

The Effects of Common Random Numbers on Stochastic Kriging Metamodels

XI CHEN, BRUCE E. ANKENMAN, and BARRY L. NELSON, Northwestern University

Ankenman et al. introduced stochastic kriging as a metamodeling tool for representing stochastic simulation response surfaces, and employed a very simple example to suggest that the use of Common Random Numbers (CRN) degrades the capability of stochastic kriging to predict the true response surface. In this article we undertake an in-depth analysis of the interaction between CRN and stochastic kriging by analyzing a richer collection of models; in particular, we consider stochastic kriging models with a linear trend term. We also perform an empirical study of the effect of CRN on stochastic kriging. We also consider the effect of CRN on metamodel parameter estimation and response-surface gradient estimation, as well as response-surface prediction. In brief, we confirm that CRN is detrimental to prediction, but show that it leads to better estimation of slope parameters and superior gradient estimation compared to independent simulation.

Categories and Subject Descriptors: I.6.6 [Computing Methodologies]: Simulation and Modeling

General Terms: Experimentation

Additional Key Words and Phrases: Simulation output analysis, simulation theory, common random numbers, metamodeling, variance reduction

ACM Reference Format:

Chen, X., Ankenman, B. E., and Nelson, B. L. 2012. The effects of common random numbers on stochastic kriging metamodels. *ACM Trans. Model. Comput. Simul.* 22, 2, Article 7 (March 2012), 20 pages. DOI = 10.1145/2133390.2133391 <http://doi.acm.org/10.1145/2133390.2133391>

1. INTRODUCTION

Beginning with the seminal papers of Kleijnen [1975] and Schruben and Margolin [1978], simulation researchers have been interested in the impact of incorporating Common Random Numbers (CRN) into experiment designs for fitting linear regression metamodels of the form

$$Y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + \varepsilon \quad (1)$$

to the output of stochastic simulation experiments. In Model (1), $Y(\mathbf{x})$ is the simulation output, $\mathbf{x} = (x_1, x_2, \dots, x_p)^\top$ is a vector of controllable design or decision variables, $\mathbf{f}(\mathbf{x})$ is a vector of known functions of \mathbf{x} (e.g., x_1, x_3^2, x_1x_7), $\boldsymbol{\beta}$ is a vector of unknown parameters of appropriate dimension, and ε represents the intrinsic variability in the simulation output assuming no bias in this metamodel.

CRN is a variance reduction technique that attempts to induce a positive correlation between the outputs of simulation experiments at distinct design points (settings of \mathbf{x} in the context of Model (1)) and thereby reduce the variance of the estimator of the

This work is supported by the National Science Foundation under grant No. CMMI-0900354.

Portions of this paper were previously published in Chen et al. [2010].

Authors' addresses: X. Chen, B. E. Ankenman, and B. L. Nelson (corresponding author), Department of Industrial Engineering and Management Sciences, Northwestern University; email: nelsonb@northwestern.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2012 ACM 1049-3301/2012/03-ART7 \$10.00

DOI 10.1145/2133390.2133391 <http://doi.acm.org/10.1145/2133390.2133391>

expected value of their difference. For $k \geq 2$ design points, a large literature has shown that, properly applied, CRN reduces the variance of “slope” parameters in (1), and therefore estimates of the response-surface gradient, while often inflating the variance of the intercept term. See, for instance, Donohue et al. [1992, 1995], Hussey et al. [1987a, 1987b], Kleijnen [1988, 1992], Nozari et al. [1987], and Tew and Wilson [1992, 1994].

It is fair to say that for Model (1) the role of CRN has been thoroughly examined. The purpose of this article is to undertake a similar analysis of the interaction of CRN and a new metamodeling technique called stochastic kriging Ankenman et al. [2008; 2010]. Stochastic kriging is an extension of kriging, which is typically applied to deterministic computer experiments (see, for instance, Santner et al. [2003]), to stochastic simulation. Kriging treats the unknown response surface as a realization of a Gaussian random field that exhibits spatial correlation, while stochastic kriging accounts for the additional uncertainty in stochastic simulation due to intrinsic sampling noise. Stochastic kriging is related to kriging with a “nugget effect” that treats the measurement errors as independent and identically distributed mean-zero random variables; stochastic kriging makes modeling additional properties of the random errors possible, namely unequal variances and correlation of the random errors across the design space. The focus of this article is the effects of introducing correlated random errors via CRN.

Ankenman et al. [2010] used a two-point problem with all parameters known and no trend model to show that CRN increases the Mean Squared Error (MSE) of the MSE-optimal predictor at a prediction point that has equal spatial correlation with the two design points. They speculated that CRN will not be helpful for prediction in general. In this article we generalize their two-point problem to allow unequal spatial correlations between the design points and the prediction point and inclusion of a linear trend model; further, we do not assume that the trend model parameters are known. Therefore we show that the detrimental effect of CRN was not an artifact of the assumptions of Ankenman et al. [2010]. We then extend the result given in Appendix EC.2 in Ankenman et al. [2010] for $k \geq 2$ spatially approximately uncorrelated design points and show that CRN inflates the MSE of prediction. In contrast to prediction, we show that CRN typically improves the estimation of trend model parameters (i.e., β) by reducing the variances of the slope parameters; CRN also improves gradient estimation in the sense that the gradient estimators from stochastic kriging are less affected by simulation noise when CRN is employed. A numerical study looks into the joint effect on prediction of using CRN and estimating the intrinsic variance; estimating the intrinsic variance is fundamental to stochastic kriging. All of these results are obtained under the assumption that the parameters of the spatial correlation model are known. Therefore, we close this article with two empirical studies in which this assumption is relaxed, and we evaluate the effects of CRN on parameter estimation, prediction, and gradient estimation in the context of estimating all the parameters of the stochastic kriging model.

2. STOCHASTIC KRIGING

In this section we briefly review stochastic kriging as developed in Ankenman et al. [2010] and the particular simplifications we exploit in this article.

In stochastic kriging we represent the simulation’s output on replication j at design point \mathbf{x} as

$$\mathcal{Y}_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \beta + M(\mathbf{x}) + \varepsilon_j(\mathbf{x}) = Y(\mathbf{x}) + \varepsilon_j(\mathbf{x}), \quad (2)$$

where M is a realization of a mean zero *Gaussian random field*; that is, we think of M as being randomly sampled from a space of functions mapping $\mathfrak{R}^p \rightarrow \mathfrak{R}$. Therefore,

$Y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x})$ represents the unknown response surface at point \mathbf{x} . In this article we will focus with only one exception on the special case

$$Y(\mathbf{x}) = \beta_0 + \sum_{d=1}^p \beta_d x_d + M(\mathbf{x}). \quad (3)$$

Finally, $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \dots$ represents the independent and identically distributed sampling noise observed for each replication taken at design point \mathbf{x} . We sometimes refer to $M(\mathbf{x})$ and $\varepsilon_j(\mathbf{x})$ as the extrinsic and intrinsic uncertainties, respectively, at design point \mathbf{x} , as they were defined in Ankenman et al. [2010].

Yin et al. [2010] also propose an extension of kriging to stochastic simulation. Their metamodel is similar to Eq. (2), except that $\varepsilon_j(\mathbf{x})$ is also modeled as a Gaussian random field that is independent of M , and they take a fully Bayesian approach by treating all of the model parameters as having prior distributions. While directly accounting for parameter uncertainty, their model does not allow the effect of CRN to be separated from the spatial structure of the intrinsic variance of the simulation output.

For most of the analysis in this article we assume that the variance $V = V(\mathbf{x}) \equiv \text{Var}[\varepsilon(\mathbf{x})]$ at all design points is equal, while allowing the possibility that $\rho(\mathbf{x}, \mathbf{x}') \equiv \text{Corr}[\varepsilon_j(\mathbf{x}), \varepsilon_j(\mathbf{x}')] > 0$ due to CRN. In most discrete-event simulation settings the variance of the intrinsic noise $V(\mathbf{x})$ depends (perhaps strongly) on the location of design point, \mathbf{x} , and one of the key contributions of stochastic kriging is to address experiment design and analysis when this is the case. However, there are a number of reasons that we will not consider heterogeneous intrinsic variance except in the empirical study: In practice, $V(\mathbf{x})$ can take many forms, making it nearly impossible to obtain useful expressions for the effect of CRN. Further, if the variance of the noise depends on \mathbf{x} , then complicated experiment design techniques (e.g., as developed in Ankenman et al. [2010]) are needed to properly counteract the effects of the nonconstant variance. Once again, this would not lead to tractable results. In some sense, the equal variance assumption used in this article is intended to represent the conditions after the proper experiment design strategy has mitigated the effects of the nonconstant variance. We do include one example in the empirical study that manifests nonconstant $V(\mathbf{x})$ as a check that our conclusions are unaffected.

In our setting an experiment design consists of n simulation replications taken at all k design points $\{\mathbf{x}_i\}_{i=1}^k$. When we assume equal variances, then taking n the same at all design points seems reasonable and again greatly simplifies the analysis; furthermore, equal n is appropriate for CRN so that replication j has a companion for all design points.

