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

THE INFLUENCE OF CORRELATION FUNCTIONS ON STOCHASTIC KRIGING METAMODELS

Wei Xie
Barry Nelson
Jeremy Staum

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

ABSTRACT

The correlation function plays a critical role in both kriging and stochastic kriging metamodels. This paper will compare various correlation functions in both spatial and frequency domains, and analyze the influence of the choice of correlation function on prediction accuracy by experimenting with three tractable examples with differentiable and non-differentiable response surfaces: the M/M/1 queue, multi-product M/G/1 queue and 3-station Jackson network. The twice or higher-order continuously differentiable correlation functions demonstrate a promising capability to fit both differentiable and non-differentiable multi-dimensional response surfaces.

1 INTRODUCTION

The kriging metamodeling method has found wide application in various fields. It is an interpolation approach used to predict the values at unobserved locations based on the observations from nearby locations. The prediction is a weighted average that gives greater weight to nearby observations.

Kriging was introduced by Sacks et al. (1989) to fit a global predictor to deterministic computer generated data. To account for the uncertainty in the unknown response surface, kriging treats the deterministic response as an observation of a random field. A kriging metamodel includes two parts, a regression model used to represent the large-scale trend, and a stationary random field used to represent the local behavior or deviation from the trend. Sacks et al. (1989) and Currin et al. (1991) demonstrated that kriging can be successfully used for the design and analysis of deterministic computer experiments.

The use of kriging in stochastic simulation was first mentioned by Mitchell and Morris (1992), with deeper study by van Beers and Kleijnen (2003). Differing from computer experiments with deterministic outputs, the outputs from stochastic simulation include sampling variability, and this variability often changes significantly across the design space. To meet the specific requirement of stochastic simulation, stochastic kriging (SK) was introduced by Ankenman et al. (). SK distinguishes the uncertainty about the response surface from the intrinsic variance inherent in stochastic simulation output. SK is a smoothing method whereas kriging is an interpolation method. Compared with the classical kriging method, SK provides a framework that can be used to effectively allocate computational effort, and it may provide better prediction of the response surface.

In both kriging and SK, choosing the correlation function of the random field plays a critical role. The choice of correlation function is usually based on considerations of the differentiability of the response surface because the sample paths of the random field inherit their differentiability from the correlation function (Santner et al. 2003). Further, in SK the extent of the smoothing due to intrinsic variance also depends on the type of correlation function used. In this paper, we will concentrate on comparing and analyzing the effect of commonly used correlation functions on SK in both the spatial and frequency domains.

In the next section we briefly describe the SK method, and demonstrate the importance of correlation function and intrinsic variance on the weight distribution for prediction in both the spatial and frequency

domains. The detailed analysis of various correlation functions is presented in Section 3. We briefly summarize the results of our empirical analysis in Section 4 before presenting details from three tractable examples—the M/M/1 queue, multi-product M/G/1 queue, and 3-station Jackson network—to illustrate the importance of choosing the correlation function in SK for fitting differentiable and non-differentiable response surfaces.

2 MODELING AND PREDICTION

Stochastic kriging treats the response surface as a realization of a random field Y with multivariate normal distribution. The stochastic simulation output on replication j at design point $\mathbf{x}$ is represented as

$$Y_j(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\boldsymbol{\beta} + M(\mathbf{x}) + \epsilon_j(\mathbf{x}). \tag{1}$$

The design variable $\mathbf{x} = (x_1, \ldots, x_d)$ is interpreted as a location in space. According to Equation (1), the variation of the response surface is divided into three uncorrelated parts: trend model, extrinsic and intrinsic uncertainties. The vector of basis functions in the trend model is represented by $\mathbf{f}(\mathbf{x})^T = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_q(\mathbf{x}))$. The trend model $\mathbf{f}(\mathbf{x})^T\boldsymbol{\beta}$ is used to capture the large-scale variation. The extrinsic uncertainty is modeled by a second-order stationary Gaussian random field M, which accounts for the local spatial dependence. The intrinsic uncertainty ϵ is the sampling variance inherent in the stochastic simulation. The sampling variances at different $\mathbf{x}$'s are independent if—as assumed here—common random numbers are not employed. The surface we try to predict is $\mathbf{f}(\mathbf{x})^T\boldsymbol{\beta} + M(\mathbf{x})$. In the following study, only constant trend β_0 is considered, so $\mathbf{f}(\mathbf{x})^T\boldsymbol{\beta} = \beta_0$.

To generate the global predictor, we choose k design points $(\mathbf{x}_1, \ldots, \mathbf{x}_k)$, then run simulations at each design point to obtain the observations. The number of replications at $\mathbf{x}_i$ is n_i, the sample mean of the simulation output at $\mathbf{x}_i$ is $\overline{Y}(\mathbf{x}_i) = n_i^{-1}\sum_{j=1}^{n_i} Y_j(\mathbf{x}_i)$, and the vector of sample means at all design points is $\overline{Y}^\top = (\overline{Y}(\mathbf{x}_1), \overline{Y}(\mathbf{x}_2), ..., \overline{Y}(\mathbf{x}_k))$. As the outputs from stochastic simulation include sampling variability, let $\sigma^2(\mathbf{x}) = \text{Var}[\epsilon(\mathbf{x})]$. The diagonal matrix $\mathbf{C}$ of intrinsic variance has ith element $\sigma^2(\mathbf{x}_i)/n_i$; kriging assumes $C = \mathbf{0}$, which is appropriate for deterministic computer experiments.

