables and the resulting outputs. The design and analysis of computer experiments have received wide attention in the past decades, see Sacks, Welch, Mitchell, and Wynn (1989); Kennedy and O'Hagan (2000); Hobert, Jones, Presnell, and Rosenthal (2002); Higdon, Gattiker, Williams, and Rightley (2008); Gramacy (2012); Kong, Ai, and Tsui (2018) among many others. Santner et al. (2018) is the most latest book summarizing the technical development of computer experiments. It introduces designs and analysis commonly used for research investigations with computer simulator platforms. One of its subsection describes methods for designing and analyzing computer experiments for quantitative and qualitative (QQ) inputs, which were proposed in literature around year 2010. Compared with this book, the article mainly focuses on reviewing several newly developed computer experimental designs with QQ inputs, as well as analysis tools, including sliced Latin hypercube designs (SLHDs), marginally coupled designs, additive Gaussian process (GP) model and so on.

Regarding the designs of computer experiments, one would like to run a set of experiments that provide information on all portions of the experimental region. This leads to the space-filling designs that spread the input points evenly throughout the full region. The Latin hypercube designs (LHDs) introduced by Mckay, Beckman, and Conover (1979) are extensively used as space-filling designs in computer experiments (Ba, Myers, & Brenneman, 2015; Hung, 2011a; Joseph, 2016; Wang, Xiao, & Xu, 2018). A good property of the LHDs is that the design points are spread evenly across the range of any individual input variable. In addition, if a few input variables are removed from the LHD, the resulting design is still a LHD, which facilitates the case where the output turns out to depend on a subset of the input variables. A number of papers studied the sampling properties of the LHDs, including Mckay et al. (1979), Stein (1987), Owen (1992), Loh (1996), and so forth. To further enhance the projection property of designs, Joseph, Gul, and Ba (2015) proposed a maximum projection (MaxPro) design that ensures good projections to all subspaces of the input variables.

However, the LHDs are designated for quantitative factors (or input variables), but not suitable for qualitative factors. Suppose that the input region is the p-dimensional unit sphere $[0, 1]^p$. Let $\boldsymbol{L} = (l_{ij})$ be an $n \times p$ Latin hypercube with each of its column a permutation on $\{1, 2, ..., n\}$ and all the columns are obtained independently. A LHD $\boldsymbol{D} = (d_{ij})$ of n runs for p factors is built as

$$d_{ij} = \left(l_{ij} - u_{ij}\right)/n, \text{ for } i = 1,2,...,n, \text{ and } j = 1,2,...,p,$$

where u_{ij} are independently generated from the uniform distribution $U[0, 1)$, d_{ij} indicate the level for the input variable j on the ith run, and u_{ij} and l_{ij} are mutually independent. When $\boldsymbol{D}$ is projected onto any one dimension of input variables, precisely one design point falls within one of the n-equally spaced intervals of (0, 1] given by $(0, 1/n]$, $(1/n, 2/n]$, ..., $(1–1/n, 1]$.

Table 1 is an example of a Latin hypercube (in transpose) with eight design points for three factors x_1, x_2, and x_3. It is seen that each row, corresponding to each factor, is a permutation of $\{1, 2, ..., 8\}$. Figure 1 shows the projection of this Latin hypercube onto any two input variables. The eight design points are space-equally distributed on any single dimensionality of factors.

Clearly, such a construction method of LHD can not be easily extended to design a computer experiment for both QQ factors, which are involved in many applications. For the example of data center thermal management (Jiang,