Let the sample mean of simulation output at \mathbf{x}_i be

$$\begin{aligned} \bar{Y}(\mathbf{x}_i) &= \frac{1}{n} \sum_{j=1}^n \mathcal{Y}_j(\mathbf{x}_i) \\ &= Y(\mathbf{x}_i) + \frac{1}{n} \sum_{j=1}^n \varepsilon_j(\mathbf{x}_i) \\ &= \beta_0 + \sum_{d=1}^p \beta_d x_d + M(\mathbf{x}_i) + \frac{1}{n} \sum_{j=1}^n \varepsilon_j(\mathbf{x}_i) \end{aligned} \quad (4)$$

and let $\bar{Y} = (\bar{Y}(\mathbf{x}_1), \bar{Y}(\mathbf{x}_2), \dots, \bar{Y}(\mathbf{x}_k))^\top$. Define $\Sigma_M(\mathbf{x}, \mathbf{x}') = \text{Cov}[M(\mathbf{x}), M(\mathbf{x}')] to be the covariance of points \mathbf{x} and \mathbf{x}' implied by the extrinsic spatial correlation model; and let the $k \times k$ matrix Σ_M be the extrinsic spatial variance-covariance matrix of the $k$$

design points $\{\mathbf{x}_i\}_{i=1}^k$. Finally, let \mathbf{x}_0 be the prediction point, and define $\Sigma_M(\mathbf{x}_0, \cdot)$ to be the $k \times 1$ vector that contains the extrinsic spatial covariances between \mathbf{x}_0 and each of the k design points; that is,

$$\Sigma_M(\mathbf{x}_0, \cdot) = (\text{Cov}[M(\mathbf{x}_0), M(\mathbf{x}_1)], \text{Cov}[M(\mathbf{x}_0), M(\mathbf{x}_2)], \dots, \text{Cov}[M(\mathbf{x}_0), M(\mathbf{x}_k)])^\top.$$

Since M is stationary, Σ_M and $\Sigma_M(\mathbf{x}_0, \cdot)$ are of the following form

$$\Sigma_M = \tau^2 \begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{pmatrix} \quad \text{and} \quad \Sigma_M(\mathbf{x}_0, \cdot) = \tau^2 \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{pmatrix},$$

where $\tau^2 > 0$ is the extrinsic spatial variance. Gradient estimation only makes sense if the response surface is differentiable. The differentiability of Gaussian process models like Eq. (3) depend on the differentiability of its spatial correlation function as the distance between design points goes to zero. See, for instance, Santner et al. [2003, Section 2.3.4]. In particular, the sample paths are infinitely differentiable if the popular Gaussian correlation function is used. Therefore we choose to adopt the Gaussian correlation function $\text{Corr}[M(\mathbf{x}_i), M(\mathbf{x}_\ell)] = \exp\{-\sum_{j=1}^p \theta_j (x_{ij} - x_{\ell j})^2\}$ in this article. To simplify notation, the spatial correlation between the design point \mathbf{x}_i and the prediction point \mathbf{x}_0 is $r_i = \text{Corr}[M(\mathbf{x}_0), M(\mathbf{x}_i)]$, and the spatial correlation between two design points \mathbf{x}_h and \mathbf{x}_i is $r_{hi} = \text{Corr}[M(\mathbf{x}_h), M(\mathbf{x}_i)]$. To obtain tractable results, the spatial correlation parameter is assumed the same across all dimensions in this article, that is, $\theta_j = \theta$, $j = 1, 2, \dots, p$. This assumption, although not always appropriate in practice, helps facilitate the analysis and demonstrate the theme of this article without introducing unnecessary technical difficulties. We remove this restriction in the empirical study.

To make the k -point models tractable, in forthcoming Sections 3.2 and 4.2 we let $\Sigma_M = \tau^2 \mathbf{I}_k$ where \mathbf{I}_k denotes the $k \times k$ identity matrix. This form of Σ_M indicates that the design points are spatially uncorrelated with one another, which might be plausible if the design points are widely separated in the region of interest. In addition, to derive results for the k -point trend model in Section 4.2, we further assume that $\Sigma_M(\mathbf{x}_0, \cdot) = \tau^2(r_0, r_0, \dots, r_0)^\top$; this scenario might be plausible if the design points are widely separated, say at the extremes of the region of interest, while \mathbf{x}_0 is central. These assumptions are useful for insight and tractability, but not necessary for stochastic kriging.

What distinguishes stochastic kriging from kriging is that we account for the sampling variability inherent in a stochastic simulation. Let Σ_ε be the $k \times k$ variance-covariance matrix implied by the sample average intrinsic noise with (h, i) element

$$\Sigma_\varepsilon(\mathbf{x}_h, \mathbf{x}_i) = \text{Cov} \left[\sum_{j=1}^n \varepsilon_j(\mathbf{x}_h)/n, \sum_{j=1}^n \varepsilon_j(\mathbf{x}_i)/n \right]$$

across all design points \mathbf{x}_h and \mathbf{x}_i . The anticipated effect of CRN is to cause the off-diagonal elements of Σ_ε to be positive. To make our results tractable in Sections 3 and 4, we let

$$\Sigma_\varepsilon = \frac{\mathbf{V}}{n} \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{pmatrix}, \quad (5)$$

where $\rho > 0$, meaning we assume equal variance and correlation. Again, these assumptions are useful for insight and tractability, but not necessary for stochastic kriging. The MSE-optimal predictor (metamodel) provided by stochastic kriging takes the form

$$\widehat{Y}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \widehat{\boldsymbol{\beta}} + \Sigma_M(\mathbf{x}_0, \cdot)^\top [\Sigma_M + \Sigma_\varepsilon]^{-1} (\bar{Y} - \mathbf{F} \widehat{\boldsymbol{\beta}}),$$

where the rows of \mathbf{F} are $\mathbf{f}(\mathbf{x}_1)^\top, \mathbf{f}(\mathbf{x}_2)^\top, \dots, \mathbf{f}(\mathbf{x}_k)^\top$. In the mathematical analysis in Sections 3 and 4, we will suppose that only $\boldsymbol{\beta}$ needs to be estimated, while $\Sigma_M, \Sigma_\varepsilon$ and $\Sigma_M(\mathbf{x}_0, \cdot)$ are known. In Section 5, we consider what happens when Σ_ε is estimated, and numerically assess its impact on prediction performance. Finally, our empirical studies in Sections 6 and 7 will estimate every parameter and reexamine the effects of CRN in this context.

3. INTERCEPT MODELS

In kriging metamodeling for deterministic computer experiments, the most common form is the intercept model (no other trend terms, better known as “ordinary kriging”) since (it is argued) the random field term M is flexible enough to account for any variation across the response surface. In this section, we study intercept models and how the use of CRN affects parameter estimation, prediction, and gradient estimation. All results are derived in the online supplement to this article.

3.1. A Two-Point Intercept Model

Consider the two-point intercept model $Y_j(x) = \beta_0 + M(x) + \varepsilon_j(x)$ with β_0 unknown, design points x_1 and x_2 with equal numbers of replications n , and prediction point x_0 , with $x_i \in \mathfrak{N}, i = 0, 1, 2$. Therefore, $Y(x_0) = \beta_0 + M(x_0)$ is the response that we want to predict, β_0 is the parameter we need to estimate, and $dY(x_0)/dx_0$ is the gradient of interest.

The Best Linear Unbiased Predictor (BLUP) of $Y(x_0)$, the stochastic kriging predictor, is

$$\widehat{Y}(x_0) = \frac{\bar{Y}(x_1) + \bar{Y}(x_2)}{2} + \frac{\tau^2 \left(\frac{\bar{Y}(x_1) - \bar{Y}(x_2)}{2} \right)}{\tau^2(1 - r_{12}) + \frac{V}{n}(1 - \rho)} (r_1 - r_2) \quad (6)$$

with MSE

$$\text{MSE}^* = \tau^2 (1 - (r_1 + r_2)) + \frac{1}{2} \left[\tau^2(1 + r_{12}) + \frac{V}{n}(1 + \rho) - \frac{\tau^4(r_1 - r_2)^2}{\tau^2(1 - r_{12}) + \frac{V}{n}(1 - \rho)} \right]. \quad (7)$$

We can show that $d\text{MSE}^*/d\rho$ is always positive, hence it follows that the use of CRN, which tends to increase ρ , increases the MSE^* of the best linear unbiased predictor for this two-point intercept model. Notice that for the spatial variance-covariance matrix of $(Y(x_0), \bar{Y}(x_1), \bar{Y}(x_2))^\top$ to be positive definite, the following condition must be satisfied: $-r_{12}^2 + 2r_1 r_2 r_{12} + 1 - (r_1^2 + r_2^2) > 0$.

The Best Linear Unbiased Estimator (BLUE) of β_0 corresponding to the BLUP of $Y(x_0)$ is

$$\widehat{\beta}_0 = \frac{\bar{Y}(x_1) + \bar{Y}(x_2)}{2} \quad (8)$$

and it is easy to see that its variance is increasing in ρ since it is a sum of positively correlated outputs. Thus, the MSE of prediction and the variance of $\widehat{\beta}_0$ are both inflated by CRN.

Let $\widehat{\nabla}_{\text{sk}}$ denote the gradient of the predictor $\widehat{Y}(x_0)$ at x_0 in the stochastic kriging setting. Under the assumptions given in Section 2, it follows that

$$\begin{aligned}\widehat{\nabla}_{\text{sk}} &= \frac{d\widehat{Y}(x_0)}{dx_0} \\ &= -2\theta[r_1(x_0 - x_1) + r_2(x_2 - x_0)] \frac{\tau^2 \left(\frac{\bar{y}(x_1) - \bar{y}(x_2)}{2} \right)}{[\tau^2(1 - r_{12}) + \frac{V}{n}(1 - \rho)]}.\end{aligned}\quad (9)$$

To assess the impact of CRN, we choose as a benchmark the gradient estimator that would be obtained if there were no simulation intrinsic variance; that is, if the response surface could be observed noise-free. We are interested in the impact of CRN on the “distance” between the noisy and noise-free gradient estimators to measure whether CRN helps mitigate the effect of intrinsic variance on gradient estimation.