Let $\Sigma(\mathbf{x}, \mathbf{x}') = \text{Cov}(M(\mathbf{x}), M(\mathbf{x}'))$ be the covariance implied by the extrinsic spatial correlation model, let Σ represent the $k \times k$ covariance matrix across all design points and let $\Sigma(\mathbf{x}, \cdot)$ be the $1 \times k$ covariance vector between prediction point $\mathbf{x}$ and all design points $\mathbf{x}_1, \ldots, \mathbf{x}_k$. The optimal mean squared error (MSE) linear predictor is (Ankenman et al.)

$$\widehat{Y}(\mathbf{x}) = \beta_0 + \Sigma(\mathbf{x}, \cdot)[\Sigma + C]^{-1}(\overline{Y} - \beta_0 \cdot \mathbf{1}_{k\times 1}) \tag{2}$$

when β_0 and the correlation structure are known. In practice, we use maximum likelihood estimates (MLEs) of the parameters for prediction. A parametric form of the spatial covariance of $M(\mathbf{x})$ is often assumed with $\Sigma(\mathbf{x}, \mathbf{x}') = \tau^2 R(\mathbf{x} - \mathbf{x}'|\boldsymbol{\theta})$, where R represents the correlation function and τ^2 is the variance. For simplification, the product form of R is used for studying multi-dimensional problems with dimensionality d: $R(\mathbf{h}|\boldsymbol{\theta}) = \prod_{j=1}^{d} R(h_j|\theta_j)$.

Using SK involves the following main steps:

1. Choose the correlation function $R(\mathbf{x}|\boldsymbol{\theta})$, design points x_i and number of replications $n_i, i = 1, 2, \ldots, k$.
2. For each design point i, estimate $\sigma^2(\mathbf{x}_i)$ from n_i replications to obtain the intrinsic variance estimate $\hat{C}$. For kriging, $\hat{C} = \mathbf{0}$.
3. Calculate the MLEs $\widehat{\boldsymbol{\theta}}$, $\widehat{\beta}_0$, $\widehat{\tau}^2$ conditional on $\hat{C}$.
4. Make predictions using Equation (2).

Our SK software is available at www.stochastickriging.net.

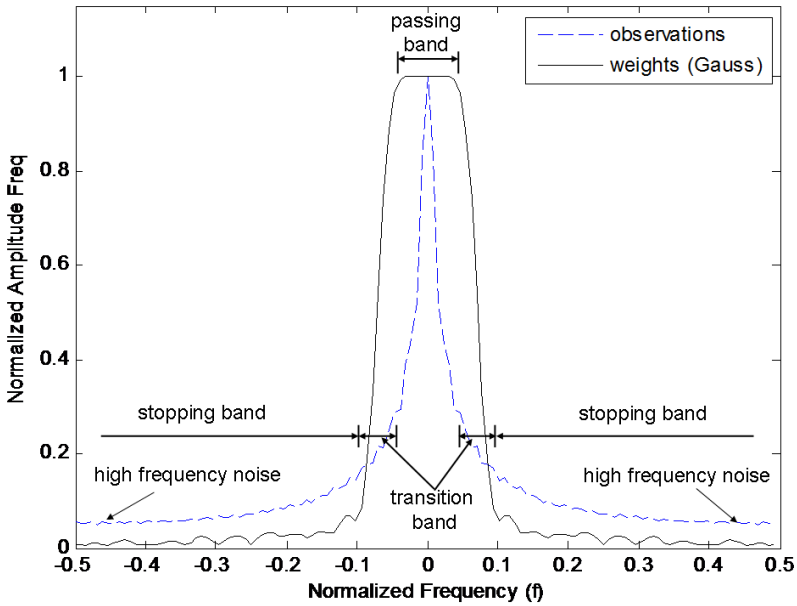

Figure 1: Frequency spectrums of weights and observations.

Viewed in the natural spatial domain of $\mathbf{x}$, the prediction at $\mathbf{x}$ is a weighted sum of observations. Equation (2) can be rewritten as

$$\widehat{Y}(\mathbf{x}) - \beta_0 = \sum_{i=1}^{k} W_i(\mathbf{x})(\overline{Y}(\mathbf{x}_i) - \beta_0) \tag{3}$$

where $W(\mathbf{x}) = \Sigma(\mathbf{x}, \cdot)[\Sigma + C]^{-1}$. By using the Fourier transformation (Oppenheim et al. 2005), this weighted sum in the spatial domain can also be represented as the pointwise product in the frequency domain. The corresponding relation in the frequency domain is

$$\tilde{Y}(f) = \tilde{X}(f)\tilde{W}(f) \tag{4}$$

where $\tilde{Y}(f)$ represents the frequency spectrum of prediction $\hat{Y}(\mathbf{x}) - \beta_0$, while $\tilde{X}(f)$ and $\tilde{W}(f)$ represent the frequency spectrums of observations $\overline{Y} - \beta_0 \cdot \mathbf{1}_{k\times 1}$ and the weight distribution $W(\mathbf{x})$, respectively.

The weight W is also called a filter. Its frequency spectrum $\tilde{W}(f)$ determines which frequency band information in the observations $\tilde{X}(f)$ can be passed to the prediction. The true response surfaces we consider are often differentiable, or at least change slowly, which is referred to as low-frequency information. High frequency noise is random and unstructured. The existence of sampling variance makes the observations differ from the true surface. Thus, the observations include both the low-frequency information from the true response surface and the high-frequency noise generated by the sampling variance. Therefore, we prefer the frequency spectrum of the weights to be a low-pass filter with a narrow transition from passing band to stopping band as shown in Figure 1. This kind of filter can effectively pass the low-frequency information from observations to the prediction and stop the high-frequency noise generated by sampling variance.

