Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

COMMON RANDOM NUMBERS AND STOCHASTIC KRIGING

Xi Chen
Bruce Ankenman
Barry L. Nelson

Department of Industrial Engineering & Management Sciences
Northwestern University
Evanston, Illinois 60208-3119 U.S.A.

ABSTRACT

We use a collection of simple models to examine the interaction between the variance reduction technique of common random numbers and a new simulation metamodeling technique called stochastic kriging. We consider the impact of common random numbers on prediction, parameter estimation and gradient estimation.

1 INTRODUCTION

Beginning with the seminal paper of Schruben and Margolin (1978), simulation researchers have been interested in the impact of incorporating common random numbers (CRN) into experiment designs for fitting linear metamodels of the form

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

to the output of stochastic simulation experiments. In Model (1), $Y(\mathbf{x})$ is the simulation output, $\mathbf{x} = (x_1, x_2, \ldots, x_p)^\top$ 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.

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 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 an intercept term. See, for instance, Donohue, Houck, and Myers (1992, 1995), Hussey, Myers and Houck (1987a,b), Kleijnen (1988, 1992), Nozari, Arnold, and Pegden (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 paper is to undertake a similar analysis of the interaction of CRN and a new metamodeling technique called stochastic kriging (Ankenman, Nelson, and Staum 2008, 2010). Stochastic kriging is an extension of kriging as applied to deterministic computer experiments (see, for instance, Santner et al. 2003) to stochastic simulation. Specifically, we explore the effects of CRN on prediction, parameter estimation and gradient estimation for stochastic kriging through variations of two-point and k-point models that are rich enough to offer some enlightening insights while still being analytically tractable.

Ankenman, Nelson, and Staum (2010) used a two-point problem with all parameters known to show that CRN increases the mean squared error (MSE) of the MSE-optimal predictor at a prediction point that has equal extrinsic spatial correlation with the two design points. They

978-1-4244-9864-2/10/$26.00 ©2010 IEEE

speculated that CRN will not be helpful for prediction in general. In this paper we generalize their two-point problem to allow unequal spatial correlations between the design points and the prediction point, and drop the assumption that the trend-model parameters are known. Even with these generalizations we are still able to show that CRN is detrimental to prediction with respect to MSE. We then extend the result given in Appendix A.2 in Ankenman, Nelson, and Staum (2010) for $k \geq 2$ spatially uncorrelated design points and show that CRN inflates the MSE of prediction. We assume that the trend parameters are unknown whereas Ankenman, Nelson, and Staum (2010) assumed that all parameters are known. In contrast to prediction, we show that CRN typically improves the estimation of trend-model parameters by reducing the variances of the slope parameters. Finally, we show that CRN improves gradient estimation in the sense that the gradient estimators from stochastic kriging are less affected by intrinsic simulation noise when CRN is employed.

2 STOCHASTIC KRIGING

In this section we briefly review stochastic kriging as developed in Ankenman, Nelson, and Staum (2010) and the particular simplifications we exploit in this paper.

In stochastic kriging we represent the simulation's output on replication j at design point $\mathbf{x}$ as $\mathcal{Y}_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top\boldsymbol{\beta} + \mathsf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x}) = \mathsf{Y}(\mathbf{x}) + \varepsilon_j(\mathbf{x})$ where M is a realization of a mean 0 *Gaussian random field*; that is, we think of M as being randomly sampled from a space of functions mapping $\Re^p \to \Re$. Therefore, $\mathsf{Y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top\boldsymbol{\beta} + \mathsf{M}(\mathbf{x})$ represents the unknown response surface at point $\mathbf{x}$. In this paper we will only consider the special case

$$\mathsf{Y}(\mathbf{x}) = \beta_0 + \sum_{d=1}^{p} \beta_d x_d + \mathsf{M}(\mathbf{x}); \tag{2}$$

that is, the trend model is linear. Finally, $\varepsilon_1(\mathbf{x}), \varepsilon_2(\mathbf{x}), \ldots$ represents the independent and identically distributed intrinsic noise observed for each replication taken at design point, $\mathbf{x}$.