TABLE 1 A Latin hypercube with three input variables

x_1	4	2	5	6	7	1	8	3
x_2	1	2	5	7	4	6	8	3
x_3	2	3	1	5	8	4	7	6

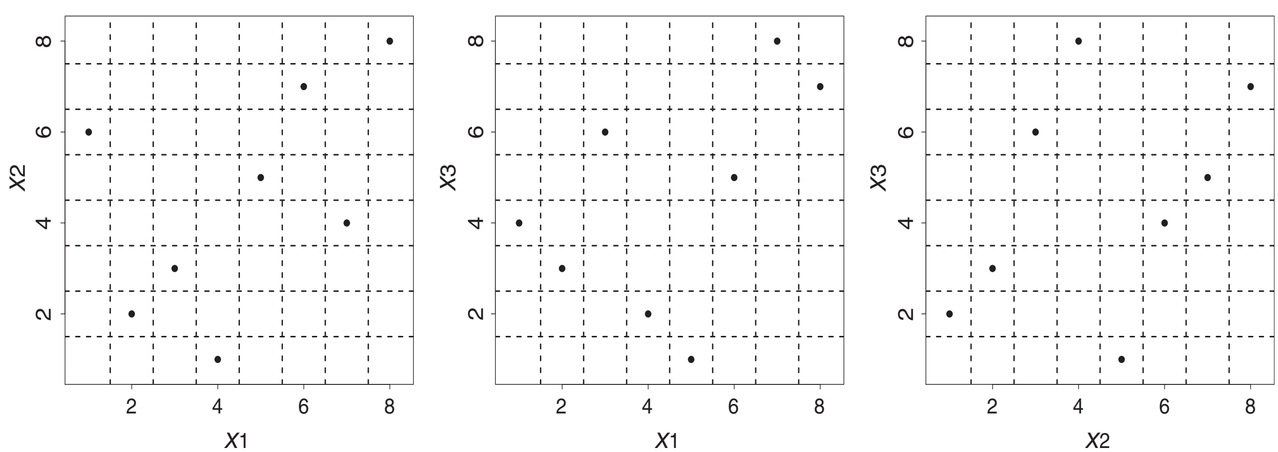

FIGURE 1 Projections of the Latin hypercube design in Table 1 onto two factors

Deng, López, & Hamann, 2016), the computational fluid dynamics for studying the temperature distribution often contains qualitative factors such as "hot air return vent location" and "power unit type" (Qian, Wu, & Wu, 2008). The investigation of wear mechanisms of total knee replacements in bioengineering (Han, Santner, Notz, & Bartel, 2009) uses the knee models with qualitative factors such as "prosthesis design" and "force pattern."

Moreover, modeling and analyzing the data coming from computer experiments with QQ factors are not straightforward. When only quantitative factors involved, the GP model is the most commonly used tool to analyze the computer experiments (Currin, Mitchell, Morris, & Ylvisaker, 1991; Hung, 2011b; Li & Sudjianto, 2005; Linkletter, Bingham, Hengartner, Higdon, & Ye, 2006; Martin & Simpson, 2004; McMillian, Sacks, Welch, & Gao, 1999). In the standard GP model, the inputs are all quantitative and the outputs can be viewed as realizations of a GP. When having qualitative factors in the computer experiments, one can not directly treat the outputs as realizations of a GP due to the discrete nature of qualitative factors.

2 | DESIGN OF COMPUTER EXPERIMENTS WITH QQ INPUTS

In the literature, there are several works investigating on how to construct designs for computer experiments with QQ factors. One simple approach is the SLHDs proposed by Qian (2012). An SLHD with $n = rm$ runs for p factors has the property that it can be divided into r LHDs of size $m \times p$, where r and m are integers. These r LHDs are called slices and the original n-run LHD is an SLHD. It is easy to see that SLHD can be used for a computer experiment with quantitative factors and a set of qualitative factors having r distinct levels of combinations. Each slice provides a space-filling design for m runs corresponding at one fixed level combination of the qualitative variables. However, as the number of qualitative factors increases, the number of level combinations would increase dramatically. As a result, the SLHD can only be suitable for a small number of qualitative factors. In Huang, Lin, Liu, and Yang (2016), one kind of SLHDs was considered with points clustered in the design region for computer experiments with QQ factors. But the run size of such designs increases significantly with the number of qualitative variables.

Another approach to accommodate the computer experiments for QQ factors is the marginally coupled design introduced by Deng, Hung, and Lin (2015), which is more economical than the SLHD with attractive space-filling properties. A marginally coupled design has two subdesigns, denoted by D_1 and D_2, with D_1 a design for qualitative factors and D_2 a design for quantitative factors. The design $D = (D_1, D_2)$ is a marginally coupled design if D_2 is a SLHD with respect to each column of D_1 (an orthogonal array is typically used for D_1). The marginally coupled design has two features: the design points for quantitative factors form a LHD; and for each level of any qualitative factors, the corresponding design points for quantitative factors form a small LHD.

Table 2 presents a marginally coupled design $D = (D_1, D_2)$ of nine runs involving two quantitative variables (x_1, x_2) and two qualitative factors (z_1, z_2) with each having three levels. Figure 2 displays the scatter plots of x_1 versus x_2. Rows of D_2 corresponding to levels 0,1,2 of z_1 or z_2 are represented by $\times$, $\circ$, and $+$. Projected onto x_1 or x_2, three points represented by $\times$ or $\circ$ or $+$ are located exactly in each of three intervals [0,1/3), [1/3,2/3), [2/3,1).

