Kriging metamodeling in multiple-objective simulation optimization

Mehdi Zakerifar, William E Biles and Gerald W Evans

Abstract
This paper describes the application of Kriging metamodeling in multiple-objective simulation optimization. An Arena-based simulation model of an \((s, S)\) inventory system is utilized to demonstrate the capabilities of Kriging metamodeling as a simulation tool. Response surface methodology and Kriging metamodeling are compared to determine the situations in which one approach might be preferred over the other. The optimization approaches described here have the objective of finding the optimal values of reorder point \(s\) and maximum inventory level \(S\) so as to minimize the total cost of the inventory system while maximizing customer satisfaction. This paper describes two alternative approaches to utilizing Kriging methodology with multiple-objective optimization in simulation studies.

Keywords
kriging metamodeling, simulation optimization, multi-objective optimization

I. An overview of Kriging metamodeling
A metamodel is an approximation of an input/output (I/O) function that is defined by an underlying simulation model (see Kleijnen1). The metamodel is a surrogate for the real-world system that is used for experimentation and analysis; that is, experimentation with the actual system is far too costly and time consuming, so computer-based experimentation, or simulation, is preferred (see Van Beers2). Most metamodeling studies focus on low-order polynomial regression using factorial-based experimental designs. For the first-order regression model with \(n\) design variables \(x_i, i = 1, \ldots, n\), the form of the regression metamodel is

\[
E(y) = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n
\]

while the second-order model is given by

\[
E(y) = \beta_0 + \sum_i \beta_i x_i + \sum_{i,j} \beta_{ij} x_i x_j
\]

The classic design used to estimate a first-order metamodel (1) is the factorial design or, in the case where \(n\) is large (say, \(n > 5\)), the fractional factorial design. For the second-order metamodel (2), the central-composite design (CCD) is the experimental design most often used.

Kleijnen3 has provided an excellent treatise on the applications of Kriging metamodeling in simulation. This paper cites more than 50 references to scholarly efforts in this domain. Kleijnen and his colleagues have also presented the most comprehensive treatment to date of the constrained optimization approach to Kriging in simulation.4 In particular, these authors describe a heuristic procedure for the solution to this problem based on Latin Hypercube Designs (LHDs),5 which was defined by McKay et al.,6 and Kriging7 to the \((s, S)\) inventory model. This recent development goes well beyond that presented by Biles et al.8 for the application of the constrained optimization approach to simulation. The formulation of that problem can be written as

\[
\text{max } \min_{x_i} f_0(x), \quad x_i, \quad i = 1, \ldots, n
\]
subject to constraints $a \leq x \leq c$, and $f_j(x) \leq d_j$, $j = 1, \ldots, m$.

Note that the notation in (3) reserves the definition of the vector $b$ as the vector of the regression coefficients suggested, but not explicitly stated, in (1) and (2).

Sometimes the decision maker is unable to identify a single objective function $f_0(x)$ that fully explains his (her) decision framework, but is confronted with several objectives $f_j(x)$, $j = 1, \ldots, m$.

The multiple-objective optimization problem can be written as follows:

\[
\max \ (\min) G(f(x))
\]

subject to $a \leq x \leq c$, where $x$ is the $n$-vector of input factors $x_i$, $i = 1, \ldots, n$, $a$ and $c$ are lower bounds and upper bounds, respectively, on $x$, and the function $G(f(x))$ is a policy for prioritizing the several objective functions $f_j(x)$, $j = 1, \ldots, m$.

Hawe and Sykulski\(^7\) discuss several algorithms by which to apply Kriging to the solution of the so-called MOOP (Multiple Objective Optimization Problem), shown in (4), but the one evaluated here is the approach of transforming the MOOP formulation (4) to a sequence of SOOP (Single Objective Optimization Problem) formulations (3). That is, alternately solve the constrained optimization problem (5) until an optimal solution is found:

\[
\max \ (\min) \ f_j(x) \text{ for some } j
\]

subject to constraints $f_k(x) \{\leq, =, \geq\} d_k$, $k = 1, \ldots, m$ for $k \neq j$ and $a \leq x \leq c$.