Throughout this paper we assume that the variance $\mathsf{V} = \mathsf{V}(\mathbf{x}) \equiv \mathrm{Var}[\varepsilon(\mathbf{x})]$ at all design points is equal, while allowing the possibility that $\rho(\mathbf{x},\mathbf{x}') \equiv \mathrm{Corr}[\varepsilon_j(\mathbf{x}), \varepsilon_j(\mathbf{x}')] > 0$ due to CRN. In most discrete-event simulation settings the variance of the intrinsic noise $\mathsf{V}(\mathbf{x})$ 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 here: In practice, $\mathsf{V}(\mathbf{x})$ 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, Nelson, and Staum 2010) are needed to properly counteract the effects of the non-constant variance. Once again, this would not lead to tractable results. In some sense, the equal variance assumption is what occurs after the proper experiment design strategy has mitigated the effects of the non-constant variance.

In our setting an experiment design consists of n simulation replications taken at all k design points $\{\mathbf{x}_i\}_{i=1}^k$. Given our equal variance assumption, assuming that n is equal for all design points seems reasonable and again greatly simplifies the analysis. Let the sample mean of simulation output at $\mathbf{x}_i$ be

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

and let $\bar{\mathcal{Y}} = \left(\bar{\mathcal{Y}}(\mathbf{x}_1), \bar{\mathcal{Y}}(\mathbf{x}_2), \ldots, \bar{\mathcal{Y}}(\mathbf{x}_k)\right)^\top$. Define $\Sigma_{\mathsf{M}}(\mathbf{x},\mathbf{x}') = \mathrm{Cov}[\mathsf{M}(\mathbf{x}), \mathsf{M}(\mathbf{x}')]$ 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 $\boldsymbol{\Sigma}_{\mathsf{M}}$ 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 $\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)$ 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,

$$\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot) = \left(\mathrm{Cov}[\mathsf{M}(\mathbf{x}_0),\mathsf{M}(\mathbf{x}_1)],\mathrm{Cov}[\mathsf{M}(\mathbf{x}_0),\mathsf{M}(\mathbf{x}_2)],\ldots,\mathrm{Cov}[\mathsf{M}(\mathbf{x}_0),\mathsf{M}(\mathbf{x}_k)]\right)^\top .$$

Since M is stationary, $\boldsymbol{\Sigma}_{\mathsf{M}}$ and $\boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)$ are of the following form:

$$\boldsymbol{\Sigma}_{\mathsf{M}} = \tau^2\begin{pmatrix} 1 & r_{12} & \ldots & r_{1k} \\ r_{21} & 1 & \ldots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \ldots & 1 \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot) = \tau^2\begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_k \end{pmatrix}$$

where $\tau^2 > 0$ is the extrinsic spatial variance.

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

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

across all design points $\mathbf{x}_h$ and $\mathbf{x}_i$. The anticipated effect of CRN is to cause the off-diagonal elements of $\boldsymbol{\Sigma}_\varepsilon$ to be positive. To make our results tractable, we let

$$\boldsymbol{\Sigma}_\varepsilon = \frac{\mathsf{V}}{n}\begin{pmatrix} 1 & \rho & \ldots & \rho \\ \rho & 1 & \ldots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \ldots & 1 \end{pmatrix} \tag{3}$$

where $\rho > 0$, meaning we assume equal variance and correlation, assumptions that are useful in this paper for drawing insight about the impact of CRN, but not necessary for stochastic kriging in general.