This frequency-domain analysis can be made rigorous when the design points are evenly distributed in the design space, but provides useful insight more generally. The maximum frequency information that can be extracted from the response surface is $f_s/2$ (Oppenheim et al. 2005) where the spatial sampling frequency f_s is proportional to the number of design points. In Figure 1, the horizontal frequency axis is normalized by f_s. The frequency around ± 0.5 corresponds to the high-frequency band, and the frequency information around the origin corresponds to the low-frequency band. Furthermore,

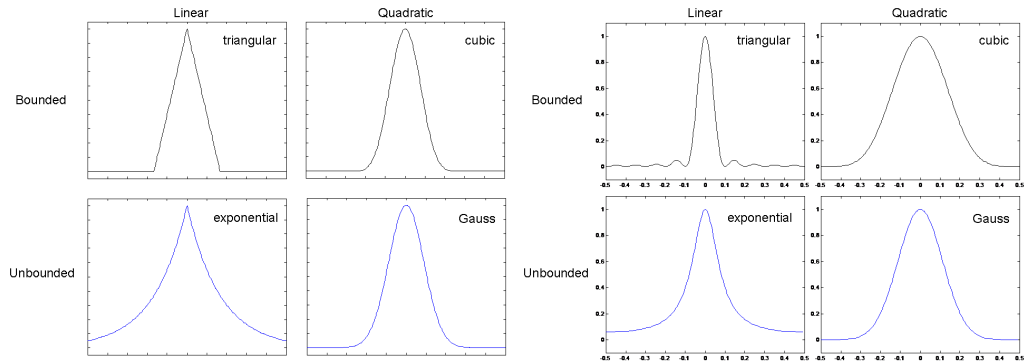

Figure 2: Correlation functions (left) and associated power spectrums (right).

as the number of design points increases and the frequency range becomes larger, the normalized frequency spectrum of both the data and the weights will shrink toward the origin.

3 CORRELATION FUNCTION ANALYSIS

In this section we will concentrate on analyzing various commonly used correlation functions including triangular, exponential, Gaussian, cubic and Matérn. For simplification, we only consider one-dimensional x in this section.

For the stationary stochastic process M, the correlation function $R(h|\theta) = \text{Cov}[M(x), M(x+h)]/\tau^2$ measures the dependence between $M(x)$ and $M(x+h)$. In the spatial domain, the correlation function provides information about the spatial dependence and differentiability of sample paths of the random field (Santner et al. 2003), which can be observed by rewriting Equation (2) as

$$\widehat{Y}(x) = \beta_0 + \sum_{i=1}^{k} \mathbf{b}_i R(x - x_i|\theta) \tag{5}$$

where $\mathbf{b} = \tau^2[\Sigma + C]^{-1}(\overline{Y} - \beta_0 \cdot \mathbf{1}_{k\times 1})$. Only the correlation function depends on x, and thus the differentiability of the metamodel is inherited from the correlation function. As the prediction point moves closer to any design point x_i, the behavior of $\hat{Y}(x)$ depends on the shape of the correlation function at the origin, especially the order of differentiability or the slope at the origin. As noted by Santner et al. (2003), the properties of the correlation function around the origin also determine the differentiability of the sample paths of the stationary process M. If the $2k$th-order partial derivative of the correlation function exists and is finite at the origin, then the kth-order partial derivative of the sample path exists.

Based on the behavior around the origin, the correlation functions can be divided into two groups as shown in Figure 2. The triangular and exponential correlation functions with the linear behavior $1-R(h|\theta) = O(h)$ are not differentiable at the origin. Thus, the sample paths generated by the triangular and exponential functions are continuous but not differentiable. The cubic and Gaussian functions with quadratic behavior $1 - R(h|\theta) = O(h^2)$ are twice or higher-order continuously differentiable at the origin, so the sample paths they generate are continuously differentiable. The shape of the Matérn correlation function changes when the value of its smoothness parameter ν changes. When ν is large, the shape of the Matérn correlation function is close to the Gaussian function; when ν is near 1, it is close to the exponential function.

The correlation functions can also be divided into two groups based on their support. The triangular and cubic correlation functions have bounded support and the range depends on the value of θ. Points located outside this range have no influence on the prediction. Thus, the covariance matrix Σ generated by the triangular or cubic function is sparse. The exponential and Gaussian functions have unbounded support; thus, the covariance matrix they generate is dense, increasing the computational complexity of matrix inversion relative to the bounded correlation functions.

The triangular correlation function,

$$R(h|\theta) = \begin{cases} 1-\theta|h|, & |h| < 1/|\theta| \\ 0, & \text{otherwise} \end{cases}$$

is unrealistic for many physical processes, according to Stein (1999), because its power spectrum does not monotonically decrease as the frequency increases (see the right side of Figure 2). Thus, the triangular correlation function is not included in our empirical study.

The power spectrum of the exponential correlation function, $R(h|\theta) = e^{-\theta|h|}$, decays quickly near the origin but is still non-zero even at high frequencies (± 0.5), which explains the rough (bumpy) metamodels generated by the exponential correlation function.