Let $\widehat{\nabla}_{\text{sk}}(n)$ be the gradient estimator when n simulation replications are used at each design point, and let $\widehat{\nabla}_{\text{sk}}(\infty)$ be the gradient estimator as $n \rightarrow \infty$, which can be obtained by simply setting the intrinsic variance $V = 0$ in Eq. (9). It follows that

$$\mathbb{E}[\widehat{\nabla}_{\text{sk}}(n) - \widehat{\nabla}_{\text{sk}}(\infty)]^2 = \frac{2\theta^2(r_1(x_0 - x_1) + r_2(x_2 - x_0))^2}{((1 - r_{12})/[\frac{V}{n}(1 - \rho)] + 1/\tau^2)(1 - r_{12})}.\quad (10)$$

From Eq. (10), we see that CRN decreases the mean squared difference between these two estimators. In the extreme case as $\rho \rightarrow 1$, even if n is not large, the gradient estimator from stochastic kriging converges to the “ideal” case because the effect of stochastic noise on gradient estimation is eliminated by employing CRN.

3.2. A k -Point Intercept Model

In the previous section we were able to show that CRN is detrimental to response surface prediction and parameter estimation, but is beneficial to gradient estimation in a two-design-point setting. In this section we are able to draw the same conclusions in a particular k -point ($k \geq 2$) intercept model, $Y_j(\mathbf{x}) = \beta_0 + \mathbf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x})$, with β_0 unknown. Under the assumptions given in Section 2, the following results can be obtained.

The BLUP of $Y(\mathbf{x}_0)$ is

$$\widehat{Y}(\mathbf{x}_0) = \frac{1}{k} \sum_{i=1}^k \bar{Y}(\mathbf{x}_i) + \frac{\tau^2}{(\frac{V}{n}(1 - \rho) + \tau^2)} \left(\sum_{i=1}^k r_i \bar{Y}(\mathbf{x}_i) - \frac{1}{k} \left(\sum_{i=1}^k \bar{Y}(\mathbf{x}_i) \right) \left(\sum_{i=1}^k r_i \right) \right)\quad (11)$$

with MSE

$$\begin{aligned}\text{MSE}^* &= \tau^2 + \frac{\tau^4}{\frac{V}{n}(1 - \rho) + \tau^2} \left(\frac{1}{k} \left(\sum_{i=1}^k r_i \right)^2 - \sum_{i=1}^k r_i^2 \right) \\ &\quad + \frac{\frac{V}{n}((k - 1)\rho + 1) + \tau^2}{k} - 2\tau^2 \left(\frac{1}{k} \sum_{i=1}^k r_i \right).\end{aligned}\quad (12)$$

Notice that for the spatial variance-covariance matrix of $(Y(\mathbf{x}_0), \bar{Y}(\mathbf{x}_1), \dots, \bar{Y}(\mathbf{x}_k))^T$ to be positive definite, it must be that $\sum_{i=1}^k r_i^2 < 1$. We show in the online supplement to this article that under this condition $d\text{MSE}^*/d\rho$ is positive for any $\rho \in [0, 1)$, hence CRN increases MSE^* .

The BLUE of β_0 corresponding to the BLUP of $Y(\mathbf{x}_0)$ is

$$\widehat{\beta}_0 = \frac{1}{k} \sum_{i=1}^k \widehat{Y}(\mathbf{x}_i) \quad (13)$$

and its variance is easily shown to be an increasing function of ρ .

Similar to the analysis of gradient estimation in Section 3.1, let $\widehat{\nabla}_{\text{sk}} = (\widehat{\nabla}_{\text{sk}_1}, \widehat{\nabla}_{\text{sk}_2}, \dots, \widehat{\nabla}_{\text{sk}_p})^\top$ denote the gradient of $\widehat{Y}(\mathbf{x}_0)$ at \mathbf{x}_0 in the stochastic kriging setting; notice that now $\widehat{\nabla}_{\text{sk}}$ is a random vector in \mathfrak{R}^p . We can show that for $j = 1, 2, \dots, p$, the j th component of the gradient is

$$\begin{aligned} \widehat{\nabla}_{\text{sk}_j} &= \frac{\partial \widehat{Y}(\mathbf{x}_0)}{\partial x_{0j}} \\ &= \frac{-2\theta\tau^2}{\tau^2 + \frac{\nu}{n}(1-\rho)} \cdot \sum_{i=1}^k \left(\left(\widehat{Y}(\mathbf{x}_i) - \frac{1}{k} \sum_{h=1}^k \widehat{Y}(\mathbf{x}_h) \right) (x_{0j} - x_{ij}) r_i \right), \end{aligned} \quad (14)$$

where the i th design point $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^\top$ is a vector in \mathfrak{R}^p , $i = 1, \dots, k$. Recall that $r_i = \exp\{-\theta \sum_{j=1}^p (x_{0j} - x_{ij})^2\}$ is the spatial correlation between \mathbf{x}_i and \mathbf{x}_0 , and that we assume that the design points are spatially approximately uncorrelated, meaning that they are separated enough that $r_{ij} \approx 0$, for $i \neq j$.

Now for $p > 2$, we continue to use $\widehat{\nabla}_{\text{sk}}(\infty)$ as the benchmark to evaluate gradient estimation in the stochastic kriging setting. We use the inner product to measure the “distance” between the two random vectors $\widehat{\nabla}_{\text{sk}}(n)$ and $\widehat{\nabla}_{\text{sk}}(\infty)$ at the prediction point $\mathbf{x}_0 \in \mathfrak{R}^p$ and call it the mean squared difference between these two gradient estimators. We can show that

$$\begin{aligned} \langle \widehat{\nabla}_{\text{sk}}(n) - \widehat{\nabla}_{\text{sk}}(\infty), \widehat{\nabla}_{\text{sk}}(n) - \widehat{\nabla}_{\text{sk}}(\infty) \rangle &= \sum_{j=1}^p \mathbf{E} \left[(\widehat{\nabla}_{\text{sk}_j}(n) - \widehat{\nabla}_{\text{sk}_j}(\infty))^2 \right] \\ &= \frac{4\theta^2}{\left(\frac{\nu}{n(1-\rho)} + \frac{1}{\tau^2} \right)} \sum_{j=1}^p \left(\sum_{i=1}^k (x_{0j} - x_{ij})^2 r_i^2 - \frac{1}{k} \left(\sum_{i=1}^k (x_{0j} - x_{ij}) r_i \right)^2 \right). \end{aligned} \quad (15)$$

As in Section 3.1, we arrive at the conclusion that for this k -point intercept model, CRN decreases the mean squared difference between these two gradient estimators.

4. TREND MODELS

Although many practitioners use intercept models for kriging, it remains to be seen what models will be most effective when noise is introduced. Also, in linear regression models, CRN is known to be most helpful for estimating slope parameters and so it seems likely that CRN will perform best under a trend model that, like a regression model, includes slope parameters. For these reasons and for completeness, we next study the effects of CRN on stochastic kriging with a linear trend model (the counterpart of “universal kriging”).

4.1. A Two-Point Trend Model

Consider the two-point trend model $\mathcal{Y}_j(x) = \beta_0 + \beta_1 x + \mathbf{M}(x) + \varepsilon_j(x)$ with β_0 and β_1 unknown, so that $Y(x_0) = \beta_0 + \beta_1 x_0 + \mathbf{M}(x_0)$ is the unknown response that we want to predict at point x_0 . Without loss of generality, suppose that $x_1 < x_2$. Then we can show the following results.

The BLUP of $Y(x_0)$ is

$$\widehat{Y}(x_0) = \frac{\bar{Y}(x_2)(x_0 - x_1) + \bar{Y}(x_1)(x_2 - x_0)}{(x_2 - x_1)} \quad (16)$$

with MSE

$$\text{MSE}^* = 2\tau^2 + \frac{V}{n} - \frac{2ab}{(a+b)^2} \left[\tau^2(1 - r_{12}) + \frac{V}{n}(1 - \rho) \right] - \frac{2\tau^2}{(a+b)}(ar_1 + br_2), \quad (17)$$

where $a = x_2 - x_0$, $b = x_0 - x_1$, $a + b = x_2 - x_1$. Eq. (17) implies that for this two-point trend model, when $x_0 \in (x_1, x_2)$, CRN increases MSE^* ; however, if we do extrapolation, that is, $x_0 \notin (x_1, x_2)$, then CRN will decrease MSE^* . Notice that the literature on kriging claims that kriging does not perform well in extrapolation, so kriging should be restricted to interpolation. Finally, if $x_0 = x_1$ or x_2 , we get $\widehat{Y}(x_0) = \bar{Y}(x_1)$ or $\bar{Y}(x_2)$, respectively; in this case MSE^* is reduced to V/n , the same with and without using CRN.