The MSE-optimal predictor (metamodel) provided by stochastic kriging takes the form

$$\widehat{\mathsf{Y}}(x_0) = \mathbf{f}(\mathbf{x}_0)^\top\widehat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\mathsf{M}}(\mathbf{x}_0,\cdot)^\top[\boldsymbol{\Sigma}_{\mathsf{M}}+\boldsymbol{\Sigma}_\varepsilon]^{-1}(\bar{\mathcal{Y}} - \mathbf{F}\widehat{\boldsymbol{\beta}})$$

where the row of $\mathbf{F}$ are $\mathbf{f}(\mathbf{x}_1)^\top, \mathbf{f}(\mathbf{x}_2)^\top,\ldots,\mathbf{f}(\mathbf{x}_k)^\top$. In our analysis we will suppose that only $\boldsymbol{\beta}$ needs to be estimated, while $\boldsymbol{\Sigma}_{\mathsf{M}}$, $\boldsymbol{\Sigma}_\varepsilon$ and $\boldsymbol{\Sigma}_{\mathsf{M}}(x_0,\cdot)$ are known. For estimation of all of the parameters see Ankenman, Nelson, and Staum (2010).

3 INTERCEPT MODELS

In kriging metamodeling for deterministic computer experiments, the most common form is the intercept model (no other trend terms) 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 will study intercept models and how the use of CRN affects parameter estimation, prediction and gradient estimation. All results are derived in Chen, Ankenman, and Nelson (2010).

3.1 A Two-Point Intercept Model

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

The best linear unbiased predictor (BLUP) of $\mathsf{Y}(x_0)$ is

$$\widehat{\mathsf{Y}}(x_0) = \frac{\bar{\mathcal{Y}}(x_1)+\bar{\mathcal{Y}}(x_2)}{2} + \frac{\tau^2\left(\frac{\bar{\mathcal{Y}}(x_1)-\bar{\mathcal{Y}}(x_2)}{2}\right)}{\tau^2(1-r_{12})+\frac{\mathsf{V}}{n}(1-\rho)}(r_1-r_2) \tag{4}$$

with MSE

$$\mathrm{MSE}^\star = \tau^2\left(1-(r_1+r_2)\right) + \frac{1}{2}\left[\tau^2(1+r_{12}) + \frac{\mathsf{V}}{n}(1+\rho) - \frac{\tau^4(r_1-r_2)^2}{\tau^2(1-r_{12})+\frac{\mathsf{V}}{n}(1-\rho)}\right]. \tag{5}$$

We can show that $d\mathrm{MSE}^\star/d\rho$ is always positive, hence it follows that CRN increases the $\mathrm{MSE}^\star$ of the best linear unbiased predictor for this two-point intercept model. Notice that for the spatial variance-covariance matrix of $(\mathsf{Y}(x_0),\bar{\mathcal{Y}}(x_1),\bar{\mathcal{Y}}(x_2))^\top$ to be positive definite, the following condition must be satisfied: $-r_{12}^2+2r_1r_2r_{12}+1-(r_1^2+r_2^2)>0$.

The best linear unbiased estimator (BLUE) of β_0 corresponding to the BLUP of $\mathsf{Y}(x_0)$ is

$$\hat{\beta}_0 = \frac{\bar{\mathcal{Y}}(x_1)+\bar{\mathcal{Y}}(x_2)}{2} \tag{6}$$

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 $\hat{\beta}_0$ are both inflated by CRN.

To analyze gradient estimation we need to impose some structure on the form of the spatial correlation function, so we assume the popular Gaussian correlation form. Thus, the spatial correlation between the design point x_i and the prediction point x_0 is $r_i = e^{-\theta(x_i-x_0)^2}$, $i=1,2$, and the spatial correlation between the two design points themselves is $r_{12}=e^{-\theta(x_1-x_2)^2}$. Let $\widehat{\nabla}_{\text{sk}}$ denote the gradient of the predictor $\widehat{\mathsf{Y}}(x_0)$ at x_0 in the stochastic kriging setting. Then

$$\begin{aligned}\widehat{\nabla}_{\text{sk}} &= \frac{d\widehat{\mathsf{Y}}(x_0)}{dx_0} \\ &= -2\theta\left[r_1(x_0-x_1)+r_2(x_2-x_0)\right]\frac{\tau^2\left(\frac{\bar{\mathcal{Y}}(x_1)-\bar{\mathcal{Y}}(x_2)}{2}\right)}{\left[\tau^2(1-r_{12})+\frac{\mathsf{V}}{n}(1-\rho)\right]}.\end{aligned} \tag{7}$$