The cubic correlation function,

$$R(h|\theta) = \begin{cases} 1-6\left(\frac{h}{\theta}\right)^2 + 6\left(\frac{|h|}{\theta}\right)^3, & |h| \le \frac{\theta}{2} \\ 2\left(1-\frac{|h|}{\theta}\right)^3, & \frac{\theta}{2} < |h| \le \theta \\ 0, & \theta < |h| \end{cases}$$

is twice continuously differentiable with bounded support. Its power spectrum has slow decay rate around the origin, leading to a smoother metamodel.

The Gaussian correlation function, $R(h|\theta) = e^{-\theta h^2}$, is infinitely differentiable with light tails. Its power spectrum has slow decay rate around the origin.

The Matérn correlation function

$$R(h|\theta,\nu) = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{2\sqrt{\nu}|h|}{\theta}\right)^{\nu} K_\nu\left(\frac{2\sqrt{\nu}|h|}{\theta}\right)$$

where K_ν is the modified Bessel function of order ν, includes both scale parameter θ and smoothness parameter ν. The degree of smoothness can be estimated, so the Matérn correlation function can provide more flexibility to fit the data. However, estimating more parameters requires more computation.

4 EMPIRICAL EVALUATION

We summarize the results obtained by studying three tractable test examples, the M/M/1 queue, multi-product M/G/1 queue and 3-station Jackson network, then present details. SK with the Gaussian correlation function is the best performer once the number of design points and the number of replications are large enough. This is even true for the 3-station Jackson network which has non-differentiable points on the response surface. When the number of design points is too small for good prediction, then the exponential correlation function may do better than Gaussian. We find Gaussian best, cubic next and exponential worst in these examples. For the Matérn correlation function, when the number of design points is relatively small its performance is close to the exponential function; when the number of design points is large enough its performance is close to the Gaussian function.

4.1 M/M/1 Queue

In this example the response is the expected waiting time prior to service for an M/M/1 queue as a function of the utilization x. Let λ be the arrival rate and μ the service rate. Then, $x = \lambda/\mu$, and the response surface is $y(x) = x/(\mu(1-x))$. We will set $\lambda = 1$ which means $x = 1/\mu$.

Our experiment design uniformly fills the design space of x, $[0.5, 0.9]$, and fixes the number of customers simulated per replication at all design points to be equal. Whitt (1989) showed that the asymptotic variance of the mean of ℓ successive customer waiting times in an M/M/1 queue is $\sigma_W^2 \equiv \lim_{\ell\to\infty} \ell \text{Var}(\overline{W}) \approx 4x/(1-x)^4$ for x close to 1. To obtain approximately constant intrinsic variance over the whole design space, we make the number of replications at design point x_i proportional to the asymptotic standard deviation, $n_i \propto \sqrt{4x_i/(1-x_i)^4}$.

The total computational budget is $B = \sum_{i=1}^{k} n_i$; let $\overline{n} = B/k$ denote the average number of replications per design point. For a fixed computational budget, we systematically change the number

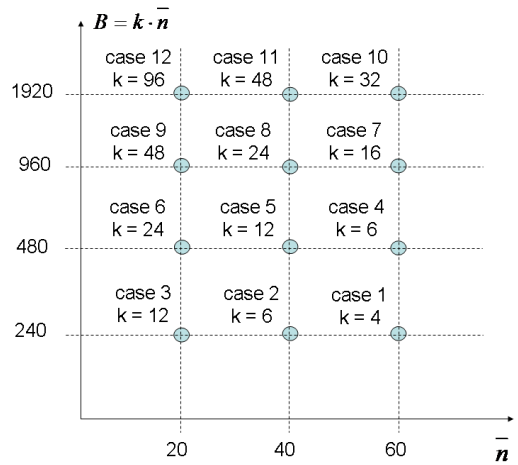

Figure 3: M/M/1 experiments.

of design points and the number of replications at each design point to examine the influence of experimental design, correlation function and intrinsic variance on the quality of the metamodel. The number of prediction points is $K = 1000$, the number of customers per replication is 1000, and the number of macro-replications is 300 in these experiments. We initialize M/M/1 queueing models in steady state.

To compare different experiment designs and the effect of various correlation functions, the root average relative MSE (RARMSE) is used as the criterion to summarize the quality of the metamodel; we estimate it by

$$\text{RARMSE} = \sqrt{\frac{1}{K}\sum_{j=1}^{K}\frac{\text{bias}_j^2 + S_j^2}{y(x_j)}} \tag{6}$$

where $y(x_j)$ is the true response value, $\bar{\hat{Y}}(x_j)$ and S_j^2 are the sample mean and sample variance of predictions over macro-replications at prediction point x_j, and $\text{bias}_j = \bar{\hat{Y}}(x_j) - y(x_j)$.

4.1.1 Results

Table 1 shows the RARMSE for the cases in Figure 3 using various correlation functions. In the table, 'K-exp' represents the fitting quality of kriging with the exponential correlation function. We include kriging to assess the importance of modeling intrinsic variance in SK. Since kriging using Gaussian, cubic and Matérn correlation functions encounters singularity problems when the number of design points increases, we only report kriging results from the exponential correlation function.

Each row in Table 1 gives the results with fixed total computational budget B, and each column gives the results with fixed average number of replications, $\bar{n}$. For the fixed $\bar{n}$, when we increase the number of design points and computational budget, the average relative root MSE decreases as shown in Table 1. This means that when the average number of replications is fixed, increasing the number of design points can extract more information and prediction accuracy is improved.

Kriging results are always at least a little bit worse than results using SK. For SK, when the number of design points is large enough, the behaviors of Gaussian and Matérn correlation functions are the best, and they are very close to each other. The cubic function is next. The exponential function is the worst of the four correlation functions. As anticipated, the Matérn correlation function demonstrates flexibility, with performance close to the exponential correlation function when the number of design points is small, and close to the Gaussian correlation function when the number of design points is larger.