The BLUE of $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ corresponding to the BLUP of $Y(x_0)$ is

$$\widehat{\boldsymbol{\beta}} = \frac{1}{(x_2 - x_1)} \begin{pmatrix} x_2 \bar{Y}(x_1) - x_1 \bar{Y}(x_2) \\ \bar{Y}(x_2) - \bar{Y}(x_1) \end{pmatrix}. \quad (18)$$

It follows that

$$\text{Var}(\widehat{\beta}_0) = \left(\tau^2 + \frac{V}{n} \right) + \frac{2x_1x_2}{(x_2 - x_1)^2} \left[\tau^2(1 - r_{12}) + \frac{V}{n}(1 - \rho) \right] \quad (19)$$

$$\text{Var}(\widehat{\beta}_1) = \frac{2 \left[\tau^2(1 - r_{12}) + \frac{V}{n}(1 - \rho) \right]}{(x_2 - x_1)^2} \quad (20)$$

and

$$\text{Cov}(\widehat{\beta}_0, \widehat{\beta}_1) = \frac{-(x_1 + x_2) \left[\tau^2(1 - r_{12}) + \frac{V}{n}(1 - \rho) \right]}{(x_2 - x_1)^2}.$$

From Eq. (20), we see that CRN reduces the variance of $\widehat{\beta}_1$. Also notice that Eq. (19) implies that if $x_1x_2 < 0$, so that 0 is interior to the design space, then CRN inflates the variance of $\widehat{\beta}_0$, while if $x_1x_2 > 0$, so that $\widehat{\beta}_0$ is an extrapolated prediction of the response at $x = 0$, then CRN decreases the variance of $\widehat{\beta}_0$.

Finally, following the analysis in Section 3.1, we can show that the mean squared difference between the gradient estimators obtained when the number of replications n is finite and when $n \rightarrow \infty$ is

$$\text{E} \left[\widehat{\nabla}_{\text{sk}}(n) - \widehat{\nabla}_{\text{sk}}(\infty) \right]^2 = \frac{2V(1 - \rho)}{n(x_1 - x_2)^2}. \quad (21)$$

Eq. (21) shows that CRN decreases the mean squared difference between these two estimators. Observe that the extrinsic spatial variance τ^2 has no influence on this mean squared difference at all.

4.2. A k-Point Trend Model

For the two-point trend model we were able to draw conclusions similar to those we found for the intercept model and an additional conclusion related to the estimation of the slope parameter. Specifically, we found that CRN is detrimental to response surface prediction at any point *inside the region of experimentation* since it increases the MSE of prediction, but CRN is beneficial to estimation of the slope parameter by decreasing the variance of its estimator and beneficial to gradient estimation since it decreases the effect of noise. As with the intercept model we can extend the conclusions of the

two-point trend model to a k -point ($k \geq 2$) trend model if additional restrictions are made.

Consider the k -point trend model $\mathcal{Y}_j(\mathbf{x}) = \beta_0 + \sum_{d=1}^p \beta_d x_d + \mathbf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x})$, where $p \geq 2$. Suppose that we have a $k \times (p + 1)$ orthogonal design matrix \mathbf{D}_k of rank $p + 1$

$$\mathbf{D}_k = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{k1} & \dots & x_{kp} \end{pmatrix}$$

which means that the column vectors of \mathbf{D}_k are pairwise orthogonal. Such an assumption on \mathbf{D}_k is not yet common for kriging, because kriging usually employs space-filling designs such as a Latin hypercube sample, but orthogonal and nearly orthogonal Latin hypercube designs are being introduced. Nevertheless, in addition to the assumptions given in Section 2 orthogonality makes the analysis tractable enough to give the following results.

The BLUE of $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ corresponding to the BLUP of $\mathcal{Y}(\mathbf{x}_0)$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{D}_k^\top \Sigma^{-1} \mathbf{D}_k)^{-1} \mathbf{D}_k^\top \Sigma^{-1} \bar{\mathcal{Y}}, \quad (22)$$

where $\Sigma = \Sigma_M + \Sigma_\varepsilon$. More explicitly,

$$\hat{\beta}_0 = \frac{1}{k} \sum_{i=1}^k \bar{\mathcal{Y}}(\mathbf{x}_i) \quad (23)$$

and

$$\hat{\beta}_j = \frac{\sum_{i=1}^k x_{ij} \bar{\mathcal{Y}}(\mathbf{x}_i)}{\sum_{i=1}^k x_{ij}^2}, \quad j = 1, 2, \dots, p. \quad (24)$$

The resulting BLUP of $\mathcal{Y}(\mathbf{x}_0)$ is

$$\hat{\mathcal{Y}}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \hat{\boldsymbol{\beta}}, \quad (25)$$

where $\mathbf{f}(\mathbf{x}_0) = (1, x_{01}, x_{02}, \dots, x_{0p})^\top$. The corresponding optimal MSE is

$$\begin{aligned} \text{MSE}^* &= \tau^2 \left(1 + \frac{1}{k} + \sum_{j=1}^p \frac{x_{0j}^2}{\sum_{i=1}^k x_{ij}^2} - 2r_0 \right) + \frac{1}{k} \frac{\mathbf{V}}{n} \left(1 + k \sum_{j=1}^p \frac{x_{0j}^2}{\sum_{i=1}^k x_{ij}^2} \right) \\ &\quad + \frac{1}{k} \frac{\mathbf{V}}{n} \rho \left((k-1) - k \sum_{j=1}^p \frac{x_{0j}^2}{\sum_{i=1}^k x_{ij}^2} \right). \end{aligned} \quad (26)$$

Notice that if

$$\frac{k-1}{k} > \sum_{j=1}^p \frac{x_{0j}^2}{\sum_{i=1}^k x_{ij}^2} \quad (27)$$

then CRN increases MSE^* .

To help interpret this result, consider a $k = 2^p$ factorial design where the design points are $x_{ij} \in \{-1, +1\}$. Then Eq. (27) reduces to $\sum_{j=1}^p x_{0j}^2 < k - 1$. Therefore, CRN will inflate the MSE^* of $\hat{\mathcal{Y}}(\mathbf{x}_0)$ at prediction points inside a sphere of radius $\sqrt{2^p - 1}$ centered at the origin (which is also the center of the experiment design). Notice that

for $p > 1$ we have $\sqrt{2^p - 1} > \sqrt{p}$, the radius of the sphere that just contains the design points and is the usual prediction region of interest. Also notice that when $p = 1$ we recover the condition for the two-point trend model, for which we have more general results available in Section 4.1 without the orthogonality assumption.

We next focus on the effect of CRN on $\text{Cov}(\hat{\boldsymbol{\beta}})$. Because of the orthogonality assumption, the expression for $\text{Cov}(\hat{\boldsymbol{\beta}})$ becomes much simpler. It can be shown that

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{D}_k^\top \Sigma^{-1} \mathbf{D}_k)^{-1} = \begin{pmatrix} \frac{\frac{V}{n}[1+(k-1)\rho]+\tau^2}{k} & 0 & \dots & 0 \\ 0 & \frac{\frac{V}{n}(1-\rho)+\tau^2}{\sum_{i=1}^k x_{i1}^2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\frac{V}{n}(1-\rho)+\tau^2}{\sum_{i=1}^k x_{ip}^2} \end{pmatrix}. \quad (28)$$

Hence we arrive at a similar conclusion to the one obtained in Section 4.1: CRN reduces the variances of $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$. Here the first diagonal term manifests that CRN increases $\text{Var}(\hat{\beta}_0)$, which is consistent with Section 4.1 since $\mathbf{0}$ is interior to the design space.

Now let

$$\hat{\nabla}_{\text{sk}} = (\hat{\nabla}_{\text{sk}_1}, \hat{\nabla}_{\text{sk}_2}, \dots, \hat{\nabla}_{\text{sk}_p})^\top$$

denote the gradient of $\hat{Y}(\mathbf{x}_0)$ at \mathbf{x}_0 in the stochastic kriging setting. We can show that for $j = 1, 2, \dots, p$, the j th component of the gradient is

$$\begin{aligned} \hat{\nabla}_{\text{sk}_j} &= \frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{0j}} \\ &= \frac{d\Sigma_{\text{M}}(\mathbf{x}_0, \cdot)}{dx_{0j}} \Sigma^{-1} (\bar{Y} - \mathbf{D}_k \hat{\boldsymbol{\beta}}) + \hat{\beta}_j \\ &= \hat{\beta}_j. \end{aligned}$$

Following the analysis in Section 3.2, we define the following inner product to measure the “distance” between the two random vectors $\hat{\nabla}_{\text{sk}}(n)$ and $\hat{\nabla}_{\text{sk}}(\infty)$ at prediction point \mathbf{x}_0 .

$$\begin{aligned} \langle \hat{\nabla}_{\text{sk}}(n) - \hat{\nabla}_{\text{sk}}(\infty), \hat{\nabla}_{\text{sk}}(n) - \hat{\nabla}_{\text{sk}}(\infty) \rangle &= \sum_{j=1}^p \mathbf{E} \left[(\hat{\nabla}_{\text{sk}_j}(n) - \hat{\nabla}_{\text{sk}_j}(\infty))^2 \right] \\ &= \frac{V}{n} (1 - \rho) \sum_{j=1}^p \left(\sum_{i=1}^k x_{ij}^2 \right)^{-1} \end{aligned} \quad (29)$$

Eq. (29) shows that CRN decreases the mean squared difference between these two gradient estimators. Similar to the result in Section 4.1, we see that only the intrinsic noise affects this mean squared difference, whereas the extrinsic spatial variance has no influence on it at all.

5. ESTIMATING THE INTRINSIC VARIANCE-COVARIANCE MATRIX

To this point in the article we have assumed that Σ_ε , the variance-covariance matrix of the intrinsic simulation noise, was known. However, a key component of stochastic kriging is estimating Σ_ε ; to examine the impact of estimating it we will need additional assumptions that are consistent with those in Ankenman et al. [2010]. This will allow

us to prove that estimating the intrinsic variance-covariance matrix does not lead to biased prediction. Then, by modifying the k -point intercept model in Section 3.2, we study the impact of estimating the common intrinsic variance when there is correlated random noise among design points induced by CRN. In the following analysis, we treat Σ_ε as unknown but everything else as known including β .

