Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors

Zhiguang Qian[†], Huaiqing Wu[‡], C. F. Jeff Wu[†]

[†]*School of Industrial and Systems Engineering,*
Georgia Institute of Technology, Atlanta, GA

[‡]*Department of Statistics*,
Iowa State University

July 3, 2006

Abstract

Modeling experiments with qualitative and quantitative factors is an important but unresolved issue in computer modeling. Some Gaussian process models that incorporate both qualitative and quantitative factors are proposed. The key to the development of these new models is an approach for constructing correlation functions with qualitative and quantitative factors. An iterative estimation procedure is developed for the proposed models. Modern optimization techniques are used in the estimation to ensure the validity of the constructed correlation functions. The proposed method is illustrated with an example involving a known function and a real example for modeling the thermal distribution of a data center.

KEY WORDS: Cokriging; Design of experiments; Kriging; Multivariate Gaussian processes; Semi-definite programming.

1 INTRODUCTION

In recent years, there has been a growing interest in the use of computer models in sciences, engineering, and business. The corresponding physical experimentation might

otherwise be time-consuming, costly, or even impossible to conduct. Because of their many attractive features, Gaussian process (GP) models have been established as a core tool for modeling computer experiments. (For detailed discussions of such models, see Santner, Williams, and Notz 2003; Fang, Li, and Sudjianto 2005.) An important but unresolved issue is how to model computer experiments with qualitative and quantitative factors. Standard methods assume that all the factors involved in a computer experiment are quantitative. However, in many situations, some factors are qualitative by nature. Consider, for instance, the data-center computer experiment to be discussed in Section 6. The configuration variables that determine the thermal properties of a data center can be either quantitative or qualitative. Examples of quantitative variables are rack temperature rise, rack power, and diffuser flow rate. Examples of qualitative variables are diffuser height (with levels "mid height of the room" and "ceiling height"), mixed power (with levels "uniform," "alt-zero," and "alt-half"), and hot-air return-vent location (with levels "perpendicular-bottom," "perpendicular-top," "parallel-bottom," and "parallel-top") (Schmidt, Cruz, and Iyengar 2005). Computer models with qualitative and quantitative factors occur frequently in business operations applications, where some social economical status of the customers, such as gender and commuting method, is inherently qualitative.

The purpose of this article is to propose new GP models to address this issue. Note that the corresponding problem for physical experimentation is much easier because no GP model is involved (Wu and Ding 1998; Wu an Hamada 2000). Quadratic models have long been used for modeling physical experiments involving quantitative and qualitative factors. While such polynomial models can provide reasonable approximations to physical phenomena, they are inapplicable to computer modeling. Unlike physical experiments, computer experiments are known to have highly non-linear input-output relationships, include many factors, and are deterministic. It is essential to develop data-driven models for computer experiments with qualitative and quantitative factors. Inspired by the success of GP models with quantitative factors, we extend them to accommodate both qualitative and quantitative factors. As a key to the development of the new GP models, a general approach for constructing correlation functions with qualitative and quantitative factors is proposed. An iterative estimation procedure is developed for the proposed model, making use of some modern optimization techniques to ensure the validity of the constructed correlation functions.

The remainder of this article is organized as follows. Section 2 presents the models used throughout the article and the motivation for this study. Section 3 gives a general approach for constructing correlation functions for GP models with qualitative and quantitative factors. Section 4 presents estimation and prediction procedures. Sections 5 and 6 illustrate the proposed method with an example involving a known function, and

with a real example for modeling temperature in a data center. Section 7 provides some discussions and concluding remarks. Proofs and computational details are deferred to the Appendix.

2 MODELS AND MOTIVATION

2.1 Gaussian Process Models With Quantitative Factors

For later development, we first briefly review GP models with quantitative factors. Suppose that an experiment involves I factors (input variables) $\mathbf{x} = (x_1, \ldots, x_I)^t$; the data consist of an $n \times I$ matrix of input variable values, denoted by $\mathbf{X} = (\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0)^t$, and the corresponding response values $\mathbf{y} = (y_1, \ldots, y_n)^t$. The GP model assumes the following structure:

$$y(\mathbf{x}) = \boldsymbol{\beta}^t \mathbf{f}(\mathbf{x}) + \epsilon(\mathbf{x}), \tag{2.1}$$

where $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_m(\mathbf{x}))^t$ is a set of m pre-specified functions, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^t$ is a vector of unknown coefficients. The residual $\epsilon(\mathbf{x})$ is assumed to be a stationary GP with mean zero and covariance

$$\mathrm{cov}(\epsilon(\mathbf{x}_1), \epsilon(\mathbf{x}_2)) = \sigma^2 K_{\boldsymbol{\phi}}(\mathbf{x}_1, \mathbf{x}_2),$$

where σ^2 is the variance, and $K_{\boldsymbol{\phi}}(\cdot, \cdot)$ is the correlation function, which depends on the unknown *correlation parameters* $\boldsymbol{\phi}$.

It is well known that the product of one-dimensional correlation functions is a valid correlation function. The use of the product correlation function allows each factor to have its own correlation parameters, which can shed light on how response values are correlated among different factors. One popular choice is the *product Gaussian correlation function* (Santner, Williams, and Notz 2003):

$$K_{\boldsymbol{\phi}}(\mathbf{x}_1, \mathbf{x}_2) = \prod_{i=1}^{I} \exp\{-\phi_{i1}(x_{i1} - x_{i2})^2\}, \tag{2.2}$$

where $\phi_{i1} \geq 0$ for $i = 1, \ldots, I$. Here, $\exp\{-\phi_{i1}(x_{i1} - x_{i2})^2\}$ is a valid correlation function for the variable x_i (Abrahamsen 1997). Note that the power correlation parameters are fixed at 2. This reduces the complication of estimating the correlation parameters. In addition, the sample path of the GP is infinitely differentiable, which is a reasonable assumption for many applications. The scale correlation parameters $\phi_{11}, \ldots, \phi_{I1}$ measure the ruggedness of the response surface (sample path) of the GP. Larger values of ϕ_{i1}'s

imply a more rugged response surface.

2.2 Gaussian Process Models With Qualitative and Quantitative Factors

To develop GP models with both qualitative and quantitative factors, we first note that a computer experiment tends to involve more quantitative factors than qualitative factors. This is because quite often more quantitative factors are needed to specify the underlying physics or mathematics of the experiment, and they are also more informative than qualitative factors. Although the number of qualitative factors is usually not large, they can determine some important properties of the experiment. For example, in the data-center experiment to be discussed in Section 6, factors that are associated with the choice of cooling materials, the method of heat transformation, and the orientation of diffusers are qualitative. To take into account the distinct natures and roles of quantitative and qualitative factors in computer modeling, three possible analysis approaches can be employed, which are described below.

The first is the *independent analysis* in which distinct Gaussian processes are used for modeling the data collected at different level combinations of the qualitative factors. This method may be relevant in some extreme situations, where different qualitative factor settings result in totally different physics of the experiments. It assumes no similarity among the experimental results obtained at the different level combinations of the qualitative factors, thus rendering the independent analysis the sole sensible choice. However, such scenarios rarely occur in practice, because a typical computer experiment contains many quantitative factors, which make the responses at different qualitative factor levels correlated to each other. Thus, the independent analysis has very limited applicability. In addition, its implementation requires fitting many GP models, and the number of unknown parameters in these models can be large, even when the number of qualitative factors is small. Consider, for example, an experiment with seven quantitative factors and three 4-level qualitative factors. The independent analysis would require fitting 64 ($= 4^3$) models, which involve 64 mean parameters (with constants used as the means of the processes), 64 variances, and 448 ($= 64 \times 7$) correlation parameters. To accurately estimate these 576 parameters would require a large number of observations, which generally cannot be afforded.

The second approach is the *collapsed analysis*, which completely ignores qualitative factors in an experiment by simply fitting a *single* GP model based on the quantitative factors. At first glance, this approach does not seem to pose much challenge in the model fitting because only one GP with the quantitative factors is involved. However, this

oversimplification of the problem ignores the contribution of qualitative factors, which can play important roles in determining the system performance.

In view of the shortcomings of the above two approaches, we introduce an *integrated analysis* in this article. It assumes a *single* GP model across different values of qualitative and quantitative factors as to borrow strengths from all the observations. Suppose that a computer experiment involves factors $\mathbf{w} = (\mathbf{x}^t, \mathbf{z}^t)^t$, where $\mathbf{x} = (x_1, \ldots, x_I)^t$ are quantitative factors, and $\mathbf{z} = (z_1, \ldots, z_J)^t$ are qualitative factors. Similar to (2.1), the response $y(\mathbf{w})$ at the input value $\mathbf{w}$ is assumed to follow the model below:

$$y(\mathbf{w}) = \mu(\mathbf{w}) + \epsilon(\mathbf{w}). \tag{2.3}$$

Here, the mean $\mu(\mathbf{w})$ will be simplified to a constant μ, as often suggested in computer experiments (Welch, Buck, Sacks, Wynn, Mitchell, and Morris 1992), thus simplifying subsequent modeling and inference procedures. The residual $\epsilon(\mathbf{w})$ is assumed to be a zero-mean GP with variance σ^2 and some correlation function. Construction of a "valid" correlation function for $\epsilon(\mathbf{w})$ is not straightforward because such a function needs to be defined in the space involving both qualitative and quantitative factors. The Gaussian correlation function used in Section 2.1 or other distance-based correlation functions (Santner, Williams, and Notz 2003) are not applicable due to the absence of the notion of "distance" for qualitative factors. A general method for constructing valid correlation functions is developed in Section 3.

3 CONSTRUCTION OF CORRELATION FUNCTIONS FOR GAUSSIAN PROCESSES WITH QUALITATIVE AND QUANTITATIVE FACTORS

In this section, we propose a general method for constructing valid correlation functions for $\epsilon(\mathbf{w})$ in model (2.3). The method does not use the normality assumption of Gaussian processes and hence applies to general stochastic processes with qualitative and quantitative factors.