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 noise and noise-free gradient estimators to measure whether CRN helps mitigate the effect of random noise 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 $n\to\infty$, which can be obtained by simply setting the intrinsic variance $\mathsf{V}=0$ in Equation (7). It follows that

$$\mathrm{E}\left[\widehat{\nabla}_{\text{sk}}(n)-\widehat{\nabla}_{\text{sk}}(\infty)\right]^2 = \frac{2\theta^2(r_1(x_0-x_1)+r_2(x_2-x_0))^2}{\left(\frac{1-r_{12}}{\frac{\mathsf{V}}{n}(1-\rho)}+\frac{1}{\tau^2}\right)(1-r_{12})}. \tag{8}$$

From Equation (8), we see that CRN decreases the mean squared difference between these two estimators. In the extreme case as $\rho\to 1$, even if n is not large, the gradient estimate from stochastic kriging converges to the "ideal" case because the effect of random 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 this section we are able to draw the same conclusions in a particular k-point ($k \geq 2$) intercept model, $\mathcal{Y}_j(\mathbf{x}) = \beta_0 + \mathsf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x})$, with β_0 unknown. To make the k-point model tractable we let $\boldsymbol{\Sigma}_{\mathsf{M}} = \tau^2\, \mathbf{I}_k$ where $\mathbf{I}_k$ denotes the $k \times k$ identity matrix. This form of $\boldsymbol{\Sigma}_{\mathsf{M}}$ 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. The following results can be shown:

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

$$\widehat{\mathsf{Y}}(\mathbf{x}_0) = \frac{1}{k}\sum_{i=1}^{k}\bar{\mathcal{Y}}(\mathbf{x}_i) + \frac{\tau^2}{\left(\frac{\mathsf{V}}{n}(1-\rho)+\tau^2\right)}\left(\sum_{i=1}^{k} r_i\bar{\mathcal{Y}}(\mathbf{x}_i) - \frac{1}{k}\left(\sum_{i=1}^{k}\bar{\mathcal{Y}}(\mathbf{x}_i)\right)\left(\sum_{i=1}^{k} r_i\right)\right) \tag{9}$$

with MSE

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

Notice that for the spatial variance-covariance matrix of $(\mathsf{Y}(\mathbf{x}_0), \bar{\mathcal{Y}}(\mathbf{x}_1), \ldots, \bar{\mathcal{Y}}(\mathbf{x}_k))^\top$ to be positive definite, it must be that $\sum_{i=1}^{k} r_i^2 < 1$. We show in Chen, et al. (2010) that under this condition $d\text{MSE}^{\star}/d\rho$ is positive for any $\rho \in [0,1]$, hence CRN increases MSE*.

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

$$\hat{\beta}_0 = \frac{1}{k}\sum_{i=1}^{k}\bar{\mathcal{Y}}(\mathbf{x}_i) \tag{10}$$

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

Similar to the analysis of gradient estimation in Section 3.1, we continue to use the Gaussian correlation function. In addition, because now $\mathbf{x} \in \Re^p$, we make a further simplifying assumption that θ, the spatial correlation parameter, is the same across all dimensions; i.e., $\theta_j = \theta,\ j = 1,2,\ldots,p$. Let $\widehat{\nabla}_{\text{sk}} = \left(\widehat{\nabla}_{\text{sk}_1}, \widehat{\nabla}_{\text{sk}_2}, \ldots, \widehat{\nabla}_{\text{sk}_p}\right)^\top$ denote the gradient of $\widehat{\mathsf{Y}}(\mathbf{x}_0)$ at $\mathbf{x}_0$ in the stochastic kriging setting; notice that now $\widehat{\nabla}_{\text{sk}}$ is a random vector in $\Re^p$. We can show that for $j = 1,2,\ldots,p$, the jth component of the gradient is

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

where the ith point $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ is a vector in $\Re^p, i = 0,1,\ldots,k$. Notice that $r_i = e^{-\theta\sum_{j=1}^{p}(x_{0j}-x_{ij})^2}$ is the spatial correlation between $\mathbf{x}_i$ and $\mathbf{x}_0$. Recall that we assume that the design points are spatially uncorrelated.