Deng et al. (2015) also studied the existence of such marginally coupled designs. For a given $n \times q$ design D_1, a marginally coupled design $D = (D_1, D_2)$ exists if there exists an $n \times p$ design D_2 with $p > 0$ such that D is a marginally coupled design. Proposition 1 in Deng et al. (2015) shows that when D_1 is an orthogonal array OA$(n, s^q, 2)$, a marginally

TABLE 2 A marginally coupled design $D = (D_1, D_2)$

x_1	x_2	z_1	z_2
0	0	0.79	0.82
0	1	0.10	0.60
0	2	0.46	0.08
1	0	0.20	0.20
1	1	0.36	0.88
1	2	0.96	0.42
2	0	0.63	0.48
2	1	0.69	0.21
2	2	0.22	0.75

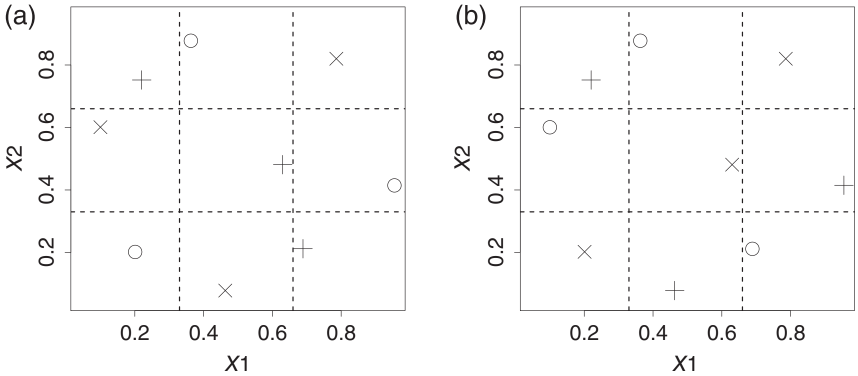

FIGURE 2 Scatter plots of x_1 versus x_2, where rows of D_2 corresponding to levels 0,1,2 of z_i are marked by $\times$, $\circ$, and $+$: (a) the levels of z_1; (b) the levels of z_2

coupled design exists if and only if D_1 is completely resolvable. Here, an s-level orthogonal array A of strength t is an $n \times m$ matrix with each column taking s distinct levels and for every $n \times t$ submatrix of A, each of all possible level combinations appears equally often (Hedayat, Sloane, & Stufken, 1999). Such an orthogonal array is denoted by OA(n, s^m, t). An OA(n, s^m, 2), say A, is said to be α-resolvable if it can be expressed as $A = ({A_1}^{\mathrm{T}}, \ldots, {A_{n/(s\alpha)}}^{\mathrm{T}})^{\mathrm{T}}$ such that each of $A_1, \ldots, A_{n/(s\alpha)}$ is an OA($s\alpha$, s^m, 1). If $\alpha = 1$, the orthogonal array is called completely resolvable.

Such a result establishes the necessary and sufficient conditions of the existence of marginally coupled designs when D_1 is an OA(n, s^q, 2). Furthermore, it reaches a conclusion on the maximum number of columns in an s-level orthogonal array of n runs for which a marginally coupled design exists. This is because an s-level completely resolvable orthogonal array of n runs has at most n/s columns (Suen, 1989). Specifically, if q^* represents the maximum value of q such that a marginally coupled design $D = (D_1, D_2)$ with D_1 = OA(n, s^q, 2) exists, then we have $q^* \le n/s$.

The details of construction methods for marginally coupled design can be found in Deng et al. (2015). The variants of marginally coupled design are also developed in other works. He, Lin, Sun, and Lv (2017) constructed marginally coupled designs when all qualitative factors have two levels. He, Lin, and Sun (2017) provided more efficient construction approaches for marginally coupled designs and derived the theoretical results. To allow flexible run size, Joseph, Gul, and Ba (2019) extended the MaxPro criterion (Joseph et al., 2015) for designing computer experiments with different types of factors, including continuous, nominal, discrete numeric, and ordinal input variables. Their proposed design is able to accommodate large number of qualitative factors with good space-filling properties.