Another approach to solving the MOOP is to weight each of the $m$ objective functions $f_j(x)$, $j = 1, \ldots, m$ with a weight $w_j$, $j = 1, \ldots, m$ to form the weighted objective function:

\[
\max \ (\min) \ \sum w_j f_j(x) \text{ such that } 0 \leq w_j \leq 1 \text{ and } \sum w_j = 1
\]

For an objective $f_j(x)$ that is to be maximized, the coefficient of $w_j$ is 1, whereas for those objectives being minimized the coefficient of $w_j$ is –1. The constraints $a \leq x \leq c$ also apply. We shall refer to this formulation as the WOOP (Weighted Objective Optimization Problem).

This paper addresses the multiple-objective decision framework and presents an extension of the constrained optimization framework discussed by Kleijnen et al.\(^4\) and Biles et al.\(^8\) The proposed approach will utilize LHDs together with a Kriging approach in the manner described by Biles et al.\(^8\) for the constrained optimization approach applied to the MOOP and WOOP formulations shown above.

Kriging is an interpolation method that predicts unknown values of a random function. The simplest Kriging model, called Ordinary Kriging, is shown in equation (7):

\[
f(x) = \beta + z(x)
\]

where $\hat{y}(x)$ denotes the Kriging predictor for the input or factor combination (point) $x$, $\beta$ is a constant, and $z(x)$ is a covariance-stationary process; see Cressie\(^7\) and Wackernagel.\(^10\)

An Ordinary Kriging prediction is a weighted linear combination of all output values already observed:

\[
\hat{y}(x_0) = \sum_{j=1}^{n} \lambda_j y(x_i) = \lambda^t Y \text{ with } \sum_{j=1}^{n} \lambda_j = 1
\]

The weights $\lambda = (\lambda_1, \ldots, \lambda_n)^t$ in (8) are not constant and depend on the distances between the input to be predicted $x_0$ and the observed inputs $x_i$. Kriging assumes that the closer the input data, the more positively correlated the prediction errors. This assumption is modeled through the correlogram. In simulation, a popular class of correlograms is given by

\[
\rho(h) = \prod_{j=1}^{k} \exp(-\theta_j |h_j|^p)
\]

where $k$ denotes the number of inputs, $h = (h_1, \ldots, h_k)^t$ is the distance vector between two inputs $x_i$ and $x_j$, $\theta_j$ represents the importance of input $x_j$ (that is, the higher $\theta_j$ the less effect input $x_j$ has), and $p_j$ represents the smoothness of the correlogram function. Often, the powers $p_j$ are chosen as $p_j = p = 2$. Then, the resulting correlogram is the infinitely differentiable Gaussian correlation function.

The criterion to select the optimal weights $\lambda$ in Equation (8) is the mean-squared prediction error $\sigma^2$, which is defined as

\[
\sigma^2 = E(\left(\hat{Y}(x_0) - \hat{y}(x_0)\right)^2)
\]

Minimizing (10) under the constraint (8) gives the optimal weights (11):

\[
\lambda_{opt} = \Gamma^{-1} \left( \gamma + \frac{1}{1 - I^{-1} \Gamma^{-1}} \gamma \right)
\]

where $\gamma$ denotes the vector of covariances $(\gamma(x_0 - x_1), \ldots, \gamma(x_0 - x_n))^t$, $\Gamma$ denotes the $n \times n$ matrix whose elements are the covariances between the outputs observed by simulating the model (i.e. $\{i, j\}$th element is $\gamma(x_i - x_j)$), the covariance between $x_i$ and $x_j$, and $I = (1, \ldots, 1)^t$ is the unit vector. Note that some of the weights $\lambda_i$ may be negative.

As we see, the optimal weights in (11) depend on the covariances between observed output data. These
optimal weights give the minimal mean-squared prediction error (also see Cressie,\(^7\) p.122):

\[
\sigma^2 = \gamma' \Gamma^{-1} \gamma - \left( \frac{1}{\Gamma^{-1} \gamma} \right)^2
\]

(12)

Actually, the covariance \(\gamma(h)\) in (11) and (12) is unknown. The usual estimator of this covariance is given by

\[
2\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Y(x_i) - Y(x_j))^2
\]

(13)

where \(|N(h)|\) denotes the number of distinct pairs in \(N(h) = \{(x_i, x_j): x_i - x_j = h; i, j = 1, \ldots, n\}\); see Matheron.\(^11\) These estimates imply estimates of the corresponding correlations, so the parameters \(\theta_j\) and \(p_j\) in (11) can be fitted. For this fitting, standard Kriging software uses Maximum Likelihood Estimation. Van Beers and Kleijnen,\(^12\) however, used Weighted Least Squares estimation for a linear correlogram function. These estimated covariances are substituted into (11) to estimate the optimal weights.