Table 1: Root average relative MSE for M/M/1 experiments.

Budget B=240	$k=4$ $\overline{n}=60$	$k=6$ $\overline{n}=40$	$k=12$ $\overline{n}=20$
K-exp	1.879	0.152	0.070
SK-exp	1.851	0.146	0.068
SK-gau	0.438	0.162	0.070
SK-cub	0.499	0.175	0.079
SK-Matern	0.605	0.141	0.070
Budget B=480	$k=6$ $\overline{n}=60$	$k=12$ $\overline{n}=40$	$k=24$ $\overline{n}=20$
K-exp	0.146	0.055	0.060
SK-exp	0.144	0.053	0.058
SK-gau	0.161	0.054	0.051
SK-cub	0.174	0.066	0.059
SK-Matern	0.141	0.054	0.051
Budget B=960	$k=16$ $\overline{n}=60$	$k=24$ $\overline{n}=40$	$k=48$ $\overline{n}=20$
K-exp	0.039	0.044	0.058
SK-exp	0.038	0.043	0.054
SK-gau	0.037	0.037	0.042
SK-cub	0.045	0.046	0.049
SK-Matern	0.037	0.038	0.043
Budget B=1920	$k=32$ $\overline{n}=60$	$k=48$ $\overline{n}=40$	$k=96$ $\overline{n}=20$
K-exp	0.036	0.042	0.058
SK-exp	0.035	0.040	0.049
SK-gau	0.029	0.030	0.036
SK-cub	0.036	0.037	0.041
SK-Matern	0.029	0.030	0.037

4.1.2 Weight Distribution and Reversion Analysis

This section aims to explain the observed performance of different correlation functions.

Recall that the weights on observations at design points are $W(x) = \hat{\Sigma}(x,\cdot)[\hat{\Sigma}+\hat{C}]^{-1}$ for SK and $W(x) = \widehat{\Sigma}(x,\cdot)\widehat{\Sigma}^{-1}$ for kriging. For kriging, when the prediction point is a design point, the weight on the response at that design point is one and the weights on the other design points are zero. When the prediction point moves away from a design point, a higher percentage of total prediction weight is assigned to the overall mean $\widehat{\beta}_0$. In SK the weights do not always sum to one even when the prediction point is a design point, but the same basic shift in weight toward the overall mean when moving away from a design point still occurs.

We numerically analyzed the influence of correlation structure on the weight distribution for Case 11. Figure 4 (left) shows the weight distributions $W(x)$ for various correlation functions for the prediction point $x = 1.550$. For kriging using the exponential correlation function, almost all of the weight is distributed to the neighboring two design points. This result agrees well with the screening effect mentioned by Cressie (1993). As we account for intrinsic variance by using SK, the screening effect becomes weaker and more weight is assigned to distant responses. For SK with the exponential function, we observe that the weight is slightly more spread out, and that all weights are positive and their sum is close to one. Using Gaussian, cubic and Matérn correlation functions even more distant design points have influence on the prediction and negative weights can be observed.

If the weights sum to less than one, then it means that a portion of the total prediction weight is assigned to the mean $\hat{\beta}_0$. Thus, the residual weight $W_r(x) = 1 - \sum_{i=1}^{k} W_i(x)$ measures reversion to the mean, and therefore also how much information is available from neighboring design points. Figure 4 (right) shows W_r as a function of the service rate $1/x$. The interesting behavior is on the left, which corresponds to heavy load on the queue and high intrinsic variance: the reversion is most substantial for the Gaussian, cubic and Matérn correlation functions, which leads to prediction bias.

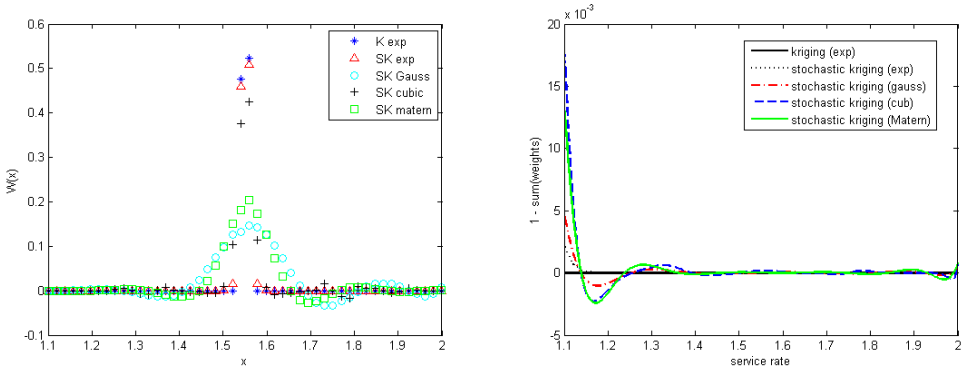

Figure 4: Left: For Case 11 ($k = 48$, $\overline{n} = 40$), weight distribution $W(x)$ for prediction point $x = 1.550$ (which is between two design points). Right: Residual sum of weights over design points.