We begin with formally stating the following assumption.

ASSUMPTION 5.1.

- (1) *The random field M is a stationary Gaussian random field.*
- (2) *For design point \mathbf{x}_i , $i = 1, 2, \dots, k$, the random noise from different replications $\varepsilon_1(\mathbf{x}_i), \varepsilon_2(\mathbf{x}_i), \dots$ are independent and identically distributed (i.i.d.) $N(0, V(\mathbf{x}))$.*
- (3) *For the j th replication, $j = 1, 2, \dots, n$, the $k \times 1$ vector of random noise across all design points $[\varepsilon_j(\mathbf{x}_1), \varepsilon_j(\mathbf{x}_2), \dots, \varepsilon_j(\mathbf{x}_k)]^\top$ has a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\tilde{\Sigma}_\varepsilon$ (with usage of CRN).*
- (4) *The random noise is independent of M .*

Notice that the only new condition added here, given those already stated in Section 2, is the multivariate normality of the random noise across all design points in the same simulation replication (Condition 3). This, along with Condition 2, will be most appropriate when the replication results $\mathcal{Y}_j(\mathbf{x}_i)$ are themselves the averages of a large number of within-replication outputs (e.g., the average customer waiting time from replication j is the average of many individual customer waiting times). Condition 1 is how we characterize our uncertainty about the true response surface and is always an approximation, while Condition 4 can be justified because we allow the variance of the random noise $V(\mathbf{x})$ to be a function of location \mathbf{x} , eliminating any remaining dependence on M .

Under Assumption 5.1, the multivariate normality of $(Y(\mathbf{x}_0), \bar{Y}(\mathbf{x}_1), \dots, \bar{Y}(\mathbf{x}_k))$ follows from a proof similar to Ankenman et al. [2010]. The stochastic kriging predictor (41) given at the beginning of Section A.2 of the online supplement of this article, is the conditional expectation of $Y(\mathbf{x}_0)$ given \bar{Y} . The $k \times 1$ vector of sample average random noise at all k design points $[\bar{\varepsilon}(\mathbf{x}_1), \bar{\varepsilon}(\mathbf{x}_2), \dots, \bar{\varepsilon}(\mathbf{x}_k)]^\top$ has a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix Σ_ε , where $\bar{\varepsilon}(\mathbf{x}_i) = n^{-1} \sum_{j=1}^n \varepsilon_j(\mathbf{x}_i)$, $i = 1, 2, \dots, k$ and $\varepsilon_j(\mathbf{x}_i)$ is the random noise at design point \mathbf{x}_i in the j th replication. It follows that $\Sigma_\varepsilon = n^{-1} \tilde{\Sigma}_\varepsilon$.

Now let S denote the sample variance-covariance matrix of the intrinsic noise across the k design points. We have

$$S = \begin{pmatrix} S_{11} & S_{12} & \dots & S_{1k} \\ S_{21} & S_{22} & \dots & S_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ S_{k1} & S_{k2} & \dots & S_{kk} \end{pmatrix}, \quad (30)$$

where

$$\begin{aligned} S_{i\ell} &= \frac{1}{n-1} \sum_{j=1}^n (\mathcal{Y}_j(\mathbf{x}_i) - \bar{Y}(\mathbf{x}_i))(\mathcal{Y}_j(\mathbf{x}_\ell) - \bar{Y}(\mathbf{x}_\ell)) \\ &= \frac{1}{n-1} \sum_{j=1}^n (\varepsilon_j(\mathbf{x}_i) - \bar{\varepsilon}(\mathbf{x}_i))(\varepsilon_j(\mathbf{x}_\ell) - \bar{\varepsilon}(\mathbf{x}_\ell)). \end{aligned} \quad (31)$$

In words, $S_{i\ell}$ is the sample covariance of the random noise at design points \mathbf{x}_i and \mathbf{x}_ℓ , $i, \ell = 1, 2, \dots, k$. We use $n^{-1}S$ to estimate Σ_ε . The next result shows that estimating Σ_ε in this way introduces no prediction bias. The proof can be found in Section A.8 of the online supplement to this article.

THEOREM 5.2. *Let $\widehat{\Sigma}_\varepsilon = n^{-1}S$, where S is specified as in Eq. (31). Define*

$$\widehat{Y}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \boldsymbol{\beta} + \Sigma_M(\mathbf{x}_0, \cdot)^\top [\Sigma_M + \widehat{\Sigma}_\varepsilon]^{-1}(\bar{Y} - \mathbf{F}\boldsymbol{\beta}), \quad (32)$$

where $\mathbf{f}(\mathbf{x}_i)$ denotes the $(q+1) \times 1$ vector of functions $\mathbf{f}(\mathbf{x}_i)$, $i = 0, 1, \dots, k$ and \mathbf{F} is the $k \times (q+1)$ model matrix of full rank

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}(\mathbf{x}_1)^\top \\ \mathbf{f}(\mathbf{x}_2)^\top \\ \vdots \\ \mathbf{f}(\mathbf{x}_k)^\top \end{pmatrix}.$$

If Assumption 5.1 holds, then $E[\widehat{Y}(\mathbf{x}_0) - Y(\mathbf{x}_0)] = 0$.

Recall the k -point intercept model $\mathcal{Y}_j(\mathbf{x}) = \beta_0 + \mathbf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x})$, where the design points $\{\mathbf{x}_i\}_{i=1}^k$ are in \mathfrak{R}^p and equal numbers of replications n are obtained from each of them. In Ankenman et al. [2010] the effect of estimating intrinsic variance was investigated assuming the intrinsic noise at each design point to be independent and identically distributed with a common intrinsic variance. Following Ankenman et al. [2010], we next focus on how much variance inflation occurs when Σ_ε is estimated under the same assumptions as in Ankenman et al. [2010] but with the addition of CRN. Suppose

$$\Sigma_M = \tau^2 \begin{pmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \vdots & \vdots & \ddots & \vdots \\ r & r & \cdots & 1 \end{pmatrix}$$

and $\Sigma_M(\mathbf{x}_0, \cdot) = \tau^2(r_0, r_0, \dots, r_0)^\top$ with $r_0, r \geq 0$. This represents a situation in which the extrinsic spatial correlations among the design points are all equal and the design points are equally correlated with the prediction point.

Notice that for the spatial variance-covariance matrix of $(Y(\mathbf{x}_0), \bar{Y}(\mathbf{x}_1), \dots, \bar{Y}(\mathbf{x}_k))^\top$ to be positive definite, the condition $r_0^2 < 1/k + r(k-1)/k$ must be satisfied. To make the analysis tractable but still interesting, we assume that $\widetilde{\Sigma}_\varepsilon$ has the following form, with ρ known and V unknown.

$$\widetilde{\Sigma}_\varepsilon = V \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

Hence it follows that $\Sigma_\varepsilon = n^{-1}\widetilde{\Sigma}_\varepsilon$. As in Ankenman et al. [2010], we suppose that there is an estimator \widehat{V} of V such that $\widehat{V} \sim V\chi_{n-1}^2/(n-1)$, namely $(n-1)\widehat{V}/V$ has a chi-squared distribution with degrees of freedom $n-1$. In Section A.9 of the online supplement to this article we show that the MSE of $\widehat{Y}(\mathbf{x}_0)$, the stochastic kriging predictor with V known, is

$$\text{MSE}^* = \tau^2 \left(1 - \frac{kr_0^2}{1 + C_{\rho\gamma} + (k-1)r} \right), \quad (33)$$

where $C_{\rho\gamma} = \frac{\gamma}{n}(1 + (k-1)\rho)$ and $\gamma = V/\tau^2$ denotes the ratio of the intrinsic variance to the extrinsic variance, which is (roughly speaking) a measure of the sampling noise relative to the response surface variation. On the other hand, the MSE of $\widehat{Y}(\mathbf{x}_0)$ obtained by substituting \widehat{V} for V is

$$\text{MSE} = \tau^2 \mathbb{E} \left[1 + \frac{kr_0^2 [1 + C_{\rho\gamma} + (k-1)r]}{\left(1 + \frac{\widehat{V}}{V} C_{\rho\gamma} + (k-1)r\right)^2} - \frac{2kr_0^2}{\left(1 + \frac{\widehat{V}}{V} C_{\rho\gamma} + (k-1)r\right)} \right]. \quad (34)$$

We assess the MSE inflation and the effect of CRN on it by evaluating the ratio of (34) to (33) numerically. The MSE inflation ratio is largest when n is small and r_0 and r are large, so in the numerical analysis we show the inflation ratio as a function of $\gamma = V/\tau^2$ and ρ for $n = 10$, $r = 0, 0.1, 0.2$ and r_0 at 95% of the maximum value it can take. We use $k = 50$ design points throughout the study for convenient comparison with the results given in Ankenman et al. [2010].

We summarize our findings as follows and refer readers to Section A.10 of the online supplement to this article for a detailed discussion. There is a penalty associated with estimating intrinsic variance; that is, doing so always inflates prediction MSE relative to using the (unknown) true value of Σ_ε . However, for a fixed value of the spatial correlation of a given response surface, CRN can either magnify or diminish this penalty depending on the ratio of the intrinsic variance to the extrinsic variance, or in other words, depending on which source of variation dominates for that particular response surface. The MSE inflation that results from estimating Σ_ε is even more substantial in the presence of CRN when spatial variation dominates intrinsic variation ($\tau^2 \gg V$). On the other hand, the MSE inflation from estimating Σ_ε is diminished by using CRN when intrinsic variation dominates spatial variation ($\tau^2 \ll V$).