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

$$\langle \widehat{\nabla}_{\text{sk}}(n) - \widehat{\nabla}_{\text{sk}}(\infty), \widehat{\nabla}_{\text{sk}}(n) - \widehat{\nabla}_{\text{sk}}(\infty) \rangle = \sum_{j=1}^{p} \mathrm{E}\left[\left(\widehat{\nabla}_{\text{sk}_j}(n) - \widehat{\nabla}_{\text{sk}_j}(\infty)\right)^2\right]$$

$$= \frac{4\theta^2}{\left(\frac{1}{\frac{\mathsf{V}}{n}(1-\rho)} + \frac{1}{\tau^2}\right)} \sum_{j=1}^{p} \left(\sum_{i=1}^{k} (x_{0j} - x_{ij})^2 r_i^2 - \frac{1}{k} \left(\sum_{i=1}^{k} (x_{0j} - x_{ij}) r_i \right)^2 \right).$$

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 trend model.

4.1 A Two-Point Trend Model

Consider the two-point trend model $\mathcal{Y}_j(x) = \beta_0 + \beta_1 x + \mathsf{M}(x) + \varepsilon_j(x)$ with β_0 and β_1 unknown, so that $\mathsf{Y}(x_0) = \beta_0 + \beta_1 x_0 + \mathsf{M}(x_0)$ 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 $\mathsf{Y}(x_0)$ is

$$\widehat{\mathsf{Y}}(x_0) = \frac{\bar{\mathcal{Y}}(x_2)(x_0 - x_1) + \bar{\mathcal{Y}}(x_1)(x_2 - x_0)}{(x_2 - x_1)} \tag{11}$$

with MSE

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

where $a = x_2 - x_0$, $b = x_0 - x_1$, $a + b = x_2 - x_1$. Equation (12) implies that for this two-point trend model, when $x_0 \in (x_1, x_2)$, CRN increases MSE*; however, if we do extrapolation, i.e., $x_0 \notin (x_1, x_2)$, then CRN will decrease MSE*.

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

$$\widehat{\boldsymbol{\beta}} = \frac{1}{(x_2 - x_1)} \begin{pmatrix} x_2\bar{\mathcal{Y}}(x_1) - x_1\bar{\mathcal{Y}}(x_2) \\ \bar{\mathcal{Y}}(x_2) - \bar{\mathcal{Y}}(x_1) \end{pmatrix}. \tag{13}$$

It follows that

$$\text{Var}(\hat{\beta}_0) = \left(\tau^2 + \frac{\mathsf{V}}{n}\right) + \frac{2x_1x_2}{(x_2 - x_1)^2}\left[\tau^2(1 - r_{12}) + \frac{\mathsf{V}}{n}(1-\rho)\right] \tag{14}$$

$$\text{Var}(\hat{\beta}_1) = \frac{2\left[\tau^2(1 - r_{12}) + \frac{\mathsf{V}}{n}(1-\rho)\right]}{(x_2 - x_1)^2} \tag{15}$$

and

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

From Equation (15), we see that CRN reduces the variance of $\hat{\beta}_1$. Also notice that Equation (14) implies that if $x_1x_2 < 0$, so that 0 interior to the design space, then CRN inflates the variance of $\hat{\beta}_0$, while if $x_1x_2 > 0$, so that $\hat{\beta}_0$ is an extrapolated prediction of the response at $x = 0$, then CRN decreases the variance of $\hat{\beta}_0$.