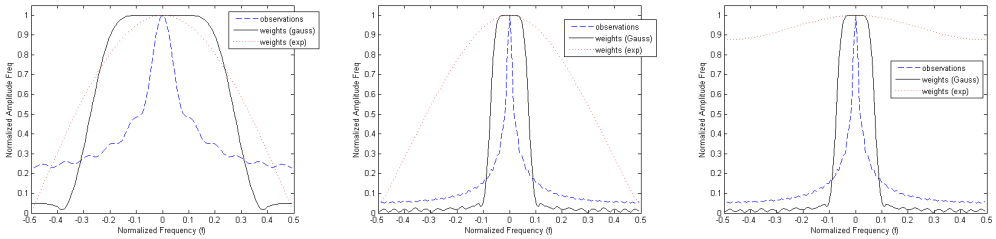

Figure 5: Frequency spectrums of observations and weight distributions at $x = 1.550$. from exponential and Gaussian correlation functions for Case 5 (left; $k = 12$, $\overline{n} = 40$) and Case 11 (center; $k = 48$, $\overline{n} = 40$) when the prediction point is between two design points. Frequency spectrums of observations and weight distributions at $x = 1.559$ from exponential and Gaussian correlation functions for Case 11 (right; $k = 48$, $\overline{n} = 40$) when the prediction point a design point.

Next we will analyze the weight distribution in the frequency domain. The weights can be considered as a filter applied to observations as shown in Equation (4). To avoid the boundary effect, we consider a prediction point that is near the middle of the design space. By applying the Fourier transform to the average weights over macro-replications, the frequency spectrum of expected weights can be estimated. The weight distribution changes when the location of the prediction point relative to the design points varies. Two locations will be considered: The prediction point is centered between two design points, and the prediction point is a design point.

When the prediction point is located between design points, the frequency spectrums of observations $\tilde{X}(f)$ and the weights $\tilde{W}(f)$ are as shown in Figure 5 (left and center) for cases with $k = 12$ and $k = 48$, respectively. The horizontal axis is the normalized frequency related to the sampling frequency. Notice the sampling frequencies of these cases are different. The vertical axis is the normalized frequency spectrum amplitude. The dashed line represents the frequency spectrum of observations. As the number of design points increases, the sampling frequency also increases. Thus, compared to Figure 5 (left), the frequency spectrum in Figure 5 (center) is observed to proportionally shrink toward the origin. With more design points, a wider band of information can be extracted. In addition, as the sampling frequency increases, the spectrum amplitude gradually decreases. The heavy tails in the frequency spectrum of the observations mainly come from the high-frequency noise generated by the intrinsic variance.

From the figures, we observe that the weights generated by the Gaussian correlation function are similar to a good low-pass filter with a narrow transition band. Thus, the Gaussian correlation function can effectively pass the low-frequency information from the observations to the predictions and stop the high-frequency noise generated by intrinsic variance.

Compared to the results obtained by the Gaussian correlation function, the frequency spectrum of the weights generated by the exponential correlation function has a wide transition band. Therefore, it can not as effectively filter out the high-frequency noise in the observations. However, if the number of design points is small, which means the sampling frequency is small, then the observations do not provide enough resolution for the Gaussian correlation function to separate the low frequency response surface signal from the high frequency stochastic noise. To capture the response surface it is best to attribute more of the observation information to the surface and less to the noise. The exponential correlation function, which passes a wider band of information, does this better than the Gaussian correlation function. This explains why SK with the exponential correlation function performs better then the Gaussian correlation function when the number of design points is small in Table 1. However, when the number of design points is large enough, the Gaussian correlation function can effectively pass the low-frequency information of interest and stop the high-frequency noise, so it delivers better predictions than the exponential correlation function.

As the position of the prediction point relative to the design points varies, the weight distribution also changes. When the prediction point moves closer to a design point, the frequency spectrum of weights from the exponential function passes more high-frequency information contained in the observations (compare Figures 5 center and right). This agrees well with the conclusion obtained in the spatial domain. When using the Gaussian correlation function, the frequency spectrum of the weights does not change much as the prediction point moves closer to design points.

4.2 Multi-product M/G/1 Queue

The M/M/1 queue example has a one-dimensional differentiable response surface. To test SK on multi-dimensional differentiable response surfaces, we will analyze the results for a multi-product M/G/1 queue.

In the multi-product M/G/1 queue the arrivals include different types of products which have different service-time requirements. The interarrival time and service time for each product have exponential distributions. In this example, the response is the expected waiting time prior to service as a function of system utilization ρ and product mix. Let λ be the total arrival rate. There are d types of products, and the fraction of product i is α_i. Thus, the design variable is $\mathbf{x} = (\rho, \alpha_1, \alpha_2, \ldots, \alpha_d)$. The mean service time for product i is set as $\tau_i = i$. Let $\overline{\tau} \equiv \sum_{i=1}^{k} \alpha_i \tau_i$ and $\tilde{\tau}^2 \equiv \sum_{i=1}^{k} \alpha_i \tau_i^2$. The steady state expected waiting time is

$$E(W_Q) \quad = \quad \frac{\lambda \sum_{i=1}^{d} \alpha_i \tau_i^2}{1 - \lambda \sum_{i=1}^{d} \alpha_i \tau_i} = \frac{\rho \tilde{\tau}^2}{(1-\rho)\overline{\tau}}.$$

The response surface for the multi-product M/G/1 queue is multi-dimensional and differentiable. The dimensionality can easily be controlled by changing the number of products. Notice dramatically different service times for different kinds of products give rise to much higher variability across the response surface compared with the M/M/1 queue.

4.2.1 Experiment Design

The uniform mixture design (Borkowski and Piepel 2009) is used to generate design points which are expected to uniformly cover the experimental region. To generate the prediction points, we first generate random uniform distributed number theoretic (NT) net points in a hypercube. Then, a transformation (Borkowski and Piepel 2009) is applied to map each NT net point into the design space to obtain prediction points.