First, consider the simple case involving only one qualitative factor, z_1, with m_1 levels, denoted by $1, \ldots, m_1$. For simplicity, let $\epsilon_u(\mathbf{x}) = \epsilon((\mathbf{x}^t, u)^t)$, for $u = 1, \ldots, m_1$. To define the correlation function of the stochastic process $\epsilon(\mathbf{w})$, where $\mathbf{w} = (\mathbf{x}^t, z_1)^t$, we envision a

mean-zero m_1-variate process

$$\boldsymbol{\epsilon}^*(\mathbf{x}) = \begin{pmatrix} \epsilon_1(\mathbf{x}) \\ \vdots \\ \epsilon_{m_1}(\mathbf{x}) \end{pmatrix}.$$

Then we only need to define correlation and cross-correlation functions for $\boldsymbol{\epsilon}^*(\mathbf{x})$. A convenient approach is to assume that $\boldsymbol{\epsilon}^*(\mathbf{x}) = \mathbf{A}\boldsymbol{\eta}(\mathbf{x})$, where $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_{m_1})^t$ is an $m_1 \times m_1$ nonsingular matrix with unit row vectors (i.e., $\mathbf{a}_u^t \mathbf{a}_u = 1$ for $u = 1, \ldots, m_1$) and $\boldsymbol{\eta}(\mathbf{x}) = (\eta_1(\mathbf{x}), \ldots, \eta_{m_1}(\mathbf{x}))^t$ is an m_1-variate stochastic process. Here, $\eta_1(\mathbf{x}), \ldots, \eta_{m_1}(\mathbf{x})$ are m_1 independent stochastic processes with the same variance σ^2 and the correlation functions $K^0_{\boldsymbol{\phi}_1}, \ldots, K^0_{\boldsymbol{\phi}_{m_1}}$, respectively. Then the correlation function for $\epsilon(\mathbf{w})$ is

$$\text{cor}(\epsilon(\mathbf{w}_1), \epsilon(\mathbf{w}_2)) = \text{cor}(\epsilon_{z_{11}}(\mathbf{x}_1), \epsilon_{z_{12}}(\mathbf{x}_2)) = \text{cor}(\mathbf{a}_{z_{11}}^t \boldsymbol{\eta}(\mathbf{x}_1), \mathbf{a}_{z_{12}}^t \boldsymbol{\eta}(\mathbf{x}_2)),$$

where $\mathbf{w}_1 = (\mathbf{x}_1^t, z_{11})^t$ and $\mathbf{w}_2 = (\mathbf{x}_2^t, z_{12})^t$ are two input values. Since

$$\text{cor}(\boldsymbol{\eta}(\mathbf{x}_1), \boldsymbol{\eta}(\mathbf{x}_2)) = \left[\prod_{u=1}^{m_1} K^0_{\boldsymbol{\phi}_u}(\mathbf{x}_1, \mathbf{x}_2) \right] \mathbf{I}_{m_1}$$

with $\mathbf{I}_{m_1}$ being the $m_1 \times m_1$ identity matrix,

$$\text{cor}(\mathbf{a}_{z_{11}}^t \boldsymbol{\eta}(\mathbf{x}_1), \mathbf{a}_{z_{12}}^t \boldsymbol{\eta}(\mathbf{x}_2)) = \mathbf{a}_{z_{11}}^t \mathbf{a}_{z_{12}} K^0_{\boldsymbol{\phi}}(\mathbf{x}_1, \mathbf{x}_2), \tag{3.4}$$

where $K_{\boldsymbol{\phi}}(\mathbf{x}_1, \mathbf{x}_2) = \prod_{u=1}^{m_1} K^0_{\boldsymbol{\phi}_u}(\mathbf{x}_1, \mathbf{x}_2)$.

Let $\tau_{u_1,u_2} = \mathbf{a}_{u_1}^t \mathbf{a}_{u_2}$, where $1 \le u_1 \le m_1$ and $1 \le u_2 \le m_1$. Then $\mathbf{T}_1 = (\tau_{u_1,u_2})_{m_1 \times m_1} = \mathbf{A}\mathbf{A}^t$ is an $m_1 \times m_1$ positive definite matrix with unit diagonal elements. In fact, any positive definite matrix with unit diagonal elements can be written as $\mathbf{B}\mathbf{B}^t$, where $\mathbf{B}$ is a nonsingular matrix with unit row vectors. Thus, the above construction shows that, for any positive definite matrix $\mathbf{T}_1 = (\tau_{u_1,u_2})_{m_1 \times m_1}$ with unit diagonal elements and any correlation function $K_{\boldsymbol{\phi}}(\mathbf{x}_1, \mathbf{x}_2)$, $\text{cor}(\epsilon(\mathbf{w}_1), \epsilon(\mathbf{w}_2)) = \tau_{z_{11},z_{12}} K_{\boldsymbol{\phi}}(\mathbf{x}_1, \mathbf{x}_2)$ is a valid correlation function. Similar correlation functions are used in Mardia and Goodall (1993) for kriging, in Brown, Le, and Zidek (1994) for assigning a prior to a covariance matrix, and in Banerjee and Gelfand (2002) for modeling a cross-covariance matrix.

If z_1 has two levels, 1 and 2, and $\tau_{12} > 0$, then $\tau_{z_{11},z_{12}}$ can be represented as $\exp\{-\theta I[z_{11} \neq z_{12}]\}$, where $\theta = \ln(1/\tau_{12}) > 0$ and $I[z_{11} \neq z_{12}]$ is the indicator function that takes 1 if $z_{11} \neq z_{12}$ and 0 otherwise. If z_1 has $m_1 > 2$ levels and $\tau_{z_{11},z_{12}}$ is positive and the same for any $z_{11} \neq z_{12}$, then $\tau_{z_{11},z_{12}} = \exp\{-\theta I[z_{11} \neq z_{12}]\}$. This would be the case if the m_1 levels of z_1 are of isotropic nature; that is, the cross-correlation between $\epsilon((\mathbf{x}^t, z_{11})^t)$ and

$\epsilon((\mathbf{x}^t, z_{12})^t)$ is the same for all $z_{11} \neq z_{12}$. As will be shown at the end of this section, $\mathbf{T}_1 = (\tau_{u_1,u_2})_{m_1 \times m_1}$ so defined is a positive definite matrix with unit diagonal elements and is thus a valid correlation matrix.

Now consider the general case with J qualitative factors $\mathbf{z} = (z_1, \ldots, z_J)^t$, where z_j has m_j levels, denoted by $1, \ldots, m_j$, for $j = 1, \ldots, J$. As an extension to (3.4), a correlation function for $\epsilon(\mathbf{w})$ can be constructed as

$$\text{cor}(\epsilon(\mathbf{w}_1), \epsilon(\mathbf{w}_2)) = \prod_{j=1}^{J} \left[\tau_{j,z_{j1},z_{j2}} K_{\boldsymbol{\phi}_j}(\mathbf{x}_1, \mathbf{x}_2) \right], \tag{3.5}$$

where $\mathbf{T}_j = (\tau_{j,u_1,u_2})_{m_j \times m_j}$ is an $m_j \times m_j$ positive definite matrix with unit diagonal elements. This is a valid correlation function as it is the product of J valid correlation functions $\tau_{1,z_{11},z_{12}} K_{\boldsymbol{\phi}_1}(\mathbf{x}_1, \mathbf{x}_2), \ldots, \tau_{J,z_{J1},z_{J2}} K_{\boldsymbol{\phi}_J}(\mathbf{x}_1, \mathbf{x}_2)$ in (3.4) for the qualitative factors $z_1, \ldots, z_J$ (Santner, Williams, and Notz 2003).

In particular, if $K_{\boldsymbol{\phi}_j}(\mathbf{x}_1, \mathbf{x}_2)$ takes the form $\exp\{-\sum_{i=1}^{I} \phi_{ij}(x_{i1} - x_{i2})^2\}$ in (2.2), it is a valid correlation function as discussed in Section 2.1. Then, the correlation function (3.5) becomes

$$\text{cor}(\epsilon(\mathbf{w}_1), \epsilon(\mathbf{w}_2)) = \left[\prod_{j=1}^{J} \tau_{j,z_{j1},z_{j2}} \right] \exp\left\{ -\sum_{i=1}^{I} \phi_i (x_{i1} - x_{i2})^2 \right\}, \tag{3.6}$$

where $\phi_i = \sum_{j=1}^{J} \phi_{ij}$ for $i = 1, \ldots, I$. Note that the correlation function in (3.6) bears some resemblance with its counterpart in (2.2) for the GP model with quantitative factors. This similarity will, in part, motivate the estimation and inference procedures in Section 4.

If, in addition, all of the qualitative factors z_j's are isotropic (which is automatic if they all have two levels), and the correlation is positive, then, $\tau_{j,z_{j1},z_{j2}} = \exp\{-\theta_j I[z_{j1} \neq z_{j2}]\}$, where $\theta_j = \ln(1/\tau_{j,1,2}) > 0$, as explained above. In this case,

$$\mathbf{T}_j = (1 - c_j)\mathbf{I}_{m_j} + c_j \mathbf{1}\mathbf{1}^t,$$

where $c_j = \exp\{-\theta_j\}$ and $\mathbf{1} = (1, \ldots, 1)^t$ is the $m_j \times 1$ vector of 1's. Then, for any non-zero $m_j \times 1$ vector $\mathbf{a}$,

$$\mathbf{a}^t \mathbf{T}_j \mathbf{a} = (1 - c_j)\mathbf{a}^t\mathbf{a} + c_j(\mathbf{a}^t\mathbf{1})^2 > 0,$$

because $0 < c_j < 1$. Thus, $\mathbf{T}_j$ is a positive definite matrix with unit diagonal elements, and is indeed a legitimate choice in (3.6). Then, the correlation function (3.6) simplifies

to

$$\mathrm{cor}(\epsilon(\mathbf{w}_1), \epsilon(\mathbf{w}_2)) = \exp\left\{ -\sum_{i=1}^{I} \phi_i (x_{i1} - x_{i2})^2 - \sum_{j=1}^{J} \theta_j I[z_{j1} \neq z_{j2}] \right\}. \tag{3.7}$$

This special correlation function leads to much simplified estimation and inference procedures, as will be seen in Section 4.