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 $n \to \infty$ is

$$\mathrm{E}\left[\widehat{\nabla}_{\mathrm{sk}}(n) - \widehat{\nabla}_{\mathrm{sk}}(\infty)\right]^2 = \frac{2\mathsf{V}(1-\rho)}{n(x_1-x_2)^2}. \tag{16}$$

Equation (16) shows that CRN decreases the mean squared difference between these two estimators. It is worth noting that here 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 parameters. 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 $\mathcal{Y}_j(\mathbf{x}) = \beta_0 + \sum_{d=1}^{p} \beta_d x_d + \mathsf{M}(\mathbf{x}) + \varepsilon_j(\mathbf{x})$, where $p \geq 2$, $\mathbf{\Sigma}_{\mathsf{M}} = \tau^2 \mathbf{I}_k$ and $\mathbf{\Sigma}_{\mathsf{M}}(\mathbf{x}_0, \cdot) = \tau^2(r_0, r_0, \ldots, r_0)^\top$. 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. Suppose that we have a $k \times (p+1)$ *orthogonal* design matrix $\mathbf{D}$:

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

which means that the column vectors of $\mathbf{D}$ are pairwise orthogonal. Such an assumption on $\mathbf{D}$ makes the analysis tractable enough to give the following results:

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

$$\widehat{\boldsymbol{\beta}} = (\mathbf{D}^\top \mathbf{\Sigma}^{-1} \mathbf{D})^{-1} \mathbf{D}^\top \mathbf{\Sigma}^{-1} \bar{\mathcal{Y}} \tag{17}$$

where $\mathbf{\Sigma} = \mathbf{\Sigma}_{\mathsf{M}} + \mathbf{\Sigma}_{\varepsilon}$. More explicitly,

$$\hat{\beta}_0 = \frac{1}{k} \sum_{i=1}^{k} \bar{\mathcal{Y}}(\mathbf{x}_i) \tag{18}$$

and

$$\hat{\beta}_j = \frac{\sum_{i=1}^{k} x_{ij} \bar{\mathcal{Y}}(\mathbf{x}_i)}{\sum_{i=1}^{k} x_{ij}^2}\ , \quad j = 1, 2, \ldots, p. \tag{19}$$

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

$$\widehat{\mathsf{Y}}(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0)^\top \widehat{\boldsymbol{\beta}} \tag{20}$$

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

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

Notice that if

$$\frac{k-1}{k} > \sum_{j=1}^{p}\frac{x_{0j}^2}{\sum_{i=1}^k x_{ij}^2} \tag{21}$$

then CRN increases MSE*.

To help interpret this result, consider a $k = 2^p$ factorial design where the design points $x_{ij} \in \{-1, +1\}$. Then Equation (21) reduces to $\sum_{j=1}^p x_{0j}^2 < k-1$. Therefore, CRN will inflate the MSE* of $\widehat{\mathsf{Y}}(\mathbf{x}_0)$ 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 Cov$(\widehat{\boldsymbol{\beta}})$. Because of the orthogonality assumption, the expression for Cov$(\widehat{\boldsymbol{\beta}})$ becomes much simpler. It can be shown that

$$\text{Cov}(\widehat{\boldsymbol{\beta}}) = (\mathbf{D}^\top\boldsymbol{\Sigma}^{-1}\mathbf{D}\ )^{-1} = \begin{pmatrix} \frac{\frac{\mathsf{V}}{n}[1+(k-1)\rho]+\tau^2}{k} & 0 & \cdots & 0\\ 0 & \frac{\frac{\mathsf{V}}{n}(1-\rho)+\tau^2}{\sum_{i=1}^k x_{i1}^2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \frac{\frac{\mathsf{V}}{n}(1-\rho)+\tau^2}{\sum_{i=1}^k x_{ip}^2}\end{pmatrix} \tag{22}$$

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

Now let

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

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

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

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

$$\langle\widehat{\nabla}_{\text{sk}}(n)-\widehat{\nabla}_{\text{sk}}(\infty), \widehat{\nabla}_{\text{sk}}(n)-\widehat{\nabla}_{\text{sk}}(\infty)\rangle = \sum_{j=1}^{p}\text{E}\left[\left(\widehat{\nabla}_{\text{sk}_j}(n)-\widehat{\nabla}_{\text{sk}_j}(\infty)\right)^2\right]$$

$$= \frac{\mathsf{V}}{n}(1-\rho)\sum_{j=1}^{p}\left(\sum_{i=1}^{k}x_{ij}^2\right)^{-1}. \tag{23}$$

Equation (23) 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 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 metamodeling technique, such as stochastic kriging.