When we simulate the multi-product M/G/1 queue, we fix the number of replications at each design point and use the formula obtained by Whitt (1989) based on a heavy-traffic approximation to plan the simulation run length to reach constant relative precision at all design points.

Here, we analyze the example with three types of products. We found similar behavior for examples with up to eight products. The range of utilization is ρ from 0.5 to 0.9; the number of design points is $k = 301$; the number of prediction points is $K = 2000$; and the number of replications at each design point is $n = 10$. We start the multi-product M/G/1 queueing model in steady state.

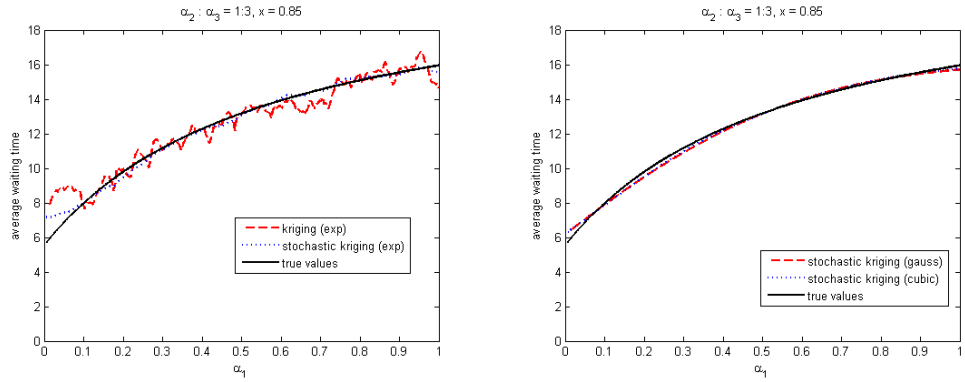

Figure 6: Metamodels for the expected waiting time in multi-product M/G/1 queue.

4.2.2 Results

One slice through the response surfaces can be obtained by fixing $\alpha_2 : \alpha_3 = 1 : 3$ and $x = 0.85$. Metamodels generated from the same output using exponential, cubic and Gaussian correlation functions are shown in Figure 6. In both panels, the horizontal axis represents the percentage of product 1, the vertical axis represents the average waiting time in the M/G/1 queue, and the solid line is the true value.

We first discuss the left panel, which shows the kriging and SK metamodels generated with the exponential correlation function. Notice that both wiggle and look noisy, which is explained by the quick decay near the origin and heavy tails in the power spectrum of the exponential correlation function (Figure 2). However, there are more obvious oscillations in the metamodel from kriging than from SK, which is because kriging does not smooth based on the intrinsic variance as SK does.

Looking at the right panel, metamodels obtained with the Gaussian and cubic correlation functions are very close to each other. Compared with results using the exponential function, the metamodels are much smoother and more accurate.

4.3 3-Station Jackson Network

For the differentiable response surfaces from the M/M/1 queue and multi-product M/G/1 queue, the Gaussian correlation function demonstrates better performance. Can it also provide a high quality metamodel for a non-differentiable response surface? To answer that question, a 3-station Jackson network with stochastic routing will be examined.

In the 3-station Jackson network both interarrival time and service time have exponential distributions. Let λ represent the total arrival rate to the system, and α_i represent the fraction of product i, $i = 1, 2, 3$. Suppose all products require the same service rate at station j represented by μ_j. Let ρ be the utilization of the bottleneck (BN) station, that is the station that has the maximum utilization. The design variable is $\mathbf{x} = (\rho,\alpha_1,\alpha_2,...,\alpha_d)$, and the response variable is the expected cycle time (CT) of different products. As the product combination changes, the BN may shift from one station to another station, which yields a non-differentiable response surface. Thus, the 3-station Jackson network provides a good test example of a non-differentiable and multi-dimensional response surface.

Let the matrix $[\delta_{ij}]$ represent the expected number of visits to station j by product i. The expected CT for product i can be calculated as (Yang et al. 2007)

$$E(CT_i) = \sum_{j=1}^{3} \frac{\delta_{ij}}{\mu_j - \lambda \sum_{m=1}^{d} \alpha_m \delta_{mj}} = \sum_{j=1}^{3} \frac{\delta_{ij}}{\mu_j \left[1 - \rho \frac{\sum_{m=1}^{d} \alpha_m \delta_{mj}/\mu_j}{\max_h(\sum_{m=1}^{d} \alpha_m \delta_{mh}/\mu_h)}\right]}.$$

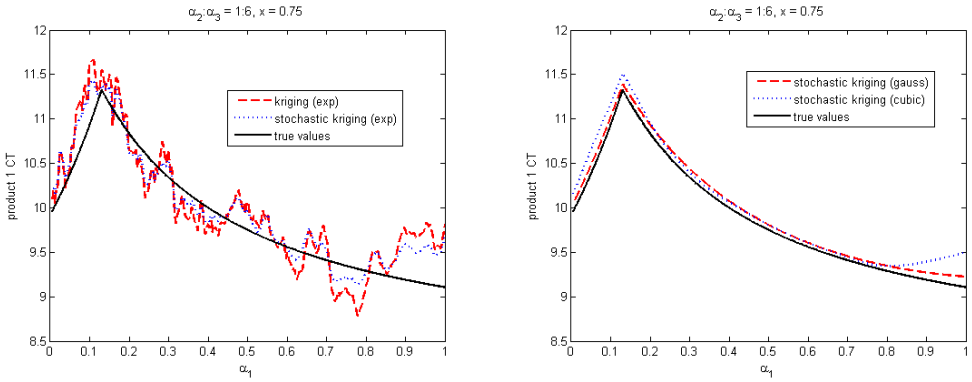

Figure 7: Metamodels for the expected cycle time for 3-station Jackson network.