These effects of CRN on MSE inflation hold for response surfaces with varying degrees of smoothness. Interestingly, we found that *the smoothness of the response surface actually matters*. Specifically, strong spatial correlation of the response surface tends to counteract the effect of CRN on MSE inflation, whatever it is. A response surface with “strong spatial correlation” tends to be smoother than one with weaker spatial dependence, since the value of the response at any point tends to be similar to—that is, strongly correlated with—other points in close proximity. When CRN magnifies the MSE inflation, then strong spatial correlation reduces the magnification. On the other hand, when CRN diminishes the MSE inflation, it is less effective at doing so when the surface exhibits strong spatial correlation.

Lastly we suggest that discretion needs to be exercised when one interprets the results preceding, because even if the MSE inflation ratio is close to 1, the MSE* itself can be large; therefore ratio = 1 does not mean that the particular setting provides a good prediction. Similarly, a large MSE inflation ratio does not necessarily imply that a particular experimental setting provides poor prediction. Finally from the discussion in Section A.10 of the online supplement to this article we conclude that even with this small value of n (recall that $n = 10$), the MSE inflation ratio is slight over an extreme range of $\gamma = V/\tau^2$. As n increases, the inflation vanishes. This suggests that the penalty for estimating V will typically be small.

6. AN EXPERIMENT WITH GAUSSIAN RANDOM FIELDS

From the two-point and k -point intercept and trend models we gained some insight into the impact of CRN on parameter estimation, prediction, and gradient estimation for stochastic kriging. However, to obtain these results we had to assume all model parameters except β (Sections 3–4) and Σ_ε (Section 5) were known. In this section, we

confirm these insights empirically when all parameters must be estimated. The factors we investigate are the strength of the correlation ρ induced by CRN; the number of design points k ; the strength of the extrinsic spatial correlation coefficient θ ; and the ratio of the intrinsic variance to the extrinsic variance $\gamma = V/\tau^2$.

We consider a one-dimensional problem where the true response surface is $Y(x) = 10 + 3x + M(x)$ with $x \in [0, 1]$. The Gaussian random field M , denoted by $\text{GRF}(\tau^2, \theta)$, has extrinsic spatial covariance between points x and x' given by $\Sigma_M(x, x') = \tau^2 \exp\{-\theta(x - x')^2\}$. A test-function instance is created by sampling $M \sim \text{GRF}(\tau^2, \theta)$, and we sample multiple instances as part of the experiment. We fix $\tau^2 = 1$ but θ is varied to obtain smooth and rough response surface instances.

The simulation response observed at point x on replication j is $\mathcal{Y}_j(x) = 10 + 3x + M(x) + \varepsilon_j(x)$, where the random noise $\varepsilon_j(x)$, $j = 1, 2, \dots, n$ is i.i.d. $\mathcal{N}(0, V)$; since we assume equal variance it is reasonable to take the same number of replications, n , at each design point. The effect of CRN is represented by specifying a common correlation $\rho = \text{Corr}[\varepsilon_j(x), \varepsilon_j(x')]$ for $x \neq x'$, $j = 1, 2, \dots, n$. We vary $\gamma = V/\tau^2 = V$ to introduce random noise of different relative intensities.

An equally spaced grid design of k design points $x \in [0, 1]$ is used, with $k \in \Omega_k = \{4, 7, 13, 25\}$. We make $n = 100$ replications at each design point, and control V so that $\gamma/n = V/n \in \Omega_\gamma = \{0.01, 0.25, 1\}$, corresponding to low, medium, and high intrinsic variance. We took $\theta \in \Omega_\theta = \{4.6052, 13.8155\}$ (or equivalently, $\exp(-\theta) \in \{0.01, 10^{-6}\}$, where $\exp(-\theta)$ is the correlation between the most distant design points in $[0, 1]$); notice that small θ tends to give a smoother response surface. We vary ρ in $\Omega_\rho = \{0, 0.4, 0.8\}$ to assess the effect of increasing correlation induced by CRN; for each $\theta \in \Omega_\theta$ we sample 10 true response surfaces, and for each response surface we run 5 macroreplications for each $\{k, \rho, \gamma\} \in \Omega_k \times \Omega_\rho \times \Omega_\gamma$ combination; for a fixed $\{\theta, k, \rho, \gamma\}$ and response surface, the macroreplications differ only in their randomly sampled $\varepsilon_j(x)$.

Thus, altogether there are $2 \times 10 \times 4 \times 3 \times 3 \times 5 = 3600$ experiments. For each one we fit a stochastic kriging metamodel using maximum likelihood estimation as described in Ankenman et al. [2010], do prediction and gradient estimation at 193 equally spaced points in $[0, 1]$, and record the values of the estimated parameters $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\tau}^2$ and $\hat{\theta}$. The stochastic kriging code used in these experiments can be found at www.stochastickriging.net.

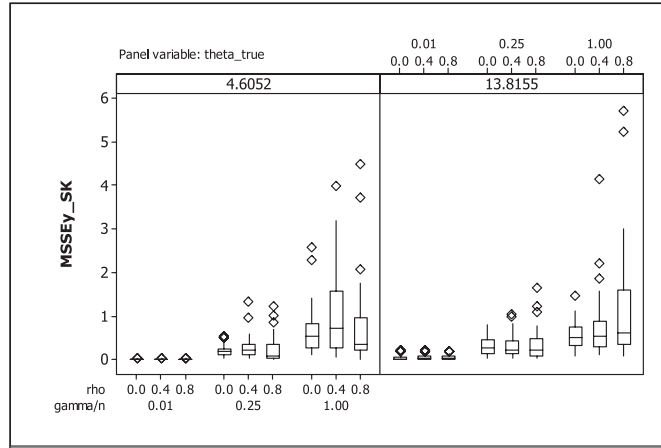
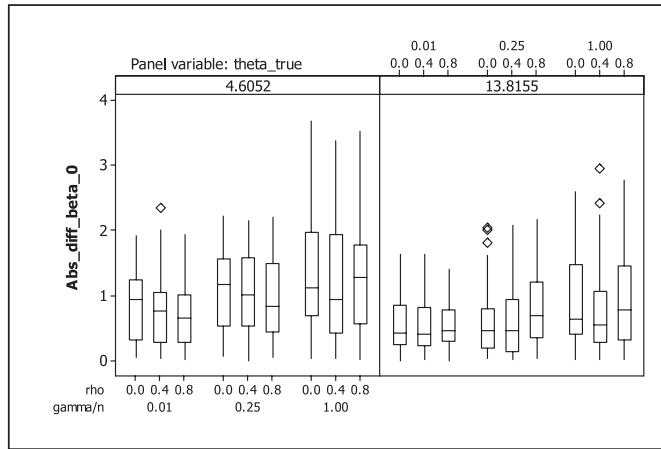
We evaluate the impact on prediction by MSSE, the mean of the sum of squared errors of the predicted values at the 193 check points, namely, $\text{MSSE}(\hat{Y}) = 193^{-1} \sum_{i=1}^{193} (Y(\mathbf{x}_i) - \hat{Y}(\mathbf{x}_i))^2$; we evaluate parameter estimation by recording the absolute difference between the true and estimated parameter on each trial; and we evaluate gradient estimation by computing the sample correlation between the true and estimated gradient across the 193 check points.

A brief preview of our findings is as follows.

- CRN does not aid prediction and instead increases the MSSE.
- CRN does reduce the variability of the slope estimator $\hat{\beta}_1$.
- CRN does improve gradient estimation in the sense that it introduces a strong positive correlation between the estimated gradient and the true gradient.

These findings are consistent with our results in the previous sections.

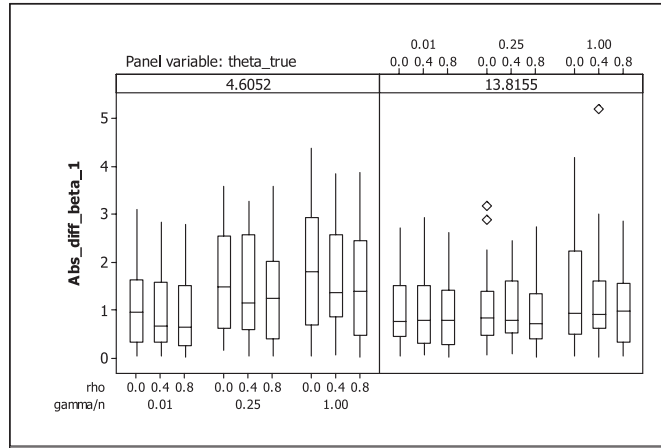
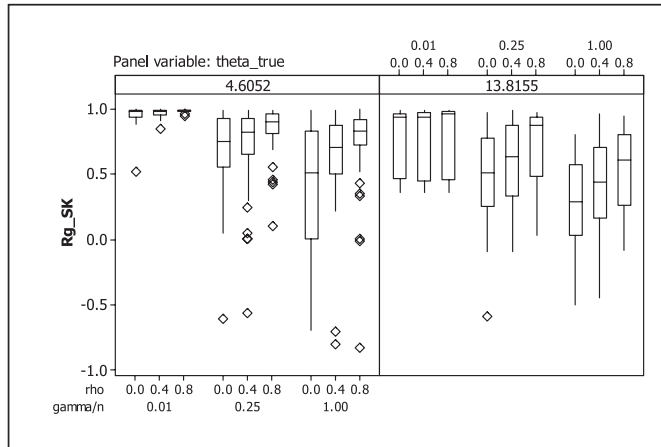
Boxplots in Figures 1–4 provide more details. For brevity, we show only graphs corresponding to the number of design points $k = 7$. In each figure, the left panel shows the sample statistics of interest obtained from the smoother response surface with $\theta = 4.6052$; while the right panel shows the statistics obtained from the rougher response surface with $\theta = 13.8155$. Within each panel from the left to the right, 3 groups of boxplots are ordered by increasing γ/n ; within each group, three individual

Fig. 1. MSSE for $k = 7$.Fig. 2. $|\beta_0 - \hat{\beta}_0|$ for $k = 7$.

boxplots are ordered by increasing ρ . Notice that each individual boxplot is a summary of 50 data points from 5 macroreplications on each of 10 surfaces.