4 ESTIMATION AND PREDICTION

Suppose the data consist of n different input values, denoted by $D_w = \{\mathbf{w}_1^0, \ldots, \mathbf{w}_n^0\}$, and the corresponding responses, denoted by $\mathbf{y} = (y_1, \ldots, y_n)^t$. Consider model (2.3) with the correlation function in (3.6). The unknown parameters that need to be estimated are μ, σ^2, $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_I)^t$, and $\mathbf{T} = \{\mathbf{T}_1, \ldots, \mathbf{T}_J\}$. The maximum likelihood method is adopted for the estimation. Denote by $\widehat{\mu}$, $\widehat{\sigma}^2$, $\widehat{\boldsymbol{\phi}}$, and $\widehat{\mathbf{T}}$ the resulting estimators.

The Gaussian process assumption implies that the log-likelihood of $\mathbf{y}$, up to a constant, is

$$-\frac{1}{2}\left[n \ln \sigma^2 + \ln |\mathbf{R}| + \frac{(\mathbf{y} - \mathbf{F}\mu)^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{F}\mu)}{2\sigma^2} \right], \tag{4.8}$$

where $\mathbf{F} = (1, \ldots, 1)^t$ is an $n \times 1$ vector; $\mathbf{R}$ is the correlation matrix, which depends on the correlation parameters $\boldsymbol{\phi}$ and $\mathbf{T}$, and its (i,j)th entry is $\mathrm{cor}(\epsilon(\mathbf{w}_i^0), \epsilon(\mathbf{w}_j^0))$ defined in (3.6).

First, given $\boldsymbol{\phi}$ and $\mathbf{T}$, $\widehat{\mu}$ and $\widehat{\sigma}^2$ can be obtained as follows:

$$\begin{aligned} \widehat{\mu} &= (\mathbf{F}^t \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathbf{R}^{-1} \mathbf{y}, \\ \widehat{\sigma}^2 &= \frac{1}{n} (\mathbf{y} - \mathbf{F}\widehat{\mu})^t \mathbf{R}^{-1} (\mathbf{y} - \mathbf{F}\widehat{\mu}). \end{aligned} \tag{4.9}$$

Substituting $\widehat{\sigma}^2$ into (4.8), we obtain a simplified form (up to a negative constant)

$$n \ln(\widehat{\sigma}^2) + \ln |\mathbf{R}|, \tag{4.10}$$

where $\widehat{\sigma}^2$ and $\mathbf{R}$ depend on the correlation parameters $\boldsymbol{\phi}$ and $\mathbf{T}$.

If all the qualitative factors are isotropic and the correlations are positive, then (3.6) becomes (3.7) and the estimation of $\boldsymbol{\phi}$ and $\mathbf{T}$ becomes the estimation of $\boldsymbol{\phi}$ and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)^t$. As discussed in Section 3, the correlation function in (3.7) is valid if the ϕ_i's

and θ_j's are positive. Then, $\widehat{\boldsymbol{\phi}}$ and $\widehat{\boldsymbol{\theta}}$ can be obtained as follows:

$$\begin{aligned}(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}) = \text{argmin}_{(\boldsymbol{\phi},\boldsymbol{\theta})} \quad & [n \ln(\widehat{\sigma}^2) + \ln |\mathbf{R}|] \\ \text{subject to} \quad & \phi_i \geq 0, \quad i = 1, \ldots, I, \\ & \theta_j \geq 0, \quad j = 1, \ldots, J. \end{aligned} \tag{4.11}$$

Next we discuss the estimation for the general case without assuming that $z_1, \ldots, z_J$ are isotropic. The absence of the isotropic assumption poses great challenges in the estimation, because the validity of (3.6) as a correlation function requires that all the $\mathbf{T}_j$'s be valid correlation matrices, that is, positive definite (or positive semi-definite) matrices with unit diagonal elements. The problem of estimating a positive definite or positive semi-definite matrix occurs in many applications in statistics, including factor analysis (Bartholomew and Knott 1999) and Gaussian graphical models (Lauritzen 1996; Edwards 2000). Two features of the present application set it apart from others. First, our problem is especially challenging because it entails estimating *multiple* correlation matrices whereas in other applications, such as a Gaussian graphical model, usually a *single* correlation matrix is involved. Second, as a major advantage in computer experiments, one can somehow "freely" choose input factor values. This flexibility is not shared by the observational studies to which factor analysis and Gaussian graphical models are usually applied. As will be discussed in Section 4.2, the use of "appropriate" experimental designs for input factors can significantly simplify the estimation procedure.

Standard methods used in statistics for maximizing a likelihood function involving a positive definite matrix work in the following manner. First, note that a matrix is positive definite if and only if all its leading principle minors are positive. These constraints then transfer to a series of nonlinear inequalities involving the elements of the matrix. Finally, an optimization problem is solved with the resulting nonlinear inequalities as the constraints and the elements of the matrix as the optimization variables. This "element-orientated" approach would involve many complicated nonlinear inequalities and a huge number of optimization variables even when the dimension of the matrix is not very large, making it computationally infeasible. To better address the optimization problem (4.10) with positive-definiteness constraints on the $\mathbf{T}_j$'s, we make use of some recently developed optimization techniques in semi-definite programming. A brief introduction of semi-definite programming is given in Section 4.1. The estimation procedure is developed in Section 4.2, and the prediction procedure is provided in Section 4.3.

4.1 Semi-definite Programming

Consider the optimization problem

$$\begin{aligned} \min_{\mathbf{X}} \quad & \mathbf{C} \bullet \mathbf{X} \\ \text{subject to} \quad & \mathbf{A}_i \bullet \mathbf{X} = \mathbf{b}_i, \quad i = 1, \ldots, m, \\ & \mathbf{X} \succ 0 \ (\succeq 0), \end{aligned} \tag{4.12}$$

where the optimization variable is $\mathbf{X}$ in the space of $n \times n$ real symmetric matrices. The inequalities $\mathbf{X} \succ 0$ and $\mathbf{X} \succeq 0$ mean that $\mathbf{X}$ is positive definite and positive semi-definite, respectively. The problem (4.12) is referred to as a *semi-definite programming* (SP) in optimization (Vandenberghe and Boyd 1996; Wolkowicz, Saigal, and Vandenberghe 2000). The notation $\mathbf{C} \bullet \mathbf{X}$ represents the inner product of the matrices $\mathbf{C}$ and $\mathbf{X}$:

$$\mathbf{C} \bullet \mathbf{X} = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}, \tag{4.13}$$

where x_{ij} and x_{ij} are the (i, j)th entries of $\mathbf{C}$ and $\mathbf{X}$, respectively. Equivalently, $\mathbf{C} \bullet \mathbf{X}$ can also be written as $\text{tr}(\mathbf{CX})$. Throughout the article, "tr" stands for the trace of a square matrix. This type of optimization problem arises in many fields, including statistics, communication theory, and machine learning. The SP problem is a convex problem, which can be solved efficiently by *interior point algorithms* (Wolkowicz, Saigal, and Vandenberghe 2000). The SP method takes a holistic view of positive semi-definite matrices and computes the solution to (4.12) within a *cone* formed by positive definite matrices (Vandenberghe and Boyd 1996). It can lead to significant computational savings, especially for large scale problems.

4.2 Estimation Procedure

The general case under consideration involves J qualitative factors $z_1, \ldots, z_J$ and I quantitative factors $x_1, \ldots, x_I$. No isotropic conditions are imposed on the qualitative factors. Without loss of generality, the number of levels of z_j, denoted by m_j, is assumed to be three or higher. If a qualitative factor has two levels, it can be grouped with the quantitative factors in the estimation because there is no need to impose positive-definiteness

conditions on it. If $\widehat{\mu}$ is given, $\widehat{\boldsymbol{\phi}}$ and $\widehat{\mathbf{T}}$ can be obtained as follows:

$$\begin{aligned}(\widehat{\boldsymbol{\phi}}, \widehat{\mathbf{T}}) = \text{argmin}_{(\boldsymbol{\phi},\mathbf{T})} \quad & [n \ln(\widehat{\sigma}^2) + \ln|\mathbf{R}|] \\ \text{subject to} \quad & \phi_i \geq 0, \;\; i = 1, \ldots, I, \\ & \mathbf{T}_j \succ 0, \;\; j = 1, \ldots, J, \\ & \text{diag}(\mathbf{T}_j) = \mathbf{1}, \;\; j = 1, \ldots, J,\end{aligned}$$

where the optimization variables are $\boldsymbol{\phi}$ and $\mathbf{T}$. Throughout this article, "diag" stands for the diagonal elements of a matrix and $\mathbf{1}$ stands for a vector of 1's. Estimating $\boldsymbol{\phi}$ and $\mathbf{T}$ can be carried out by iterating between the following ϕ-step and T-step.

ϕ-step: Given $\widehat{\mathbf{T}}$, $\widehat{\boldsymbol{\phi}}$ is obtained as follows:

$$\begin{aligned}\widehat{\boldsymbol{\phi}} = \text{argmin}_{\boldsymbol{\phi}} \quad & [n \ln(\widehat{\sigma}^2) + \ln|\mathbf{R}|] \\ \text{subject to} \quad & \phi_i \geq 0, \;\; i = 1, \ldots, I,\end{aligned} \tag{4.14}$$

where $\widehat{\mu} = (\mathbf{F}^t\mathbf{R}^{-1}\mathbf{F})^{-1}\mathbf{F}^t\mathbf{R}^{-1}\mathbf{y}$ and $\widehat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{F}\widehat{\mu})^t\mathbf{R}^{-1}(\mathbf{y} - \mathbf{F}\widehat{\mu})$.