4.3.1 Experiment Design

We follow a similar approach to that used in the multi-product M/G/1 queue to generate design and prediction points. The setting of this experiment is as follows: the range for utilization of the BN station is ρ from 0.6 to 0.8, the number of design points is $k = 201$, the number of prediction points is $K = 1000$; and the number of replications at each design point is $n = 10$. We start the 3-station Jackson network empty and delete the initial transit. The service rate at station j is $[\mu_j] = [1, 1.667, 0.667]$. The expected number of visits to station j by product i is

$$[\delta_{ij}] = \begin{bmatrix} 1.05 & 0.53 & 1.13 \\ 1 & 1 & 1 \\ 1.11 & 0.11 & 0.61 \end{bmatrix}.$$

To plan the simulation run length at each design point, we first pick the product that has the highest cycle time variance, which is product 1. For simplification, we ignore the positive correlation between different stations. Then, based on the selected product, we plan the run length to reach similar relative precision across the design space using the approach in Whitt (1989).

4.3.2 Results

To make it easy to demonstrate the results, Figure 7 displays one slice of the metamodel from each correlation function based on the same simulation output, fixing $\alpha_2 : \alpha_3 = 1 : 6$ and setting $\rho = 0.75$. In both panels, the horizontal axis is the percentage of product 1, and the vertical axis is the expected cycle time of product 1. There is obvious noise in both metamodels using the exponential correlation function, although the extent of the noise in the metamodel from SK is a little bit smaller than that from kriging. The difference between metamodels from kriging and SK using the same correlation function is due to SK accounting for intrinsic variance. Comparing SK with Gaussian and cubic correlation functions, the Gaussian function provides an obviously better metamodel than the cubic function, and both are substantially better than the exponential.

Figure 7 demonstrates the metamodel quality on one slice of the response surface. The prediction quality using kriging and SK with various correlation functions over the entire space is in Table 2. We observe that SK provides better predictions than kriging. Comparing results obtained from different correlation functions, the Gaussian function performs best, and it is followed by the cubic function.

ACKNOWLEDGMENT

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

Table 2: Root average relative squared error (RARSE) for 3-station Jackson network.

RARSE	product 1	product 2	product 3
K-exp	0.039	0.038	0.041
SK-exp	0.032	0.031	0.030
SK-gau	0.007	0.006	0.007
SK-cub	0.020	0.016	0.027

REFERENCES

Ankenman, B., B. L. Nelson, and J. Staum. Stochastic kriging for simulation metamodeling. *Operations Research* 58 (2): 371–382.

Borkowski, J. J., and G. F. Piepel. 2009. Uniform designs for highly constrained mixture experiments. *Journal of Quality Technology* 41 (1): 35–47.

Cressie, N. A. C. 1993. *Statistics for spatial data*. New York: John Wiley & Sons.

Currin, C., T. Mitchell, M. Morris, and D. Ylvisaker. 1991. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association* 86 (416): 953–963.

Mitchell, T. J., and M. D. Morris. 1992. The spatial correlation function approach to response surface estimation. In *Proceedings of the 1992 Winter Simulation Conference*, ed. R. C. Crain, J. R. Wilson, J. J. Swain, and D. Goldsman, 565–571: Piscataway, N.J.: Institute of Electrical and Electronics Engineers.

Oppenheim, A. V., R. W. Schafer, and J. R. Buck. 2005. *Discrete-time signal processing*. Upper Saddle River, NJ: Pearson Education.

Sacks, J., S. B. Schiller, and W. J. Welch. 1989. Designs for computer experiments. *Technometrics* 31 (1): 41–47.

Santner, T. J., B. J. Williams, and W. I. Notz. 2003. *The design and analysis of computer experiments*. New York: Springer.

Stein, M. L. 1999. *Interpolation of spatial data: Some theory for kriging*. New York: Springer-Verlag.

van Beers, W. C. M., and J. Kleijnen. 2003. Kriging for interpolation in random simulation. *Journal of the Operational Research Society* 54:255–262.

Whitt, W. 1989. Planning queueing simulations. *Management Science* 35 (11): 1341–1366.

Yang, F., J. Liu, M. Tongarlak, B. E. Ankenman, and B. L. Nelson. 2007. Metamodeling for cycle time-throughput-product mix surfaces using progressive model fitting. In *Proceedings of the 2007 Winter Simulation Conference*, ed. S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 322–330: Piscataway, N.J.: Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHIES

WEI XIE is a Ph.D. student in the Department of Industrial Engineering and Management Sciences at Northwestern University. Her research interests are in simulation metamodeling and input uncertainty. Her e-mail address is ⟨WeiXie2013@u.northwestern.edu⟩.

BARRY L. NELSON is the Charles Deering McCormick Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University, and a Fellow of INFORMS. 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/⟩.

JEREMY STAUM is Associate Professor of Industrial Engineering and Management Sciences and holds the Pentair- Nugent Chair at Northwestern University. He coordinated the Risk Analysis track of the 2007 Winter Simulation Conference and serves as department editor for financial engineering at *IIE Transactions* and as an associate editor at *Operations Research*. His website is ⟨users.iems.northwestern.edu/~staum⟩.