To evaluate prediction, we calculate the MSSE by averaging the squared difference between the true response $Y(x_0)$ and the predicted value $\hat{Y}(x_0)$ at x_0 across 193 check points. A summary of the MSSE for $k = 7$ is shown in Figure 1. It is easy to see that increasing ρ increases MSSE, and leads to wider interquartile range. This is especially true when θ is large, or equivalently, when the extrinsic spatial correlation is small. As we expected, for fixed ρ , increasing γ/n will increase MSSE. On the other hand, we mention (without showing graphs) that for fixed ρ , increasing the number of design points k leads to narrower interquartile range when γ/n is not large. Finally, by observing the sets of three boxplots that are grouped close together to show the effect of increasing ρ , we conclude that CRN does not help prediction.

For parameter estimation, we use the Absolute Deviation (AD), that is, $|\beta_j - \hat{\beta}_j|$. A summary of the statistical dispersion of $|\beta_j - \hat{\beta}_j|$, $j = 0$ and 1 for $k = 7$ is shown in Figures 2 and 3. For $|\beta_1 - \hat{\beta}_1|$, we see in Figure 3 that increasing ρ decreases $|\beta_1 - \hat{\beta}_1|$;

Fig. 3. $|\beta_1 - \hat{\beta}_1|$ for $k = 7$.Fig. 4. $\text{Corr}(\hat{\nabla}_{\text{true}}(n), \hat{\nabla}_{\text{sk}}(n))$ for $k = 7$.

this effect is more evident when θ is small. The effect of ρ on $|\beta_0 - \hat{\beta}_0|$ is not as obvious as on $|\beta_1 - \hat{\beta}_1|$. As we expected, for fixed ρ , increasing γ/n leads to increased ADs and wider interquartile range for both parameters. Finally, we mention (without showing graphs) that increasing the number of design points k moves the interquartile range closer to 0 and helps to estimate slope parameter even better. We conclude that CRN improves the estimation of the slope parameter, but its effect on estimating the intercept parameter is not as clear.

To evaluate gradient estimation, we use the correlation between the true gradient and the gradient estimated in the stochastic kriging setting instead of using the mean squared difference between them, since in most applications it is far more important to find the correct direction of change rather than the magnitude of this change. Therefore, $\text{Corr}(\hat{\nabla}_{\text{true}}(n), \hat{\nabla}_{\text{sk}}(n))$ gives a better view of the effect of ρ on gradient estimation under the influence of θ and k . We use the finite-difference gradient estimate from the noiseless response data as the true gradient $\hat{\nabla}_{\text{true}}(n)$. A summary of the correlations between $\hat{\nabla}_{\text{sk}}(n)$ and $\hat{\nabla}_{\text{true}}(n)$ for $k = 7$ is shown in Figure 4. It is obvious that

increasing ρ consistently increases $\text{Corr}(\widehat{\nabla}_{\text{true}}(n), \widehat{\nabla}_{\text{sk}}(n))$ for all γ/n values, and makes the interquantile range narrower as well as moving them toward 1; in fact, this effect is more manifest when γ/n is large. Typically, for fixed ρ , increasing γ/n decreases $\text{Corr}(\widehat{\nabla}_{\text{true}}(n), \widehat{\nabla}_{\text{sk}}(n))$ and leads to wider interquantile range. Furthermore, we mention (without showing graphs) that increasing the number of design points k increases $\text{Corr}(\widehat{\nabla}_{\text{true}}(n), \widehat{\nabla}_{\text{sk}}(n))$ and makes the interquantile range narrower. We conclude that CRN improves gradient estimation by introducing a strong positive correlation between the estimated gradient and the true gradient.

For each parameter set $\{\rho, \gamma, \theta, k\}$, we also estimated τ^2 and θ for the 10 response surfaces, each with 5 macro-replications. The estimates $\widehat{\theta}$ and $\widehat{\tau}^2$ obtained are not as good as $\widehat{\beta}_0$ and $\widehat{\beta}_1$ when compared to their known true values. For brevity, we choose not to present these results.

7. M/M/ ∞ QUEUE SIMULATION

In this section, we move a step closer to realistic system simulation problems. Let $Y(\mathbf{x})$ be the expected steady-state number of customers in an M/M/ ∞ queue with arrival rate x_1 and mean service time x_2 ; it is known that $Y(\mathbf{x}) = x_1 x_2$ and the distribution of the steady-state queue length is Poisson with mean $Y(\mathbf{x})$. Notice that the variance of the response is $x_1 x_2$ which changes across the design space.

Therefore, given values for x_1 and x_2 we can simulate a steady-state observation by generating a Poisson variate with mean $x_1 x_2$. Given a set of design points $\{x_{i1}, x_{i2}\}_{i=1}^{k'}$, we induce correlation across design points by using the inverse CDF method [Law and Kelton 2000], where k' denotes the number of design points used. Specifically, for replication j

$$\mathcal{Y}_j(\mathbf{x}_i) = F_{\mathbf{x}_i}^{-1}(\mathcal{U}_j), \quad i = 1, 2, \dots, k', \quad (35)$$

where $\mathcal{U}_1, \mathcal{U}_2, \dots, \mathcal{U}_n$ are i.i.d. $U(0,1)$; n is the number of simulation replications, and $F_{\mathbf{x}_i}^{-1}(\cdot)$ represents the inverse CDF of a Poisson distribution with mean $x_{i1} x_{i2}$. Notice that our experiment differs from what would occur in practice because we only take a single observation of the queue length on each replication, rather than the average queue length over some period of time. This allows us to compute the correlation induced by CRN, values of which typically are greater than 0.9 in this setting.

In stochastic kriging when the response surface $Y(\mathbf{x})$ is unknown, we assume that it takes the form $Y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} + M(\mathbf{x})$. Three different specifications of $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$ are considered to evaluate the effects of CRN: they are

Model 1. An intercept-only model, $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0$;

Model 2. A misspecified trend model, $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$;

Model 3. A correctly specified trend model, $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$.

By “correctly specified” we mean that Model 3 can recover the true response surface $x_1 x_2$ while the other two cannot.

Our experiment design is as follows. We consider the design space $1 \leq x_d \leq 5$, $d = 1, 2$. For design points we use a Latin hypercube sample of $k \in \{5, 20, 50\}$ points, and augment the design with the four corner points (1, 1), (1, 5), (5, 1) and (5, 5) to avoid extrapolation. Thus, there are $k' = k + 4$ design points in total. At each design point $n = 400$ simulation replications are made either using CRN as in Eq. (35), or sampled independently. We then fit stochastic kriging metamodels with trend terms specified as Models 1, 2, and 3 and make 100 macroreplications of the entire experiment.

For each model specification, we evaluate the impact on predication by $\widehat{\text{MISE}}(\widehat{Y})$, the approximated mean integrated squared error of \widehat{Y} . We evaluate gradient estimation by

Table I. The Scaled Mean Integrated Squared Error of Predictions Obtained for the Three Response Models with and Without Using CRN

k'	Model 1		Model 2		Model 3	
	Indep.	CRN	Indep.	CRN	Indep.	CRN
9	82.5 (9.8)	68 (7)	501 (112)	267 (99)	12 (1)	13 (2)
24	20 (1)	55 (9)	25 (2)	65 (10)	6.2 (0.4)	13 (2)
54	9.5 (0.6)	72 (11)	12 (1)	73 (10)	4.0 (0.3)	7.0 (1.2)

Table II. The Scaled Mean Integrated Squared Error of Gradient Estimates Obtained for the Three Response Models with and without Using CRN

k'	Model 1				Model 2				Model 3			
	Indep.		CRN		Indep.		CRN		Indep.		CRN	
	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{01}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{02}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{01}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{02}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{01}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{02}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{01}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{02}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{01}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{02}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{01}}$	$\frac{\partial \hat{Y}(\mathbf{x}_0)}{\partial x_{02}}$
9	84 (10)	90 (15)	8.4 (1.4)	8.8 (1.5)	748 (204)	782 (209)	388 (177)	367 (169)	3.7 (1.1)	3.4 (0.4)	0.48 (0.07)	0.50 (0.08)
24	30 (3)	31 (4)	2.8 (0.4)	2.8 (0.3)	39 (4)	64 (15)	3.2 (0.3)	3.4 (0.4)	2.2 (0.4)	3.4 (0.6)	0.37 (0.05)	0.39 (0.05)
54	15 (2)	19 (3)	3.0 (0.5)	2.9 (0.4)	23 (4)	26 (3)	2.9 (0.4)	2.9 (0.4)	6.7 (2.1)	13 (3)	0.26 (0.04)	0.29 (0.05)