T-step: Given $\widehat{\boldsymbol{\phi}}$, $\widehat{\mathbf{T}}$ is obtained as follows:

$$\begin{aligned}\widehat{\mathbf{T}} = \text{argmin}_{\mathbf{T}} \quad & [n \ln(\widehat{\sigma}^2) + \ln|\mathbf{R}|] \\ \text{subject to} \quad & \mathbf{T}_j \succ 0, \;\; j = 1, \ldots, J, \\ & \text{diag}(\mathbf{T}_j) = \mathbf{1}, \;\; j = 1, \ldots, J,\end{aligned} \tag{4.15}$$

In optimization, such an iterative algorithm is called *block coordinate descent* or *nonlinear Gaussian-Seidel* method (Bertsekas 1999). It is well known that this type of algorithm will converge under mild conditions (e.g., the objective function is differentiable). The optimization problem in (4.14) is a standard nonlinear program, which can be solved by quasi-Newton algorithms. The major difficulty in the implementation of the procedure lies in the T-step because of the complex objective function and constraints involved. The details for implementing the T-step are given below. For easier presentation, let $\mathbf{e} = \mathbf{y} - \mathbf{F}\widehat{\mu}$ and $\mathbf{E} = \mathbf{e}\mathbf{e}^t/n$ throughout this article. Then,

$$\widehat{\sigma}^2 = \text{tr}(\mathbf{e}^t\mathbf{R}^{-1}(\mathbf{e}/n)) = \text{tr}((\mathbf{e}/n)\mathbf{e}^t\mathbf{R}^{-1}) = \text{tr}(\mathbf{E}\mathbf{R}^{-1}).$$

Thus, the objective function in (4.15) can be written as follows:

$$f(\mathbf{T}) = n \ln[\text{tr}(\mathbf{E}\mathbf{R}^{-1})] + \ln|\mathbf{R}|. \tag{4.16}$$

For computational convenience, the optimization problem in (4.15) can be approximated by the following linear problem:

$$\begin{aligned} \widehat{\mathbf{T}} = \text{argmin}_{\mathbf{T}} \quad & \left[f(\mathbf{T}_0) + \textstyle\sum_{j=1}^{J} (\frac{\partial f(\mathbf{T}_0)}{\partial \mathbf{T}_j} \bullet \mathbf{T}_j) \right] \\ \text{subject to} \quad & \mathbf{T}_j \succ 0, \;\; j = 1, \ldots, J, \\ & \text{diag}(\mathbf{T}_j) = \mathbf{1}, \;\; j = 1, \ldots, J, \end{aligned} \tag{4.17}$$

where $\frac{\partial f(\mathbf{T}_0)}{\partial \mathbf{T}_j}$ is the partial derivative of $f(\mathbf{T})$ with respect to $\mathbf{T}_j$, evaluated at some given value of $\mathbf{T}$, $\mathbf{T}_0 = \{\mathbf{T}_{0,1}, \ldots, \mathbf{T}_{0,J}\}$. That is,

$$\frac{\partial f(\mathbf{T}_0)}{\partial \mathbf{T}_j} = \left[\frac{n}{\text{tr}(\mathbf{E}\mathbf{R}^{-1})} \frac{\partial \text{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \mathbf{T}_j} + \frac{1}{|\mathbf{R}|} \frac{\partial |\mathbf{R}|}{\partial \mathbf{T}_j} \right] |_{\mathbf{T}=\mathbf{T}_0}. \tag{4.18}$$

The formulas for $\frac{\partial \text{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \mathbf{T}_j}$ and $\frac{\partial |\mathbf{R}|}{\partial \mathbf{T}_j}$ are given in the Appendix. Such a linear approximation has been shown to be reasonable and is widely used to approximate SP problems with nonlinear objective function constraints (Wolkowicz, Saigal, and Vandenberghe 2000). If necessary, this kind of linear approximation can be repeated a few times. Now define the following block diagonal matrices

$$\mathbf{W} = \text{bkdiag}(\mathbf{T}_1, \ldots, \mathbf{T}_J), \;\; \text{and} \;\; \mathbf{C} = \text{bkdiag}\left(\frac{\partial f(\mathbf{T}_0)}{\partial \mathbf{T}_1}, \ldots, \frac{\partial f(\mathbf{T}_0)}{\partial \mathbf{T}_J} \right). \tag{4.19}$$

Note that $\mathbf{W} \succ 0$ if and only if $\mathbf{T}_1 \succ 0, \ldots, \mathbf{T}_J \succ 0$. Then, the optimization problem in (4.17) can be recast as the following SP problem:

$$\begin{aligned} \widehat{\mathbf{W}} = \text{argmin}_{\mathbf{W}} \quad & \mathbf{C} \bullet \mathbf{W} \\ \text{subject to} \quad & \mathbf{W} \succ 0, \\ & \text{diag}(\mathbf{W}) = \mathbf{1}. \end{aligned} \tag{4.20}$$

Some structures of D_w can significantly simplify the computation in the T-step. This convenience is only possible for the present experimental situation but not for an observational study. Consider first the simple case involving *one* qualitative factor z_1 with more than two levels, denoted by $1, \ldots, m_1$. Assume D_w is a *cross array* (Wu and Hamada 2000) of D_x and D_z, where D_x is a $p \times I$ design matrix for the quantitative factors $\mathbf{x}$, and $D_z = (1, \ldots, m_1)^t$ is an $m_1 \times 1$ design matrix for the qualitative factor z_1. Consequently, D_w consists of all level combinations between those in D_x and those in D_z. Hence D_w has $n = pm_1$ rows (runs). As shown in the following proposition, this cross-array structure of D_w simplifies the optimization problem in (4.15) and also makes it free of $\widehat{\boldsymbol{\phi}}$. Consequently, estimating $\boldsymbol{\phi}$ and $\mathbf{T}_1$ can be done *separately* by carrying out a simplified T-step

and then the ϕ-step. This is much simpler than the general estimation procedure, which iterates between the ϕ-step and the T-step.

Let $\mathbf{H} = (h_{j_1 j_2})$ denote a $p \times p$ matrix with its (j_1, j_2)th entry given as

$$h_{j_1 j_2} = \exp\left\{ -\sum_{i=1}^{I} \phi_i (x_{ij_1} - x_{ij_2})^2 \right\}. \tag{4.21}$$

With the above assumption on experimental design, the optimization problem in (4.15) can be simplified as follows:

Proposition 1. *Suppose that D_w is a cross array of D_x and D_z. Then, the problem in (4.15) is equivalent to*

$$\begin{aligned} \widehat{\mathbf{T}}_1 = \mathit{argmin}_{\mathbf{T}_1} \quad & \left(m_1 \ln[\mathit{tr}(\mathbf{T}_1^{-1})] + \ln|\mathbf{T}_1|\right) \\ \mathit{subject\ to} \quad & \mathbf{T}_1 \succ 0, \\ & \mathit{diag}(\mathbf{T}_1) = \mathbf{1}. \end{aligned} \tag{4.22}$$

The proof of this proposition is given in the Appendix. With the proposition, the linear approximation (4.17) becomes

$$\begin{aligned} \widehat{\mathbf{T}}_1 = \mathrm{argmin}_{\mathbf{T}_1} \quad & [f(\mathbf{T}_{1,0}) + \nabla_{\mathbf{T}_1} f(\mathbf{T}_{1,0}) \bullet \mathbf{T}_1] \\ \text{subject to} \quad & \mathbf{T}_1 \succ 0, \\ & \mathrm{diag}(\mathbf{T}_1) = \mathbf{1}, \end{aligned} \tag{4.23}$$

where

$$f(\mathbf{T}_{1,0}) \;=\; m_1 \ln[\mathrm{tr}(\mathbf{T}_{1,0}^{-1})] + \ln|\mathbf{T}_{1,0}|,$$

and (from Dattorro 2005, app. D.2.3 and D.2.4)

$$\nabla_{\mathbf{T}_1} f(\mathbf{T}_{1,0}) \;=\; -\frac{m_1}{\mathrm{tr}(\mathbf{T}_{1,0}^{-1})} \mathbf{T}_{1,0}^{-2} + \mathbf{T}_{1,0}^{-1}.$$

These expressions significantly simplify the optimization problem in (4.23). This is a SP described in Section 4.1 and can be solved by efficient interior point algorithms.

We now consider the general case with J qualitative factors $z_1, \ldots, z_J$. Assume D_w is a cross array of D_x, the $p \times I$ design matrix for $\mathbf{x}$, and D_z, the $q \times J$ design matrix for $\mathbf{z}$. Hence D_w has $n = pq$ rows (runs). Let $\mathbf{z}_1^0, \ldots, \mathbf{z}_q^0$ denote the q input values for $\mathbf{z}$, and $\mathbf{T}^*$

be a $q \times q$ matrix with its (r,s)th entry given as

$$t_{r,s} = \prod_{j=1}^{J} \tau_{j,z_{jr}^0,z_{js}^0}. \tag{4.24}$$

Using the argument for establishing Proposition 1, we have the following result:

Proposition 2. *Suppose that D_w is a cross array of D_x and D_z, where D_x is a $p \times I$ design matrix for $\mathbf{x}$ and D_z is a $q \times J$ design matrix for $\mathbf{z}$. Then, the problem in (4.15) is equivalent to*

$$\begin{aligned} \widehat{\mathbf{T}} = \mathit{argmin}_{\mathbf{T}} \quad & (q \ln[\mathit{tr}((\mathbf{T}^*)^{-1})] + \ln|\mathbf{T}^*|) \\ \mathit{subject\ to} \quad & \mathbf{T}_j \succ 0, \ \ j = 1, \ldots, J, \\ & \mathit{diag}(\mathbf{T}_j) = \mathbf{1}, \ \ j = 1, \ldots, J. \end{aligned} \tag{4.25}$$

Again, the cross-array structure of D_w reduces the estimation of $\boldsymbol{\phi}$ and $\mathbf{T}$ to separately carrying out the simplified T-step and then the ϕ-step. This is much simpler than the general estimation procedure that iterates between the ϕ-step and the T-step.