The models analyzed in this paper provide compelling evidence that CRN leads to less precise predictions, better gradient estimation, and better estimation of the "slope" terms in any trend model.

Throughout this paper we have assumed that the parameters of the random field M—specifically τ^2 and θ—were known, as well as the intrinsic variance V. In Chen et al. (2010) we assess empirically the impact of CRN when these parameters are estimated as well, which it turns out does not change the general conclusions above.

ACKNOWLEDGMENT

This article is based upon work supported by the National Science Foundation under Grant No. CMMI-0900354.

REFERENCES

Ankenman, B. E., B. L. Nelson and J. Staum. 2008. Stochastic kriging for simulation metamodeling. *2008 Winter Simulation Conference Proceedings*, 362–370.

Ankenman, B. E., B. L. Nelson and J. Staum. 2010. Stochastic kriging for simulation metamodeling. *Operations Research*, in press.

Chen, X., B. E. Ankenman and B. L. Nelson. 2010. The effects of common random numbers on stochastic kriging. Working Paper, Department of Industrial Engineering & Management Sciences, Northwestern University.

Donohue, J. M., R. C. Houck and R. H. Myers. 1992. Simulation designs for quadratic response surface models in the presence of model misspecification. *Management Science* 38:1765-1791.

Donohue, J. M., E. C. Houck and R. H. Myers. 1995. Simulation designs for the estimation of quadratic response surface gradients in the presence of model misspecification. *Management Science* 41:244-262.

Hussey, J. R., R. H. Myers, and E. C. Houck. 1987a. Pseudorandom number assignment in quadratic response surface designs. *IIE Transactions* 19:395-403.

Hussey, J. R., R. H. Myers, E. C. Houck. 1987b. Correlated simulation experiments in first-order response surface design. *Operations Research* 35: 744-758.

Kleijnen, J. P. C. 1988. Analyzing simulation experiments with common random numbers. *Management Science* 34:65-74.

Kleijnen, J. P. C. 1992. Regression metamodels for simulation with common random numbers: Comparison of validation tests and confidence intervals. *Management Science* 38:1164-1185.

Nozari, A., S. F. Arnold and C. D. Pegden. 1987. Statistical analysis for use with the Schruben and Margolin correlation induction strategy. *Operations Research* 35:127-139.

Santner, T. J., B. J. Williams and W. I. Notz. 2003. *The Design and Analysis of Computer Experiments*. Springer, NY.

Schruben, L. W., and B. H. Margolin. 1978. Pseudorandom number assignment in statistically designed simulation and distribution sampling experiments. *Journal of the American Statistical Association* 73:504-525.

Tew, J. D., and J. R. Wilson. 1992. Validation of simulation analysis methods for the Schruben-Margolin correlation-induction strategy. *Operations Research* 40:87-103.

Tew, J. D., and J. R. Wilson. 1994. Estimating simulation metamodels using combined correlation-based variance reduction techniques. *IIE Transactions* 26:2-16.

AUTHOR BIOGRAPHIES

XI CHEN is a Ph.D. student in the Department of Industrial Engineering & Management Sciences at Northwestern University. Her e-mail address is ⟨xichen2013@u.northwestern.edu⟩.

BRUCE ANKENMAN is an Associate Professor in the Department of Industrial Engineering & Management Sciences at Northwestern University. His research interests include the statistical design and analysis of experiments. Although much of his work has been concerned with physical experiments, recent research has focused on computer simulation experiments. Professor Ankenman is currently the director of the Masters of Engineering Management Program and the director of the Manufacturing and Design Engineering Program. He co-directs the freshman engineering and design course (EDC), and is the director of undergraduate programs for the Segal Design Institute. His e-mail and web addresses are ⟨ankenman@northwestern.edu⟩ and ⟨users.iems.northwestern.edu/~bea/⟩.

BARRY L. NELSON is the Charles Deering McCormick Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. His e-mail and web addresses are ⟨nelsonb@northwestern.edu⟩ and ⟨www.iems.northwestern.edu/~nelsonb/⟩.