2. The constrained optimization problem

A general statement of the constrained optimization problem is as follows:

\[
\min f_0(x)
\]

subject to \(f_1(x_1, x_2) \leq 20\), \(f_2(x_1, x_2) \leq 8\), \(x_1 \geq 0\), \(x_2 \geq 0\), and \(x_1 \geq x_2\).

The Arena-based \((s, S)\) model was simulated for \(r = 5\) replications, (see Law,\(^14\)) at each of the \(k = 13\) design points in the CCD illustrated in Figure 1, yielding the results shown in Table 1. The surface graphs for holding cost, shortage cost, and total cost are shown in Figure 2. Note that the response surfaces for holding cost and shortage cost are both highly planar in the experimental region, while the total cost surface is nearly quadratic.

As stated earlier, most metamodeling studies focus on low-order polynomial regression using factorial-based experimental designs. The CCD is a popular experimental design for estimating a quadratic regression in a local, unimodal region of the design space. Figure 1 illustrates a spreadsheet showing the deployment of \(k = 2^n + 2n + c\) design points, where \(n\) is the number of input parameters or design variables, and \(c\) is the number of replications of the center point of the design space that is selected to follow the recommended values in the literature (i.e. 3–5).\(^15\) In the case shown in Figure 1, \(n = 2\) and \(c = 5\), therefore \(k = 13\). There are \(2^2 = 4\) points arrayed in a square (generally, an \(n\)-dimensional hypercube with \(2^n\) corner points), \(2 \times 2 = 4\) points along the two axes (generally, \(2n\)), and \(c = 5\) replications of the center point. In the case of simulation experimentation, \(r\) replications of the simulation are carried out at each of the design points using common random numbers, except in the case of

\[
\begin{array}{cccccccccccccccc}
S & 10 & 12 & 14 & 16 & 18 & 20 & 22 & 24 & 26 & 28 & 30 & 32 & 34 & 36 & 38 & \text{Sum} \\
40 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
45 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
50 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
55 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
60 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
65 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
70 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
75 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
80 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
85 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
90 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
95 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
100 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
105 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
110 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\text{Sum} & 1 & 0 & 2 & 0 & 0 & 0 & 0 & 7 & 0 & 0 & 0 & 0 & 2 & 0 & 1 & 13 \\
\end{array}
\]

Figure 1. A central-composite design for the \((s, S)\) inventory system.
the $c$ repetitions of the center point, where independent random number streams must be employed. The mean values for each of the $m$ simulation responses are recorded and multiple regression is applied to fit the response functions $y_j(x), j = 1, \ldots, m$. An optimization procedure, such as Matlab,$^{16}$ LINGO,$^{17}$ or MS-Excel Solver$^{18}$ is used to find an optimal solution to the problem formulation in (15).

<table>
<thead>
<tr>
<th>Trial</th>
<th>$s$</th>
<th>$S$</th>
<th>$s^2$</th>
<th>$S^2$</th>
<th>$s^2S$</th>
<th>Holding cost</th>
<th>Shortage cost</th>
<th>Total cost</th>
<th>$v(h)$</th>
<th>$v(s)$</th>
<th>$v(z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>25.57</td>
<td>7.79</td>
<td>119.78</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>50</td>
<td>1156</td>
<td>2500</td>
<td>1700</td>
<td>17.56</td>
<td>4.63</td>
<td>125.24</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>100</td>
<td>1156</td>
<td>10,000</td>
<td>3400</td>
<td>42.50</td>
<td>1.68</td>
<td>128.74</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>50</td>
<td>196</td>
<td>2500</td>
<td>700</td>
<td>11.64</td>
<td>22.15</td>
<td>123.66</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>100</td>
<td>196</td>
<td>10,000</td>
<td>1400</td>
<td>32.75</td>
<td>13.53</td>
<td>130.34</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>40</td>
<td>576</td>
<td>1600</td>
<td>960</td>
<td>9.73</td>
<td>15.04</td>
<td>128.18</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>110</td>
<td>576</td>
<td>12,100</td>
<td>2640</td>
<td>41.47</td>
<td>5.90</td>
<td>131.55</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>75</td>
<td>100</td>
<td>5625</td>
<td>750</td>
<td>21.11</td>
<td>22.16</td>
<td>129.23</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>38</td>
<td>75</td>
<td>1444</td>
<td>5625</td>
<td>2850</td>
<td>31.40</td>
<td>1.95</td>
<td>123.37</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>26.07</td>
<td>6.77</td>
<td>118.98</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>26.07</td>
<td>7.43</td>
<td>119.50</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>26.49</td>
<td>7.05</td>
<td>119.38</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>25.94</td>
<td>7.01</td>
<td>120.30</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Figure 2.** Surface graphs for holding cost, shortage cost, and total cost for the $(s, S)$ inventory system resulting from a 13-point central-composite experimental design.
The three rightmost columns in Table 1 (\(v(h)\) for holding cost, \(v(s)\) for shortage cost, and \(v(c)\) for any constraint violation) are Boolean vectors in which the value 1 for a given element indicates a constraint violation, whereas a 0 represents a feasible solution. Hence, among the experimental trials, only trial 2 yielded a feasible solution. However, the application of an optimization tool could possibly predict other feasible solutions. Here, MS-Excel Solver\(^{18}\) was applied to a constrained optimization problem formed by using the fitted regression metamodels for the responses \(f_j(x), j = 0,1,2\).