If D_z also has some cross-array structure, the optimization problem in (4.25) can be further simplified. Suppose the qualitative factors $\mathbf{z}$ are grouped into $d \geq 2$ disjoint sets, $\{z_j : j \in A_k\}$, for $k = 1, \ldots, d$, where $\bigcup_{j=1}^{d} A_j = \{1, \ldots, J\}$ and the size of A_k is $J_k \geq 1$. Suppose further that D_z is a cross array of $D_1, \ldots, D_d$, where D_k is a $q_k \times J_k$ design matrix for the factors in $\{z_j : j \in A_k\}$. (However, D_k is not required to be a cross array among its constituent factors $z_j, j \in A_k$.) Thus, $\prod_{k=1}^{d} q_k = q$ and $\sum_{k=1}^{d} J_k = J$. Let $\mathbf{T}_{A_k} = \{\mathbf{T}_j : j \in A_k\}$ and $\mathbf{T}_k^*$ be a $q_k \times q_k$ matrix with its (r,s)th entry given as

$$t_{r,s}^{(k)} = \prod_{j \in A_k} \tau_{j,z_{jr}^0,z_{js}^0}. \tag{4.26}$$

Again, using the argument for establishing Proposition 1, we have the following result:

Proposition 3. *Suppose that D_z in Proposition 2 is a cross array of $D_1, \ldots, D_d$, where D_k is a $q_k \times J_k$ design matrix for the factors in $\{z_j : j \in A_k\}$. Then, solving the problem (4.25) is equivalent to solving the following d simpler problems separately:*

$$\begin{aligned} (P_{A_k}): \quad \widehat{\mathbf{T}}_{A_k} = \mathit{argmin}_{\mathbf{T}_{A_k}} \quad & (q_k \ln[\mathit{tr}((\mathbf{T}_k^*)^{-1})] + \ln|\mathbf{T}_k^*|) \\ \mathit{subject\ to} \quad & \mathbf{T}_j \succ 0, \ \ j \in A_k, \\ & \mathit{diag}(\mathbf{T}_j) = \mathbf{1}, \ \ j \in A_k, \end{aligned} \tag{4.27}$$

for $k = 1, \ldots, d$. In particular, if $d = J$, $A_k = \{k\}$, and $q_k = m_k$, then $\mathbf{T}_{A_k} = \mathbf{T}_k$ and

$\mathbf{T}_k^* = \mathbf{T}_k$. *Then (4.27) simplifies to*

$$(P_k): \quad \widehat{\mathbf{T}}_k = \textit{argmin}_{\mathbf{T}_k} \quad (m_k \ln[\textit{tr}((\mathbf{T}_k)^{-1})] + \ln|\mathbf{T}_k|)$$
$$\textit{subject to} \qquad \mathbf{T}_k \succ 0,$$
$$\textit{diag}(\mathbf{T}_k) = \mathbf{1}. \tag{4.28}$$

The method proposed to tackle the problem in (4.15) can be used to solve the problems in the above two propositions.

4.3 Prediction

The fitted GP model can be used in predicting the response value y at any untried point in the design space. The empirical best linear unbiased predictor (BLUP) of y at the point $\mathbf{w}_0$ is

$$\widehat{y}(\mathbf{w}_0) = \widehat{\mu} + \widehat{\mathbf{r}}_0^t \widehat{\mathbf{R}}^{-1}(\mathbf{y} - \widehat{\mu}), \tag{4.29}$$

where $\widehat{\mathbf{R}}$ is the estimated correlation matrix of $\mathbf{y}$ and

$$\widehat{\mathbf{r}}_0 = (\widehat{\text{cor}}(y(\mathbf{w}_0), y(\mathbf{w}_1^0)), \ldots, \widehat{\text{cor}}(y(\mathbf{w}_0), y(\mathbf{w}_n^0)))^t.$$

The empirical BLUP in (4.29) enjoys such nice properties as its counterpart for the Gaussian process model with quantitative factors in Section 2.1, and smoothly interpolates all the observed data points.

5 AN EXAMPLE INVOLVING A KNOWN FUNCTION

In this section, we consider an experiment involving one qualitative factor, z_1, and one quantitative factor, x_1. The output of this experiment is assumed to come from the following function:

$$y = \begin{cases} \exp(1.4x_1)\cos(7\pi x_1/2) & \text{if } z_1 = 1, \\ \exp(3x_1)\cos(7\pi x_1/2) & \text{if } z_1 = 2. \end{cases}$$

Figure 1 depicts the two curves of the function values with $z_1 = 1$ and $z_1 = 2$, respectively. The overall similarity of the curves makes the independent analysis inappropriate for this example, whereas their notable differences in some regions also make the collapsed analysis inappropriate. The integrated analysis seems to be most appropriate among the

three, and is expected to outperform the other two.

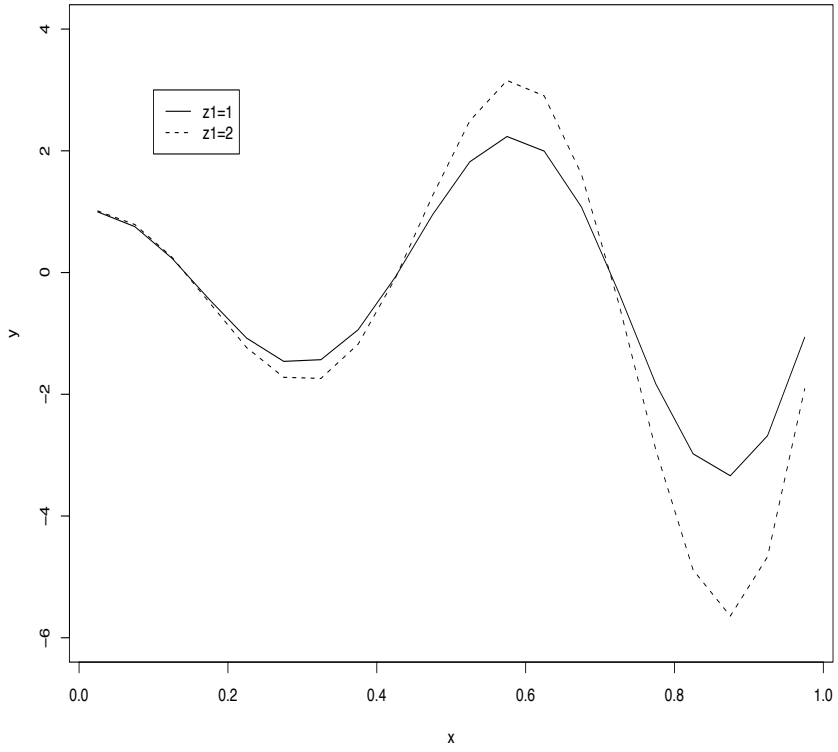

Figure 1: Two curves of the function values with $z_1 = 1$ and $z_1 = 2$.

Table 1 lists the training data used for model building, including four runs with $z_1 = 1$ and another four runs with $z_1 = 2$. For comparison, the data is analyzed using the three approaches discussed in Section 2.2.

x_1	z_1	y
0.0625	1	0.84
0.1875	1	-0.61
0.3125	1	-1.48
0.4375	1	0.18
0.5625	2	5.38
0.6875	2	2.28
0.8125	2	-10.09
0.9375	2	-10.56

Table 1: The training data for the example involving a known function

The independent analysis fits two separate GP models, one for the four runs with

$z_1 = 1$ and one for the four runs with $z_1 = 2$. Table 2 lists the estimated parameters in the fitted GP models. Note that $\widehat{\phi} = 1225.495$ for the GP model with $z_1 = 1$ is so large that the prediction based on the fitted models will be rugged, potentially producing inaccurate results. The collapsed analysis fits a GP model to the eight runs using the values of x_1 only, ignoring the difference between $z_1 = 1$ and $z_1 = 2$. Table 3 reports the estimated parameters in the resulting model. The integrated analysis fits a GP model that incorporates both x_1 and z_1 to the eight runs, and the estimated parameters are given in Table 4.

	$\widehat{\phi}$	$\widehat{\sigma}^2$	$\widehat{\mu}$
$z_1 = 1$	1225.495	0.76	-0.27
$z_1 = 2$	31.44	50.35	-2.90

Table 2: The estimated parameters of the fitted GP models with $z_1 = 1$ and $z_1 = 2$ in the independent analysis

$\widehat{\phi}$	$\widehat{\sigma}^2$	$\widehat{\mu}$
24.15	33.50	-1.60

Table 3: The estimated parameters of the fitted GP model in the collapsed analysis

$\widehat{\phi}$	$\widehat{\theta}_{12}$	$\widehat{\sigma}^2$	$\widehat{\mu}$
30.69	20.00	27.50	-1.36

Table 4: The estimated parameters of the fitted GP model in the integrated analysis

Next we assess the prediction accuracy of the three methods. The testing data consists of 40 data points. For $z_1 = 1$ and $z_1 = 2$, x_1 takes 20 equally-spaced values $0.025, 0.075, \ldots, 0.975$. The root mean squared errors (RMSEs) for these three prediction methods are calculated. The RMSE for the integrated analysis is 7.71, which is respectively 45% and 41% of the RMSEs (17.30 and 18.67) for the independent and collapsed analyses. Clearly, the prediction accuracy of the integrated analysis is much better than that of the other two analyses.

6 A DATA-CENTER COMPUTER EXPERIMENT

In this section, the proposed method is illustrated using a data-center computer experiment. With the increasing need for storing, manipulating, and managing data sets, data centers are widely used to provide application services or management for various data processing, such as web hosting internet, intranet, telecommunication, and information technology. Figure 2 shows a schematic layout of an internet data center using

Sun Microsystems (Lawrence Berkeley National Laboratory 2002). Driven by advances in hardware and data-storage techniques, data centers now can be very large, sprawling over thousands of square feet.

In designing and running a reliable data center, it is essential to maintain the system operating environment at a temperature within a functional range. Data-center facilities are extremely energy intensive, with many computer equipments constantly generating heat. Monitoring and studying the temperature of a data center is a difficult task, because it is largely unknown how different configurations affect the thermal distribution of the data center. The physical thermal process is complex, depending on many factors, and detailed temperatures at different locations cannot be actually measured. A computer experiment, built on computational fluid-dynamics models, implemented in professional software like Flotherm (Flometrics 2005) and FLUENT (Fluent 1998), is often used as a proxy to study the air movement and thermal distribution of a data center.