3 | ANALYSIS OF COMPUTER EXPERIMENTS WITH QQ INPUTS

In this section, we review the current literature on how to model the computer experiments for the QQ factors. Consider an n-run computer experiment with p quantitative factors and q qualitative factors. Denote the ith quantitative factor as $x^{(i)}$ ($i = 1, \ldots, p$) and the jth qualitative factor as $z^{(j)}$ ($j = 1, \ldots, q$). There are m_j levels ($1, \ldots, m_j$) of the qualitative factor $z^{(j)}$. Let the kth ($k = 1, \ldots, n$) input data be $\mathbf{w}_k = \left(\mathbf{x}_k^{\mathrm{T}}, \mathbf{z}_k^{\mathrm{T}}\right)^{\mathrm{T}}$, where $\mathbf{x}_k = (x_{k1}, \ldots, x_{kp})^{\mathrm{T}} \in \mathbb{R}^p$ is the quantitative part and $\mathbf{z}_k = (z_{k1}, \ldots, z_{kq})^{\mathrm{T}} \in \mathbb{R}^q$ is the qualitative part of the input. Denote $Y(\mathbf{w}_k)$ as the output from the input $\mathbf{w}_k$ and the response (or output) vector $\mathbf{y} = (Y(\mathbf{w}_1), \ldots, Y(\mathbf{w}_n))^{\mathrm{T}}$.

The Gaussian process model is a common tool to analyze computer experiments. A standard GP (Kriging) model quantifying the relationship between output $Y(\mathbf{x})$ and quantitative inputs $\mathbf{x}$ assumes

$$Y(\mathbf{x}) = \mu + G(\mathbf{x}),$$

where μ is the constant trend. Here, $G(\mathbf{x})$ is a GP with mean zero and the covariance function $\phi(\cdot) = \sigma^2 R(\cdot|\boldsymbol{\theta})$, where σ^2 is the variance, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^T$ is the range (or correlation) parameters with all $\theta_k > 0$ ($k = 1, \ldots, p$), and $R(\cdot|\boldsymbol{\theta})$ is the correlation determined by a stationary correlation function, for example, Gaussian, power-exponential, Matérn (Rasmussen & Williams, 2006), and lifted Brownian (Plumlee & Apley, 2017) correlation functions. A popular choice for $R(\cdot|\boldsymbol{\theta})$ is the Gaussian correlation function

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

which represents the correlation between $G(\mathbf{x}_1)$ and $G(\mathbf{x}_2)$ for any two input data $\mathbf{x}_i = (x_{i1}, ..., x_{ip})^T$ and $\mathbf{x}_j = (x_{j1}, ..., x_{jp})^T$. The parameters $\boldsymbol{\theta}$ can be estimated via maximum likelihood estimation, along with μ and σ^2 in the model. For more details on standard GP models, see Sacks et al. (1989) and Kleijnen (2009).

However, the standard GP model does not work directly for computer experiments with QQ factors. This is because the correlation function $R(\cdot|\boldsymbol{\theta})$ does not take into account the fact that different level combinations of qualitative factors may not have specific distance measurement. To deal with QQ inputs, several papers proposed to use a multiplicative covariance function for any two inputs $\mathbf{w}_1$ and $\mathbf{w}_2$ as (Qian et al., 2008; Zhou, Qian, & Zhou, 2011)

$$\phi(Y(\mathbf{w}_1), Y(\mathbf{w}_2)) = \mathrm{Cov}(Y(\mathbf{w}_1), Y(\mathbf{w}_2)) = \sigma^2 \prod_{j=1}^{q} \tau_{z_{1j}z_{2j}}^{(j)} R(\mathbf{x}_1, \mathbf{x}_2 | \boldsymbol{\theta}), \tag{2}$$

where the parameter $\tau_{z_{1j}z_{2j}}^{(j)}$ represents the correlation between two levels (z_{1j} and z_{2j}) in the *jth* qualitative factor $z^{(j)}$, and $R(\mathbf{x}_1, \mathbf{x}_2|\boldsymbol{\theta})$ is defined as in (1). Denote the correlation matrix for $z^{(j)}$ ($j = 1, ..., q$) by $\mathbf{T}_j = \left(\tau_{z_{1j}z_{2j}}^{(j)}\right)_{m_j \times m_j}$. Three different functions of $\tau_{z_{1j}z_{2j}}^{(j)}$ were suggested in literature as follows:

a. The exchangeable correlation (EC) function: $\tau_{z_{1j}z_{2j}}^{(j)} = c$ $(0 < c < 1)$ when $z_{1j} \neq z_{2j}$; otherwise, $\tau_{z_{1j}z_{2j}}^{(j)} = 1$ (Joseph & Delaney, 2007).
b. The multiplicative correlation (MC) function: $\tau_{z_{1j}z_{2j}}^{(j)} = \exp\{-(\theta_{z_{1j}} + \theta_{z_{2j}})\}$ $(\theta_{z_{1j}}, \theta_{z_{2j}} > 0)$ when $z_{1j} \neq z_{2j}$; otherwise, $\tau_{z_{1j}z_{2j}}^{(j)} = 1$ (McMillian et al., 1999).
c. The unrestrictive correlation (UC) function: define $T_j = L_j L_j^{\mathrm{T}}$ where L_j is a lower triangular matrix; for the *rth* row $(l_{r1}^{(j)}, ..., l_{rr}^{(j)})$ in L_j, $l_{11}^{(j)} = 1$, and for $r = 2, ..., m_j$.