MS-Excel Solver\(^{18}\) was used to obtain a solution to the constrained optimization problem as stated in (11), with the solution \(x_1 = 25.42, x_2 = 61.42, f_0(x) = \$118.49, f_1(x) = \$20.00,\) and \(f_2(x) = \$8.00\). Inasmuch as real-valued solutions for \(s\) and \(S\) are infeasible, simulation trials were run at the \(2^2\) design points surrounding the predicted \(x^*\). The design point \(s = 26\) and \(S = 61\) yielded the solution total cost \(f_0(x) = \$116.24\), holding cost \(f_1(x) = \$19.97\), and shortage cost \(f_2(x) = \$7.69\). Hence, the holding cost constraint is nearly binding at this solution.

### 2.2. A generalized Kriging approach to constrained simulation optimization

The key to applying Kriging to an experimental situation is the LHD. The difficulty with factorial-based designs, such as the CCD illustrated above, is that they are non-space filling, as was shown in Figure 1. An experimental design is said to be space filling if the \(k\) design points are deployed throughout the design space in such a way that they are as near as possible equidistant from one another. Clearly, the CCD design points in Figure 1 do not possess this feature. The ‘sum’ values at the right-hand end of each row and the bottom of each column in Figure 1 show this situation quite clearly; that is, the many zeros among these ‘sum’ values demonstrate the non-space filling characteristic of CCDs. The presence of ‘sum’ values greater than 1 signify another undesirable feature of CCDs from the standpoint of Kriging; that is, CCDs are collapsing, and as such are unacceptable for the application of Kriging.

### 2.3. Latin Hypercube Designs

An alternative to factorial-based CCDs is a space-filling design such as the LHD, which is discussed by Kleijnen.\(^1\) These designs have particularly favorable characteristics when applied in a constrained optimization experimental setting in that there is a high probability that one or more design points will not only fall within the feasible region (which is not known a priori), but will actually fall close to a constrained optimum solution \((x^*, y^*)\). LHDs are also particularly well suited for Kriging in that they can be made to cover the design space in such a way that design points are almost equidistant from one another in \(n\)-dimensional space. Figure 3 illustrates a non-collapsing, space-filling LHD. Non-collapsing, space-filling designs, such as that illustrated in Figure 3, are preferred when applying Kriging to simulation experimentation. The fact that the ‘sum’ values in the rows and columns of Figure 3 are all 1 is an indication of the non-collapsing and space-filling features of LHDs.

A generalized Kriging heuristic approach to constrained simulation optimization is as follows.