The experiment considered in this section models an air-cooled cabinet, implemented in Flotherm (Flometrics 2005), for predicting the airflow and heat transfer in the electronic equipments. Each run in this experiment takes several days to complete. We consider part of the data from Qian (2006). Table 5 lists eight configuration variables and their levels used in the experiment. In the table, x_1, x_2, x_3, x_4, and x_5 are quantitative factors; z_1, z_2, and z_3 are 2-, 4- and 3-level qualitative factors, respectively. The response of interest, denoted by y, is the temperature at one selected location of the system. More details for the engineering background of this data center can be found in Schmidt (2003), Schmidt, Cruz, and Iyengar (2005), and Qian (2006).

x_1: Rack temperature rise (C)	10	15	20		
x_2: Rack power (KW)	4	12	22	28	36
x_3: Diffuser angle	0	30			
x_4: Diffuser flow rate (%)	100	80	60		
x_5: Ceiling height (ft)	12	17	22		
z_1: Diffuser location/configuration.	Even	Odd			
z_2: Hot-air return-vent location	Bot-Per	Top-Per	Bot-Par	Top-Par	
z_3: Remove/mixed power	Uniform	Alt-Zero	Alt-Half		

Table 5: Configuration variables for the data center example

The five quantitative factors in Table 5 are of distinct scales, and their values are standardized first. The standardization of each variable is carried out by subtracting its lower design bound from its values, and then dividing the results by its design range. All results and plots given hereafter are associated with the standardized variables, which take values in $[0, 1]$. This experiment has 73 observations. There are 24 level combinations for the three qualitative factors. Hence, on average, each of these combinations has about

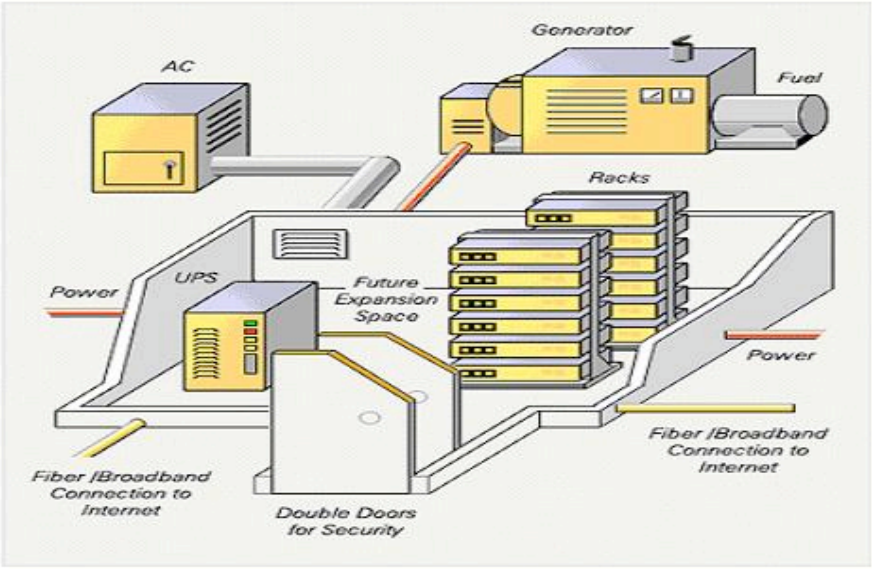

Figure 2: Schematic layout of an internet data center (Sun Microsystems) (Lawrence Berkeley National Laboratory 2002).

three observations, making the independent analysis infeasible. The data will be analyzed using the integrated analysis, and the results will be compared with those obtained from the collapsed analysis at the end of this section.

For the integrated analysis, the major difficulty with the model fitting is to estimate the correlation matrices for z_2 and z_3. The estimation is carried out using the two-step procedure in Section 4, implemented in *Matlab* (The MathWorks 2006) and making use of a semi-definite programming package *CVX* (Grant, Boyd, and Ye 2006). Note that this data set does not have a cross-array structure. Thus the general estimation procedure in Section 4, where the T-step involves the SP problem (4.20), was used. The procedure was found to converge after 100 iterations. Table 6 lists the estimated mean, variance, and correlation parameters for the quantitative factors x_1, x_2, x_3, x_4, and x_5, and the estimated correlation parameter for z_1. As shown in the table, the estimated correlation parameters vary significantly from one quantitative factor to another, and the values for x_3 and x_4 are much larger than the rest, indicating that the responses may be rugged in the dimensions of x_3 or x_4. The estimated correlation for z_1 between its two levels is small (0.00005), indicating that the responses at the two levels of z_1 are not significantly correlated. This is consistent with the known fact that different placements of diffusers (z_1) in a data center lead to distinct data-center thermal distributions (Schmidt 2003; Schmidt, Cruz, and Iyengar 2005).

$\widehat{\mu}$	$\widehat{\sigma}^2$	$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\phi}_3$	$\widehat{\phi}_4$	$\widehat{\phi}_5$	$\widehat{\tau}_1$
14.06	24.94	5.72	1.03	19.99	12.92	1.45	0.00005

Table 6: The estimated mean, variance, and correlation parameters for the quantitative factors x_1, x_2, x_3, x_4, and x_5, and the estimated correlation parameter for z_1

Tables 7 and 8 give the estimated correlation matrices for z_2 and z_3. Both matrices are symmetric with unit diagonal elements. Also, their eigenvalues are all positive [(0.154, 0.247, 0.311, 3.287) and (0.187, 0.454, 2.359), respectively]. Thus, the estimated correlation matrices are positive definite with unit diagonal elements, and are indeed valid correlation matrices.

	$i=1$	$i=2$	$i=3$	$i=4$
$j=1$	1.0000	0.7519	0.8019	0.6971
$j=2$	0.7519	1.0000	0.8187	0.7137
$j=3$	0.8019	0.8187	1.0000	0.7876
$j=4$	0.6971	0.7137	0.7876	1.0000

Table 7: The estimated correlation matrix for z_2

Following Welch, Buck, Sacks, Wynn, Mitchell, and Morris (1992), a functional ANOVA decomposition is conducted to study the factor effects in the fitted GP model. Since the

	$i=1$	$i=2$	$i=3$
$j=1$	1.0000	0.5960	0.8119
$j=2$	0.5960	1.0000	0.6230
$j=3$	0.8119	0.6230	1.0000

Table 8: The estimated correlation matrix for z_3

present example contains some qualitative factors, the procedure is modified accordingly. For example, in calculating the mean effect function of a qualitative factor, evaluation of the integral in the original ANOVA decomposition simplifies to averaging over some discrete values. First, we consider the first-order effects. Figure 3 depicts the mean-effect functions of the quantitative factors x_1, x_2, x_3, x_4, and x_5. Table 9 lists the main effects of the qualitative factors z_1, z_2, and z_3.

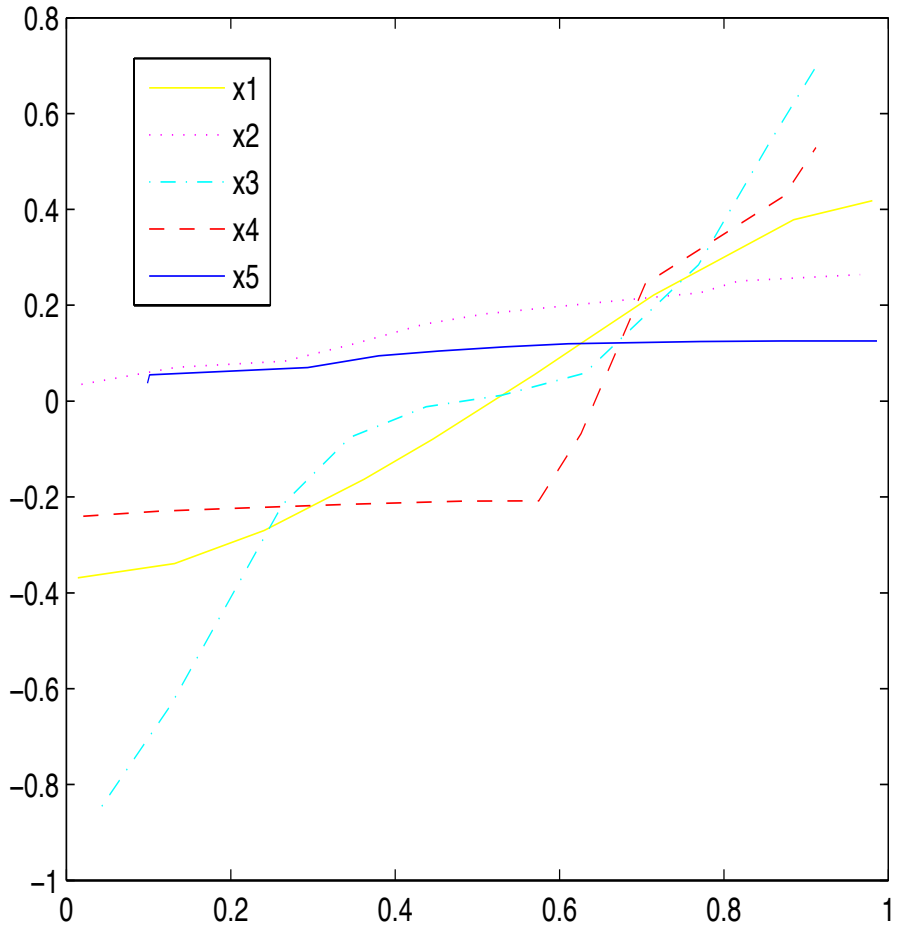

Figure 3: The main-effect functions of x_1, x_2, x_3, x_4, and x_5

.

z_1	Even: -0.01	Odd: 0.005		
z_2	Bot-Per: 0.08	Top-Per: 0.06	Bot-Par: 0.09	Top-Par: 0.03
z_3	Uniform: 0.05	Alt-Zero: 0.22	Alt-Half: 0.06	

Table 9: Estimated main effects of z_1, z_2, and z_3

Next, we investigate the second-order effects. Figure 4 displays the two-way interaction plots for some selected pairs of the quantitative factors. Note the large and complex interaction patterns of (x_2, x_3) and of (x_4, x_5). Figures 5-7 show some selected two-way interaction functions between the quantitative and qualitative factors. Note the different interaction patterns of (x_1, z_3) and (x_5, z_3). While the interactions of (x_1, z_3) are mainly due to z_3 ="alt-zero", the interactions of (x_5, z_3) are mainly due to z_3 ="alt-half." Figure 6 indicates that the interactions between x_4 and z_2 become larger as the values of x_4 move away from the middle. As illustrated by these figures, the interactions among the variables are rather intricate. Such nonlinear relationships cannot be captured by quadratic models.