$$\begin{cases} l_{r1}^{(j)} = \cos\left(\varphi_{j,r,1}\right) \\ l_{rs}^{(j)} = \sin\left(\varphi_{j,r,1}\right) ... \sin\left(\varphi_{j,r,s-1}\right) \cos\left(\varphi_{j,r,s}\right) \quad \text{for } s = 2, ..., r-1 \\ l_{rr}^{(j)} = \sin\left(\varphi_{j,r,1}\right) ... \sin\left(\varphi_{j,r,r-1}\right), \end{cases}$$

where $\varphi_{j,\, r,\, s} \in (0, \pi)$ for $s = 1, ..., r - 1$ (Qian et al., 2008; Zhou et al., 2011).

A thorough discussion on the choices of the correlation matrix $R(\cdot|\boldsymbol{\theta})$ in the GP model for QQ factors can be found in Roustant, Padonou, Deville, Clément, and Wynn (2018) and Zhang and Notz (2015).

Note that, in the use of the multiplicative correlation function, a zero value of any $\tau_{z_{1j}, z_{2j}}^{(j)}$ would result in the overall correlation $\mathrm{Cov}(Y(\mathbf{w}_1), Y(\mathbf{w}_2))$ being equal to zero. To overcome this problem, Deng, Lin, Liu, and Rowe (2017) proposed an additive GP model as

$$Y(\mathbf{x}, z_1, ..., z_q) = \mu + G_1(z_1, \mathbf{x}) + \cdots + G_q(z_q, \mathbf{x}), \tag{3}$$

where μ is the overall mean, G_j's are independent GPs with mean zeroes and the covariance function ϕ_j, for $j = 1, ..., q$. The additive form is employed in (3) to quantify the contribution of q qualitative input factors to the output. Such an additive form emphasizes the effect of each qualitative factor coupled with quantitative factors. Besides, the additive formulation enables to infer the significance of each individual qualitative factor in the model. The model (3) contains interactions between qualitative factors and quantitative factors by each GP component $G_i(z_i, \boldsymbol{x})$, and interactions among quantitative factors by the correlation function $R(\cdot|\boldsymbol{\theta}^{(i)})$ in each $G_i(z_i, \boldsymbol{x})$. One can also easily extend its current form to accommodate the interactions among qualitative factors.

Given the model (3), the response Y follows a GP with mean zero and the covariance function ϕ specified by

$$\phi(Y(\mathbf{w}_1), Y(\mathbf{w}_2)) = \mathrm{Cov}(Y(\mathbf{w}_1), Y(\mathbf{w}_2)) = \sum_{j=1}^{q} \sigma_j^2 \tau_{z_{1j}z_{2j}}^{(j)} R\left(\mathbf{x}_1, \mathbf{x}_2 | \boldsymbol{\theta}^{(j)}\right), \tag{4}$$

where σ_j^2 and $\boldsymbol{\theta}^{(j)}$ ($j = 1, ..., q$) are process variance parameters and range parameters corresponding to $z^{(j)}$, respectively. Same as above, three different choices of $\tau_{z_{1j}z_{2j}}^{(j)}$, the exchangeable, multiplicative, and unrestricted correlation functions, are adopted in (4). Note that if any $\tau_{z_{1j}z_{2j}}^{(j)}$ has a zero (or near zero) value, the overall covariance in (2) will be zero (or near zero). While such problems are avoided in the additive model structure in (4). To elaborate the advantage of the additive covariance function compared with the multiplicative covariance structure, the following example will reinvestigate the real data analysis in Deng et al. (2017).

Example 1 A fully three-dimensional (3D) coupled finite element model (FEM) is used to capture the deformations and stresses of full-scale embankments involving unreinforced, piled, and two different reinforced and piled sections (Rowe & Liu, 2015). The computer experiments consider one quantitative factor and three qualitative factors on improving the performance of reinforced embankments with floating columns over soft clay. The quantitative factor x is the distance from the embankment centerline to the embankment shoulder, taking 29 values uniformly in [0, 14]. Three qualitative factors are embankment construction rate (z_1 = 1, 5, 10 m/month), Young's modulus of columns (z_2 = 50, 100, and 200 MPa), and reinforcement stiffness (z_3 = 1,578, 4,800, and 8,000 kN/m). The response variable is the final embankment crest settlement.