Table III. Results for the Slope Parameters for the Correctly Specified Trend Model with and without Using CRN

k'	$\hat{\beta}_0(\beta_0 = 0)$		$\hat{\beta}_1(\beta_1 = 0)$		$\hat{\beta}_2(\beta_2 = 0)$		$\hat{\beta}_3(\beta_3 = 1)$	
	Indep.	CRN	Indep.	CRN	Indep.	CRN	Indep.	CRN
9	-0.009 (0.0095)	-0.003 (0.0024)	-0.001 (0.0040)	-0.001 (0.0009)	0.001 (0.0043)	-0.001 (0.0009)	1.000 (0.0016)	1.000 (0.0003)
24	-0.016 (0.0090)	-0.004 (0.0028)	0.004 (0.0031)	-0.001 (0.0007)	0.006 (0.0036)	-0.001 (0.0007)	0.998 (0.0012)	1.000 (0.0002)
54	0.008 (0.0082)	-0.006 (0.0021)	-0.003 (0.0030)	0.000 (0.0005)	-0.004 (0.0029)	0.000 (0.0005)	1.001 (0.0010)	1.000 (0.0002)

computing $\widehat{\text{MISE}}(\widehat{\nabla}_{\text{sk}_d})$, $d = 1$ and 2 . In both cases, we approximate MISE by using a 2500 check-point grid in $[1, 5]^2$. Formally,

$$\widehat{\text{MISE}}(\hat{Y}) = \frac{1}{100} \sum_{\ell=1}^{100} \frac{1}{2500} \sum_{i=1}^{2500} (Y(\mathbf{x}'_i) - \hat{Y}_\ell(\mathbf{x}'_i))^2$$

and

$$\widehat{\text{MISE}}(\widehat{\nabla}_{\text{sk}_d}) = \frac{1}{100} \sum_{\ell=1}^{100} \frac{1}{2500} \sum_{i=1}^{2500} (\nabla_d(\mathbf{x}'_i) - \widehat{\nabla}_{\text{sk}_d}(\mathbf{x}'_i, \ell))^2,$$

where the Integrated Squared Error (ISE) is approximated by averaging the sum of squared errors over the 2500 check points and the Mean Integrated Squared Error (MISE) is approximated by averaging the approximated ISE over 100 macroreplications. Notice that for better presentation of the results, values shown in Tables I and II are calculated without the scaling factor $1/2500$. Finally we give summary statistics for the parameter estimates of the correctly specified trend model (Model 3).

The effects of CRN on prediction, gradient estimation, and parameter estimation can be found in Tables I–III. The values in parentheses are the corresponding standard errors. In brief, we found that the results derived in the previous sections still hold;

that is, CRN improves gradient estimation and estimation of slope parameters, but does not aid prediction.

In Table I, we observe for all three model specifications that the $\widehat{\text{MISE}}(\hat{Y})$ is smaller with independent sampling than with CRN, with the exception of Models 1 and 2 when the number of design points is very small ($k' = 9$); increasing the number of design points make this effect even more apparent. Notice that the misspecified trend model (Model 2) gives even worse prediction results than the intercept model (Model 1), while the correctly specified trend model is much better.

In Table II, it is observed that CRN improves gradient estimation for all three response models. The values for the sample means and standard errors obtained on the correctly specified trend model are much smaller than the corresponding values from the other two response models. We observe once again that the misspecified trend model (Model 2) gives much worse gradient estimates than the intercept model (Model 1) does.

Lastly, we are interested in knowing how CRN affects estimates of the slope parameters for the correctly specified trend model. The results given in Table III manifest that CRN reduces variances of the slope parameter estimates to a great extent; increasing the number of design points does not improve the results much in this case. Notice that if the correctly specified trend model is assumed, one is able to successfully recover the true response model with moderately large number of replications.

8. CONCLUSIONS

CRN is one of the most widely used variance reduction techniques; in fact, with most simulation software one would have to carefully program the simulation to avoid using it. Therefore, it is important to understand its effect on a new metamodeling technique such as stochastic kriging. Previous research with other metamodels, such as linear regression, has shown that CRN often leads to more precise parameter estimation, especially with slope parameters that are essentially gradients. However, since CRN can inflate the variability of parameters such as the intercept, it can reduce the precision of the actual prediction.

The parameters, the form, and even the underlying assumptions of stochastic kriging are substantially different from traditional metamodels. Nevertheless, in this article we have provided compelling evidence that CRN has effects on the stochastic kriging metamodel that are similar or at least analogous to the effects seen in more traditional metamodel settings. Specifically, we have used a variety of tractable models to show that CRN leads to: (1) less precise prediction of the response surface in terms of MSE, (2) better estimation of the slope terms in any trend model, and (3) better gradient estimation.

In addition, we are able to show that under Assumption 5.1 in Section 5, estimating the intrinsic variance-covariance matrix Σ_ε introduces no prediction bias to the plug-in BLUP. A thorough numerical analysis of the MSE inflation that is induced by estimating Σ_ε revealed that stronger spatial correlation counteracts the effect of CRN on MSE inflation.

Finally, through an experiment with Gaussian random fields and an M/M/ ∞ queue example we assessed the impact of CRN on prediction, parameter estimation, and gradient estimation when the parameters of the trend model β , of the random field τ^2 and θ , and the intrinsic variance-covariance matrix Σ_ε are all estimated as would be required in actual application. The conclusions given by the empirical evaluation were consistent with our analytical results.

The implications of our results are that when the actual prediction values matter, CRN is not recommended. Such scenarios might occur in financial risk analysis or tactical decision making where the primary purpose of the metamodel is to produce

predictions of the response in places where no actual simulations have been made and the predictions are needed quickly (long before actual simulation runs would finish). CRN is recommended for use in gradient estimation for simulation optimization or if the metamodel is a physics-based model where better parameter estimates are of great value to, say, establish sensitivities. Sensitivity analysis is also particularly useful for verification and validation of simulation models. Since CRN substantially improves the performance of stochastic kriging gradient estimators, a fruitful area for future research is applying stochastic kriging and CRN to simulation optimization.

ACKNOWLEDGMENTS

We would like to thank the referees and editors for comments and suggestions for improvement of our article.

REFERENCES

- ANKENMAN, B., NELSON, B. L., AND STAUM, J. 2010. Stochastic kriging for simulation metamodeling. *Oper. Res.* 58, 371–382.
- ANKENMAN, B. E., NELSON, B. L., AND STAUM, J. 2008. Stochastic kriging for simulation metamodeling. In *Proceedings of the Winter Simulation Conference*, S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, Eds., IEEE, Los Alamitos, CA, 362–370.
- CHEN, X., ANKENMAN, B., AND NELSON, B. L. 2010. Common random numbers and stochastic kriging. In *Proceedings of the Winter Simulation Conference*, B. Johansson, S. Jain, J. Montoya-Torres, J. Hagan, and E. Yücesan, Eds., IEEE, Los Alamitos, CA, 947–956.
- DONOHUE, J. M., HOUCK, R. C., AND MYERS, R. H. 1992. Simulation designs for quadratic response surface models in the presence of model misspecification. *Manag. Sci.* 38, 1765–1791.
- DONOHUE, J. M., HOUCK, R. C., AND MYERS, R. H. 1995. Simulation designs for the estimation of quadratic response surface gradients in the presence of model misspecification. *Manag. Sci.* 41, 244–262.
- GRAYBILL, F. A. 1969. *Matrices with Applications in Statistics* 2nd Ed. Wadsworth, Belmont, CA.
- HUSSEY, J. R., MYERS, R. H., AND HOUCK, E. C. 1987a. Correlated simulation experiments in first-order response surface design. *Oper. Res.* 35, 744–758.
- HUSSEY, J. R., MYERS, R. H., AND HOUCK, E. C. 1987b. Pseudorandom number assignment in quadratic response surface designs. *IIE Trans.* 19, 395–403.
- KLEIJNEN, J. P. C. 1975. Antithetic variates, common random numbers and optimal computer time allocation in simulation. *Manag. Sci.* 21, 1176–1185.
- KLEIJNEN, J. P. C. 1988. Analyzing simulation experiments with common random numbers. *Manag. Sci.* 34, 65–74.
- KLEIJNEN, J. P. C. 1992. Regression metamodels for simulation with common random numbers: Comparison of validation tests and confidence intervals. *Manag. Sci.* 38, 1164–1185.
- LAW, A. M. AND KELTON, W. D. 2000. *Simulation Modelling and Analysis* 3rd Ed. McGraw Hill, New York.
- NOZARI, A., ARNOLD, S. F., AND PEGDEN, C. D. 1987. Statistical analysis for use with the Schruben and Margolin correlation induction strategy. *Oper. Res.* 35, 127–139.
- SANTNER, T. J., WILLIAMS, B. J., AND NOTZ, W. I. 2003. *The Design and Analysis of Computer Experiments*. Springer, New York.
- SCHRUBEN, L. W. AND MARGOLIN, B. H. 1978. Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *J. Amer. Statist. Assoc.* 73, 504–525.
- STEIN, M. L. 1999. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.
- TEW, J. D. AND WILSON, J. R. 1992. Validation of simulation analysis methods for the Schruben-Margolin correlation-induction strategy. *Oper. Res.* 40, 87–103.
- TEW, J. D. AND WILSON, J. R. 1994. Estimating simulation metamodels using combined correlation-based variance reduction techniques. *IIE Trans.* 26, 2–16.
- YIN, J., NG, S. H., AND NG, K. M. 2010. A Bayesian metamodeling approach for stochastic simulations. In *Proceedings of the Winter Simulation Conference*, B. Johansson, S. Jain, J. Montoya-Torres, J. Hagan, and E. Yücesan, Eds. IEEE, Los Alamitos, CA, 1055–1066.

Received May 2010; revised September 2011; accepted October 2011