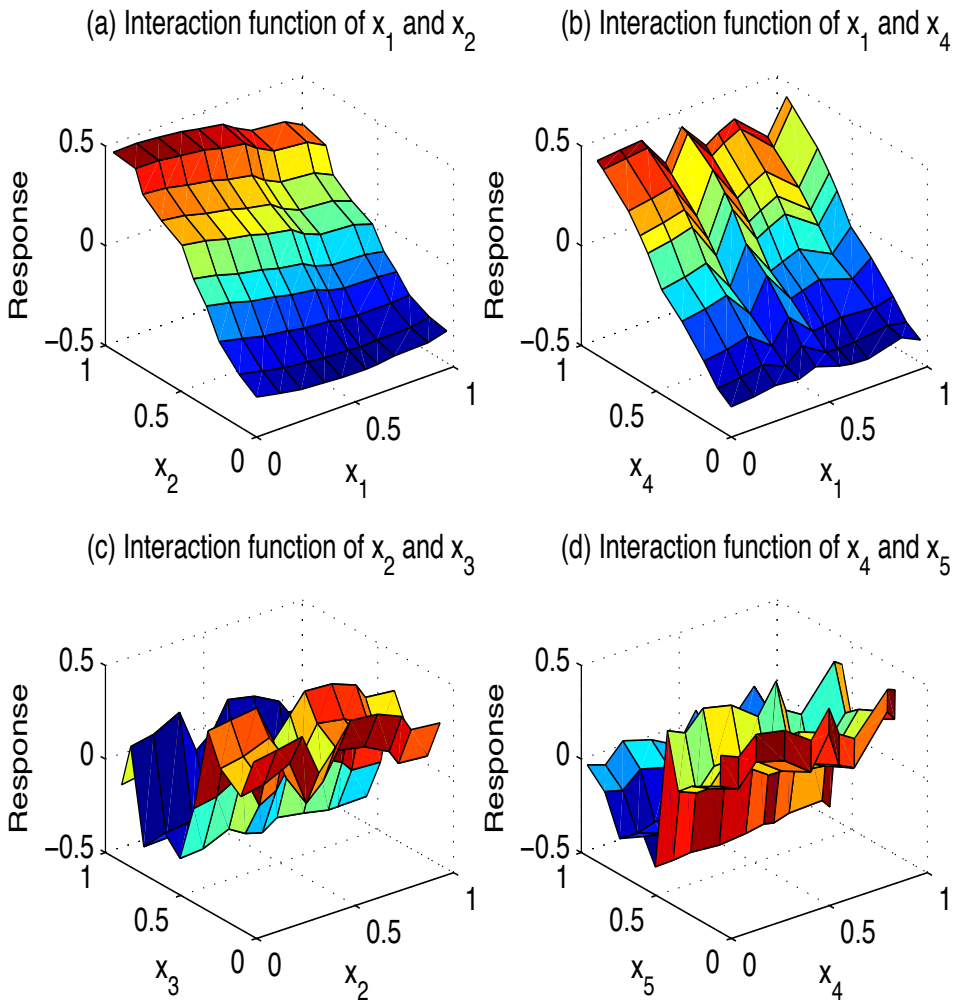

Figure 4: The two-way interaction functions for some selected pairs of x_1, x_2, x_3, x_4, and x_5.

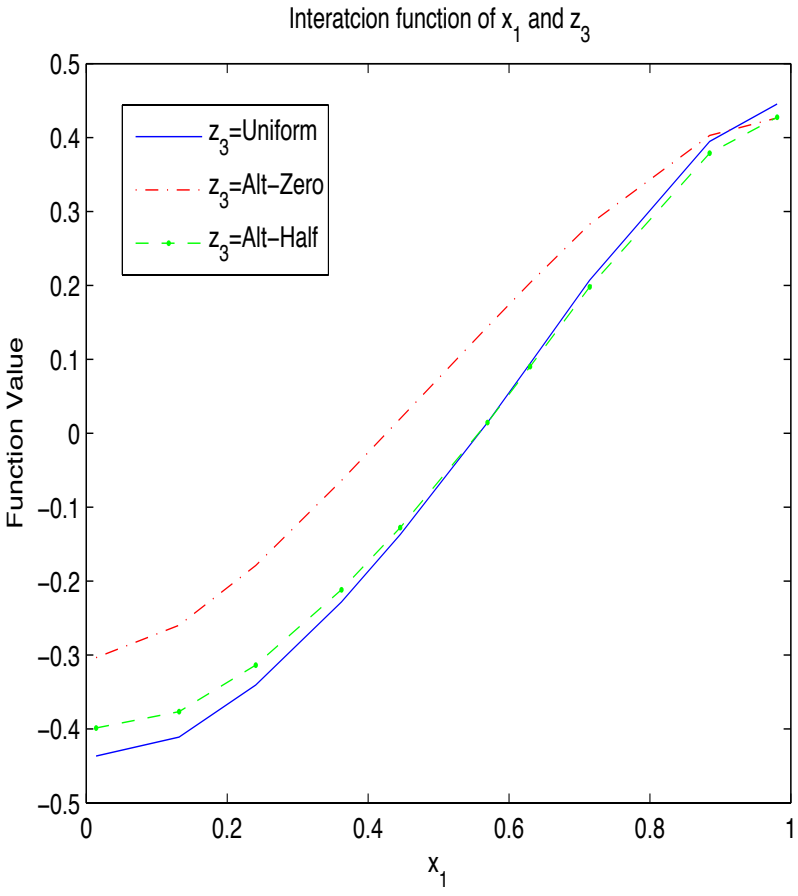

Figure 5: The two-way interaction function of x_1 and z_3.

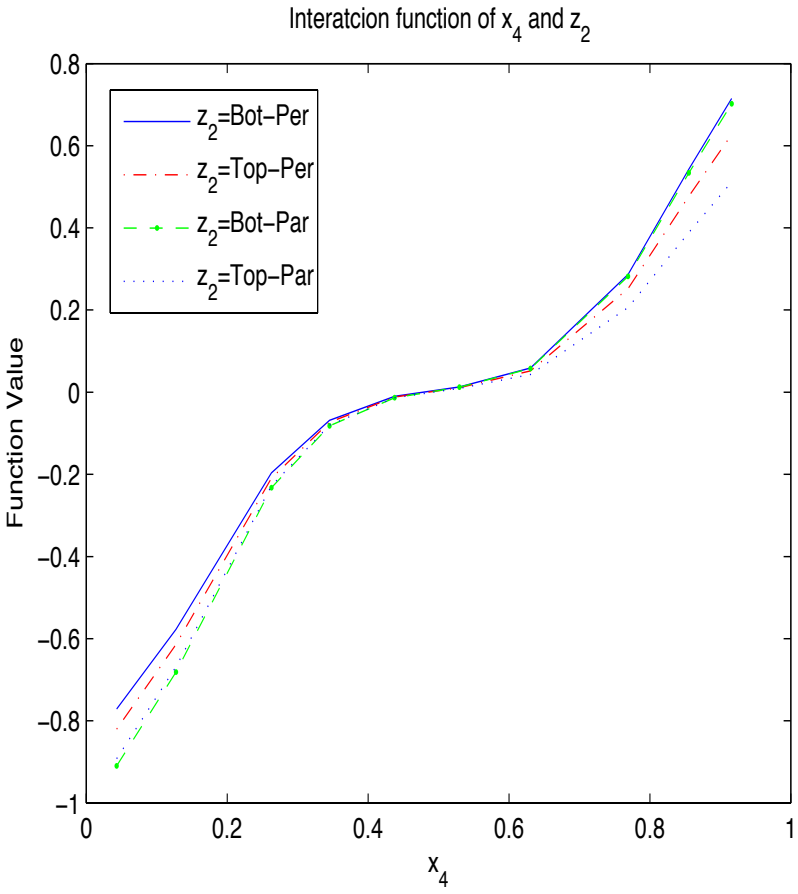

Figure 6: The two-way interaction function of x_4 and z_2.

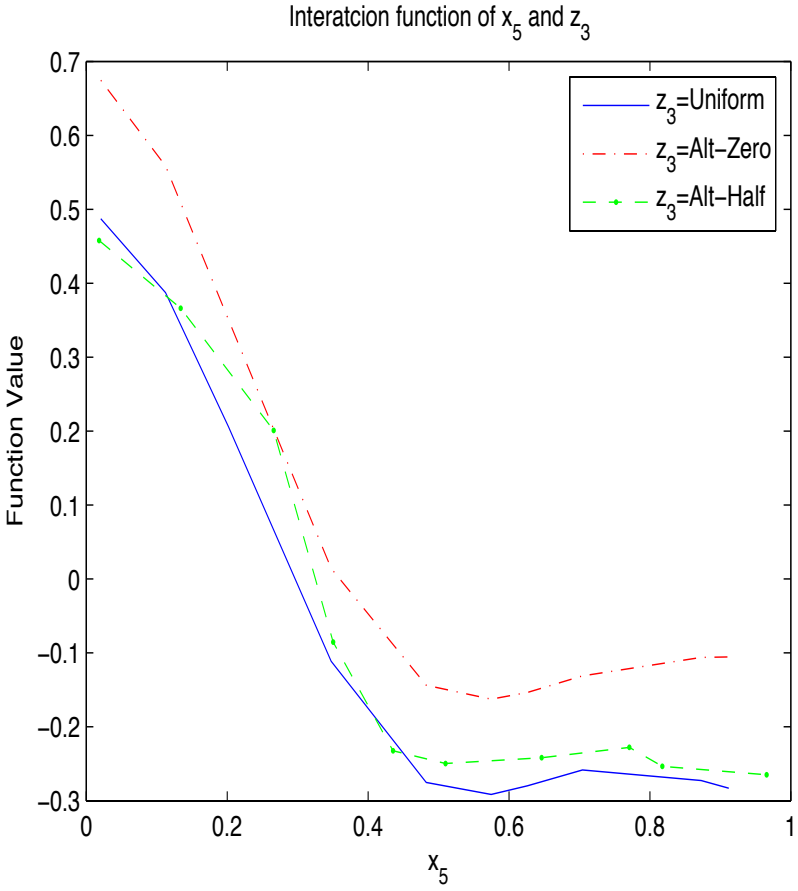

Figure 7: The two-way interaction function of x_5 and z_3

.