The settings of training data and testing data are the same as those in Section 5 of Deng et al. (2017). Specifically, the testing set contains 29 data points in which the values of quantitative factor are taken uniformly from [0, 14], and $(z_1, z_2, z_3) = (5, 100, 4{,}800)$. We compare four methods, denoted as AD, EC, MC, and UC, which are the additive GP model with covariance structure (4), and the GP models with multiplicative covariance function (2) under the exchangeable, multiplicative, and unrestrictive correlation functions for the qualitative factors, respectively. For methods in comparison, we make prediction at randomly chosen 20 input settings out of those 29 ones and compute the corresponding Nash-Sutcliffe efficiency (NSE). We repeat this process 100 times. Figure 3 displays the boxplots of the logarithm of NSEs for the four methods, showing that the AD method outperforms the EC, MC, and UC methods.

There is also some work of Bayesian approaches for analyzing computer experiments with QQ factors. Han, Santner, Notz, and Bartel (2009) proposed a Bayesian hierarchical quantitative–qualitative variable model to fit QQ inputs. But they assumed a strong assumption that the outputs corresponding to different levels of qualitative factors are draws from Gaussian stochastic processes with similar correlation structures and magnitude of variation. They described the "similarities" in the model parameters by a prior distribution. A few other directions of analyzing computer experiments for QQ factors are also developed recently. Roustant et al. (2018) proposed "group kernels" for the GP model with QQ factors to accommodate a potentially large number of combination levels of qualitative factors. Zhang, Tao, Chen, and Apley (2018) introduced an approach that maps each qualitative factor to an underlying quantitative latent

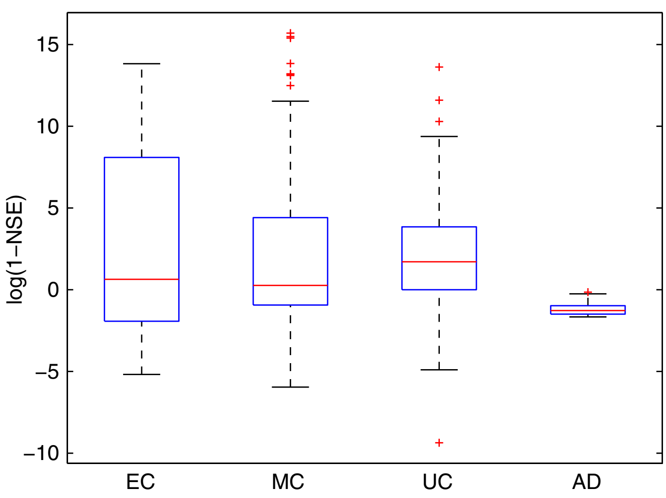

FIGURE 3 Boxplots of the log(1 − NSE) associated with "EC," "MC," "UC," and "AD" for the computer model in the real application over 100 randomly chosen prediction sets of 20 input settings. Abbreviation: EC, exchangeable correlation; MC, multiplicative correlation; NSE, Nash-Sutcliffe efficiency; UC, unrestrictive correlation

variable, such that the GP model with QQ factors can be transformed to a standard GP model with only quantitative factors.

4 | CONNECTION TO DATA MINING

Machine learning and data mining are methods of data analysis that automate model prediction. They have attracted great attention resulting with wide applications in various industrial sectors. The GP model is one of the popular tools in machine learning (Williams & Rasmussen, 2006), which gives easy interpretation of model prediction and provides a well-founded framework for supervised learning and model selection. Moreover, the data with QQ inputs often occur in the machine learning and data mining. In this regard, the methodologies reviewed in this article are potentially useful for machine learning to handle the QQ inputs. The methods and techniques reviewed in this work can be much more flexible than the one-hot encoding technique used in the machine learning to deal with qualitative variables. In addition, the design strategies reviewed in this article can be used to facilitate the adaptive selection of design points in the context of active learning (Deng, Joseph, Sudjianto, & Wu, 2009) and Bayesian optimization (Conti & O'Hagan, 2010; Han, Santner, Notz, & Bartel, 2009) when both QQ inputs are presented in the data. For example, in the Bayesian optimization with QQ factors, the marginal coupled design can serve as a good initial design to make the optimization procedure efficiently search for the optimal setting.