For comparison, the collapsed analysis of the 73 observations is also conducted. As shown in Table 10, the estimated correlation parameters for x_1, x_2, x_3, x_4, and x_5 are all very small. This seems to imply that all the quantitative factors have small significant effects on the thermal distribution of the data center. However, this is largely due to ignoring the contributions of the qualitative factors, which resulted in an ill-posed correlation matrix in the estimation procedure that led to very small estimated correlation parameters.

$\widehat{\phi}_1$	$\widehat{\phi}_2$	$\widehat{\phi}_3$	$\widehat{\phi}_4$	$\widehat{\phi}_5$
0.1540	0.0443	0.3683	0.1059	0.0498

Table 10: The estimated correlation parameters for x_1, x_2, x_3, x_4, and x_5 from the collapsed analysis

For this example the integrated analysis outperformed the collapsed analysis and successfully uncovered some intricate input-and-output relationships.

7 Discussions and Concluding Remarks

Ever since the publication of Sacks, Welch, Mitchell, and Wynn (1989), Gaussian process models have enjoyed great popularity in computer modeling. To date, one important but still unsettled problem is how to model computer experiments with qualitative and quantitative factors. Here, we give a systematic treatment of building Gaussian process models with qualitative and quantitative factors. The proposed methodology has two major contributions. First, it is a general method for constructing correlation functions with qualitative and quantitative factors. It makes use of some underlying multivariate Gaussian processes. The second is an iterative procedure used for estimation. The validity of the constructed correlation functions in the estimation is ensured using some recently developed optimization techniques.

The proposed method is successfully applied to an example involving a known function and a real example for designing data centers. Although the primary focus is on modeling and estimation, some suggestions for selecting designs for computer experiments with qualitative and quantitative factors are also given. Research on the design issue is currently ongoing and will be reported elsewhere.

Acknowledgment

Z. Qian is grateful to Dr. Zhaosong Lu and Dr. Renato D.C. Monteiro for stimulating discussions on some optimization issues.

Appendix: Proofs and Computational Details

Definitions and Formulas for $\frac{\partial \mathbf{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \mathbf{T}_j}$ and $\frac{\partial |\mathbf{R}|}{\partial \mathbf{T}_j}$

The definitions and results below follow from Graham (1981, chap. 4).

(1): Define $\frac{\partial \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \mathbf{T}_j}$ as

$$\frac{\partial \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \mathbf{T}_j} = \begin{pmatrix} \frac{\partial \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \tau_{j,1,1}} & \cdots & \frac{\partial \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \tau_{j,1,m_j}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \tau_{j,m_j,1}} & \cdots & \frac{\partial \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \tau_{j,m_j,m_j}} \end{pmatrix}.$$

For $1 \le r \le m_j$, $1 \le s \le m_j$, it is clear that $\frac{\partial \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1})}{\partial \tau_{j,r,s}} = \mathrm{tr}\left(\frac{\partial (\mathbf{E}\mathbf{R}^{-1})}{\partial \tau_{j,r,s}}\right)$. Furthermore, $\mathrm{tr}(\frac{\partial (\mathbf{E}\mathbf{R}^{-1})}{\partial \tau_{j,r,s}}) = \mathrm{tr}(\mathbf{E}\frac{\partial \mathbf{R}^{-1}}{\partial \tau_{j,r,s}}) = \mathrm{tr}(-\mathbf{E}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \tau_{j,r,s}}\mathbf{R}^{-1})$.

(2): Define $\frac{\partial |\mathbf{R}|}{\partial \mathbf{T}_j}$ as

$$\frac{\partial |\mathbf{R}|}{\partial \mathbf{T}_j} = \begin{pmatrix} \frac{\partial |\mathbf{R}|}{\partial \tau_{j,1,1}} & \cdots & \frac{\partial |\mathbf{R}|}{\partial \tau_{j,1,m_j}} \\ \vdots & \ddots & \vdots \\ \frac{\partial |\mathbf{R}|}{\partial \tau_{j,m_j,1}} & \cdots & \frac{\partial |\mathbf{R}|}{\partial \tau_{j,m_j,m_j}} \end{pmatrix}.$$

Let ρ_{uv} be the (u,v)th element of $\mathbf{R}$ and $\mathbf{R}_{uv}$ be the cofactor of element ρ_{uv} in $|\mathbf{R}|$. Then, for $1 \le r \le m_j$, $1 \le s \le m_j$,

$$\frac{\partial |\mathbf{R}|}{\partial \tau_{j,r,s}} = \mathrm{tr}(\mathbf{A}\mathbf{B}_{jrs}^t),$$

where $A = [\mathbf{R}_{uv}]$ and $\mathbf{B}_{jrs} = [b_{uv}^{(jrs)}]$ are $n \times n$ matrices. Here, $b_{uv}^{(jrs)} = \frac{\partial \rho_{uv}}{\partial \tau_{j,r,s}}$, for $1 \le u \le n$ and $1 \le v \le n$.

Proof of Proposition 1

Using the Kronecker product notation (Graham 1981), we have $\mathbf{R} = \mathbf{H} \otimes \mathbf{T}_1$. Basic facts on Kronecker product (Graham 1981, chap. 2) imply that

$$\begin{aligned} |\mathbf{H} \otimes \mathbf{T}_1| &= |\mathbf{H}|^{m_1} |\mathbf{T}_1|^p, \\ \mathbf{E}\mathbf{R}^{-1} = \mathbf{E}(\mathbf{H}^{-1} \otimes \mathbf{T}_1^{-1}) &= (\mathbf{E} \otimes 1)(\mathbf{H}^{-1} \otimes \mathbf{T}_1^{-1}) = (\mathbf{E}\mathbf{H}^{-1}) \otimes \mathbf{T}_1^{-1}, \\ \mathrm{tr}(\mathbf{E}\mathbf{R}^{-1}) = \mathrm{tr}(\mathbf{E}\mathbf{H}^{-1} \otimes \mathbf{T}_1^{-1}) &= \mathrm{tr}(\mathbf{E}\mathbf{H}^{-1})\mathrm{tr}(\mathbf{T}^{-1}). \end{aligned} \tag{7.30}$$

Since $\mathbf{E}$ and $\mathbf{H}$ are independent of $\mathbf{T}_1$, the problem (4.15) [using equation (4.16)] simplifies to (4.22).

References

[1] Abrahamsen, P. (1997), "A Review of Gaussian Random Fields and Correlation Functions," Report No. 917, http://publications.nr.no/917Rapport.pdf, Norwegian Computer Center, Oslo, Norway.

[2] Bartholomew, D. J., and Knott, M. (1999), *Latent Variable Models and Factor Analysis*, London: Arnold.

[3] Banerjee, S., and Gelfand, A. E. (2002), "Prediction, Interpolation and Regression for Spatially Misaligned Data," *Sankhya*, 64, 227-245.

[4] Bertsekas, D. P. (1999), *Nonlinear Programming*, Nashua, NH: Athena Scientific.

[5] Brown, P. J., Le, N. D., and Zidek, J. V. (1994), "Multivariate Spatial Interpolation and Exposure to Air Pollutants," *The Canadian Journal of Statistics*, 22, 489-509.

[6] Dattorro, J. (2005), *Convex Optimization and Euclidean Distance Geometry*, Palo Alto, CA: Meboo Publishing.

[7] Edwards, D. M. (2000), *Introduction to Graphical Modelling*, New York: Springer.

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

[9] Flometrics (2005),"Flotherm: Fluid Dynamics Based Thermal Analysis Software," http://www.flomerics.com/.

[10] Fluent, Inc. (1998), FLUENT, Release 5.5.14 (3d, segregated, laminar).

[11] Graham, A. (1981), *Kronecker Products and Matrix Calculus With Applications*, Chichester, England: Ellis Horwood Limited.

[12] Grant, M., Boyd, S., and Ye, Y. (2006), *CVX: Matlab Software for Disciplined Convex Programming*, Version 1.0 beta 3, http://www.stanford.edu/ boyd/cvx/.

[13] Lauritzen, S. L. (1996), *Graphical Models*, Oxford: Clarendon Press.

[14] Lawrence Berkeley National Laboratory (2002), "Data Center Energy Use: Truth versus Myth," http://www.lbl.gov/Science-Articles/Archive/data-center-energy-myth.html.

[15] Mardia, K. V., and Goodall, C. R. (1993), "Spatial-Temporal Analysis of Multivariate Environmental Monitoring Data," In: *Multivariate Environmental Statistics*, G. P. Patil and C. R. Rao, eds., 347-386, Amsterdam: Elsevier.

[16] Qian, Z. (2006), "Computer Experiments: Design, Modeling and Integration," Georgia Tech Ph.D. Thesis.

[17] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments," *Statistical Science*, 4, 409-435.

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

[19] Schmidt, R. (2003), "Hot Spots in Data Centers," Online Forum of Electrical Cooling, www.electronics-cooling.com.

[20] Schmidt, R. R., Cruz, E. E., and Iyengar, M. K. (2005), "Challenges of Data Center Thermal Management," *IBM Journal of Research and Development*, 49,709-723.

[21] The MathWorks, Inc. (2006), *Matlab: The Language of Technical Computing*, Version 6.5.1.

[22] Vandenberghe, L., and Boyd, S. (1996), "Semidefinite Programming," *SIAM Review*, 38, 49-95.

[23] Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992), "Screening, Predicting and Computer Experiments," *Technometrics*, 34, 15-25.

[24] Wolkowicz, H., Saigal, R., and Vandenberghe, L.(eds.) (2000), *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, Boston: Kluwer Academic.

[25] Wu, C. F. J., and Ding, Y. (1998), "Construction of Response Surface Designs for Qualitative and Quantitative Factors," *Journal of Statistical Planning and Inference*, 71, 331-348.

[26] Wu, C. F. J., and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.