5 | CONCLUSION

In this work, we review a set of design and analysis methods for computer experiments, especially with both QQ factors. There are several topics deserving further investigation. Space-filling designs are popular in practice, but better designs may well exist. Sequential designs appear to be particularly appropriate for expensive computer experiments. On the other hand, the existing techniques in modeling computer experiments are computationally intensive when the number of observations is large. To overcome this challenge, development of efficient modeling techniques is called for. Some ideas of local Gaussian process (Bilionis & Zabaras, 2012; Fang et al., 2019; Nguyen-Tuong, Peters, & Seeger, 2008; Yan, Li, Bai, Deng, & Foley, 2017) can be used to facilitate the computation. Finally, the model calibration for computer experiments is important (Chandra & Lin, 2012; Chang, Kerns, Lee, & Stanek, 2009; Han, Santner, & Rawlinson, 2009). However, calibrating a computer model with QQ factors would call for careful investigation.

ACKNOWLEDGMENT

The authors thank the Associate Editor and referees for their helpful comments to improve the quality of the paper. This research is supported by the National Natural Science Foundation of China (71903024, 71531004).

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

AUTHOR CONTRIBUTIONS

Xiaoning Kang: Investigation; methodology; writing-original draft, review, and editing. **Xinwei Deng**: Investigation; methodology; supervision; visualization; writing-review and editing.

ORCID

Xinwei Deng https://orcid.org/0000-0002-1560-2405

RELATED WIRES ARTICLE

Experimental design

REFERENCES

Ba, S., Myers, W. R., & Brenneman, W. A. (2015). Optimal sliced Latin hypercube designs. *Technometrics*, *57*, 479–487.

Bilionis, I., & Zabaras, N. (2012). Multi-output local Gaussian process regression: Applications to uncertainty quantification. *Journal of Computational Physics*, *231*, 5718–5746.

Chandra, W., & Lin, D. K. J. (2012). Parameter calibration and inadequacy modeling in a computer experiment. *Quality and Reliability Engineering International*, *28*, 35–46.

Chang, G. A., Kerns, G. J., Lee, D., & Stanek, G. L. (2009). Calibration experiments for a computer vision oyster volume estimation system. *Journal of Statistics Education*, *17*. https://doi.org/10.1080/10691898.2009.11889528

Conti, S., & O'Hagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *Journal of Statistical Planning and Inference*, *140*, 640–651.

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

Deng, X., Hung, Y., & Lin, C. D. (2015). Design for computer experiments with qualitative and quantitative factors. *Statistica Sinica*, *25*, 1567–1581.

Deng, X., Joseph, V. R., Sudjianto, A., & Wu, C. J. (2009). Active learning through sequential design, with applications to detection of money laundering. *Journal of the American Statistical Association*, *104*, 969–981.

Deng, X., Lin, C. D., Liu, K.-W., & Rowe, R. (2017). Additive Gaussian process for computer models with qualitative and quantitative factors. *Technometrics*, *59*, 283–292.

Fang, G., Wang, X., Wang, K., Lee, K. H., Ho, J. D. L., Fu, H. C., … Kwok, K. W. (2019). Vision-based online learning kinematic control for soft robots using local Gaussian process regression. *IEEE Robotics and Automation Letters*, *4*, 1194–1201.

Fang, K.-T., Li, R., & Sudjianto, A. (2005). *Design and modeling for computer experiments*. London: Chapman and Hall/CRC.

Gramacy, R. B. (2012). Cases for the nugget in modeling computer experiments. *Statistics and Computing*, *22*, 713–722.

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

Han, G., Santner, T. J., & Rawlinson, J. J. (2009). Simultaneous determination of tuning and calibration parameters for computer experiments. *Technometrics*, *51*, 464–474.

He, Y., Lin, C. D., & Sun, F. (2017). On construction of marginally coupled designs. *Statistica Sinica*, *27*, 665–683.

He, Y., Lin, C. D., Sun, F., & Lv, B. (2017). Marginally coupled designs for two-level qualitative factors. *Journal of Statistical Planning and Inference*, *187*, 103–108.

Hedayat, A. S., Sloane, N. J. A., & Stufken, J. (1999). *Orthogonal arrays*. New York, NY: Springer.

Higdon, D., Gattiker, J., Williams, B., & Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, *103*, 570–583.

Hobert, J. P., Jones, G. L., Presnell, B., & Rosenthal, J. S. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, *89*, 769–784.

Huang, H., Lin, D. K., Liu, M.-Q., & Yang, J.-F. (2016). Computer experiments with both qualitative and quantitative variables. *Technometrics*, *58*, 495–507.

Hung, Y. (2011a). Adaptive probability-based Latin hypercube designs. *Journal of the American Statistical Association*, *106*, 213–219.

Hung, Y. (2011b). Penalized blind kriging in computer experiments. *Statistica Sinica*, *21*, 1171–1190.

Jiang, H., Deng, X., López, V., & Hamann, H. F. (2016). Online updating of computer model output using real-time sensor data. *Technometrics*, *58*, 472–482.

Joseph, V. R. (2016). Space-filling designs for computer experiments: A review. *Quality Engineering*, *28*, 28–35.

Joseph, V. R., & Delaney, J. D. (2007). Functionally induced priors for the analysis of experiments. *Technometrics*, *49*, 1–11.

Joseph, V. R., Gul, E., & Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, *102*, 371–380.

Joseph, V. R., Gul, E., & Ba, S. (2019). Designing computer experiments with multiple types of factors: The MaxPro approach. *Journal of Quality Technology*, in press. https://doi.org/10.1080/00224065.2019.1611351

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

Kleijnen, J. P. (2009). Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, *192*, 707–716.

Kong, X., Ai, M., & Tsui, K. L. (2018). Flexible sliced designs for computer experiments. *Annals of the Institute of Statistical Mathematics*, *70*, 1–16.

Li, R., & Sudjianto, A. (2005). Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. *Technometrics*, *47*, 111–120.

Linkletter, C., Bingham, D., Hengartner, N., Higdon, D., & Ye, K. Q. (2006). Variable selection for Gaussian process models in computer experiments. *Technometrics*, *48*, 478–490.

Loh, W. L. (1996). On Latin hypercube sampling. *Annals of Statistics*, *24*, 2058–2080.

Martin, J. D., & Simpson, T. W. (2004). Use of Kriging models to approximate deterministic computer models. *AIAA Journal*, *43*, 853–863.

Mckay, M. D., Beckman, R. J., & 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.

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

Nguyen-Tuong, D., Peters, J., & Seeger, M. (2008). Local Gaussian process regression for real time online model learning and control. In *International conference on neural information processing systems*.

Owen, A. B. (1992). A central limit theorem for Latin hypercube sampling. *Journal of the Royal Statistical Society*, *54*, 541–551.

Plumlee, M., & Apley, D. W. (2017). Lifted Brownian Kriging models. *Technometrics*, *59*, 165–177.

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

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

Rasmussen, C., & Williams, C. (2006). *Gaussian processes for machine learning*. Cambridge, MA: The MIT Press.

Roustant, O., Padonou, E., Deville, Y., Clément, A., & Wynn, H. (2018). Group kernels for Gaussian process metamodels with categorical inputs. *arXiv.org/abs/1802.02368*.

Rowe, R. K., & Liu, K.-W. (2015). Three-dimensional finite element modelling of a full-scale geosynthetic-reinforced, pile-supported embankment. *Canadian Geotechnical Journal*, *52*, 2041–2054.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, *4*, 409–423.

Santner, T. J., Williams, B. J., & Notz, W. I. (2018). *The design and analysis of computer experiments*. New York, NY: Springer.

Stein, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, *29*, 143–151.

Suen, C. Y. (1989). Some resolvable orthogonal arrays with two symbols. *Communications in Statistics*, *18*, 3875–3881.

Wang, L., Xiao, Q., & Xu, H. (2018). Optimal maximin L_1-distance Latin hypercube designs based on good lattice point designs. *The Annals of Statistics*, *46*, 3741–3766.

Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning*. Cambridge, MA: MIT Press.

Yan, J., Li, K., Bai, E. W., Deng, J., & Foley, A. M. (2017). Hybrid probabilistic wind power forecasting using temporally local Gaussian process. *IEEE Transactions on Sustainable Energy*, *7*, 87–95.

Zhang, Y., & Notz, W. I. (2015). Computer experiments with qualitative and quantitative variables: A review and reexamination. *Quality Engineering*, *27*, 2–13.

Zhang, Y., Tao, S., Chen, W., & Apley, D. W. (2018). A latent variable approach to Gaussian process modeling with qualitative and quantitative factors. Technometrics. https://doi.org/10.1080/00401706.2019.1638834

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

How to cite this article: Kang X, Deng X. Design and analysis of computer experiments with quantitative and qualitative inputs: A selective review. *WIREs Data Mining Knowl Discov*. 2020;10:e1358. https://doi.org/10.1002/widm.1358