1. Formulate the constrained optimization problem as stated in (14).
2. Enumerate the design space \(a \leq x \leq c\) for the \(n\) input factors \(x_i, i = 1,\ldots,n\).
3. Place a LHD comprised of \(8n \leq k \leq 12n\) design points in the design space.
4. Run \(r\) replications of the simulation model at each of these \(k\) design points.
5. Enter the values of the \(n\) input factors \(x_i, i = 1,\ldots,n\) and the \(m + 1\) responses \(f_j(x), j = 0,1,\ldots,m\) for each of the \(k\) simulation trials into a table of results. Save this table of results as a text file titled LHD.txt.
6. Compute Kriging predictions for a grid of \(Z = p^m\) prediction points, where \(p\) is the number of levels of each design variable in the LHD and \(n\) is the number of design variables with using the program krig.m from the Matlab\(^{19}\) toolkit, DACE\(^{20}\), and the input file LHD.txt. This operation results in Kriging predictions at each of the \(Z\) points for each of the \(m + 1\) simulation responses.
7. Test whether a prediction point in the \(p^m\) grid satisfies the constraint function \(f_j(x), j = 1,\ldots,m\). If constraint \(f_j(x)\) is satisfied, set \(v_{jz} = 0\); if constraint \(f_j(x)\) is violated, then set \(v_{jz} = 1\). If \(v_{jz} = 1\) for any constraint \(f_j(x)\), then by the Boolean OR operation \(v_z = 1\). If a prediction point violates none of the \(m\) constraints, then by the Boolean AND operation \(v_z = 0\). The set of \(p^m\) binary values \(v_z, z = 1,\ldots,p^m\) is called the ‘constraint violation vector.’ (Note: it is necessary to represent the binary constraint violation results as a vector since it is impractical to view all possible two-variable surface graphs of the constraint violation function \(v_z\).)
8. Evaluate the objective function \(f_0(x)\) at each point in the \(p^m\) prediction grid for which \(v_z = 0\); that is, for each feasible prediction point.
9. Select the grid point \(x_z\) at which the objective function \(f_0(x)\) is minimized (or maximized) as a predicted optimum solution \([x^*, f^*(x)]\).
10. Using the Kriging prediction program krig.m, evaluate the \(2^n\) design points bounding the predicted optimum \([x^*, f^*(x)]\). Select the point that minimizes \(f_0(x)\). Using \(r\) replications of the simulation model at each of these \(2^n\) design points, verify the new predicted optimum \([x^*, f^*(x)]\).

Although this 10-step approach cannot be proven to produce a globally optimal solution, owing to the stochastic nature of the responses \(f(x)\), it is nevertheless very likely to do so because of the exhaustive search in each modal region of the design space. A modal region is identified by a 'patch' of zeros along the constraint violation vector \(v\).

2.4. An illustrative example

To evaluate a space-filling LHD with the constrained optimization problem, the Arena-based \((s, S)\) model of Kelton et al.\(^{13}\) was modified through the addition of some variables. The model was run for \(r = 5\) replications at each design point in the 20-point LHD shown in Table 2 and Figure 3. Note that the bottom 'sum' row of 1s and the rightmost 'sum' column of 1s in Figure 3 verify that the LHD satisfies the requirement that there be one and only one design point at each of the selected values of the input variables \(x_i, i = 1, \ldots, n\). The model was run for a period of 120 weeks in each replication, with a starting inventory of \(S\) units.

The three rightmost columns in Table 2 show whether a design point violates either the holding cost constraint \(v_h\) or the shortage cost constraint \(v_s\), with the

![Table 2. Latin Hypercube Design-based optimization experiments for the (s, S) inventory system](image)

<table>
<thead>
<tr>
<th>Trial</th>
<th>s</th>
<th>S</th>
<th>(s^2)</th>
<th>(s^2S)</th>
<th>Holding cost</th>
<th>Shortage cost</th>
<th>Total cost</th>
<th>(v_h)</th>
<th>(v_s)</th>
<th>(v_z)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>48</td>
<td>400</td>
<td>960</td>
<td>11.6</td>
<td>18.84</td>
<td>124.99</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>58</td>
<td>441</td>
<td>1218</td>
<td>16.53</td>
<td>13.33</td>
<td>119.63</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>68</td>
<td>484</td>
<td>1496</td>
<td>21.45</td>
<td>10.49</td>
<td>121.41</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>78</td>
<td>529</td>
<td>2068</td>
<td>26.69</td>
<td>8.88</td>
<td>123.27</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>40</td>
<td>576</td>
<td>1600</td>
<td>9.73</td>
<td>15.04</td>
<td>128.18</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>25</td>
<td>50</td>
<td>625</td>
<td>2500</td>
<td>14.18</td>
<td>12.23</td>
<td>122.06</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>26</td>
<td>60</td>
<td>676</td>
<td>3600</td>
<td>18.89</td>
<td>6.68</td>
<td>118.24</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>27</td>
<td>70</td>
<td>729</td>
<td>4900</td>
<td>25.03</td>
<td>6.27</td>
<td>120.27</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>56</td>
<td>784</td>
<td>3136</td>
<td>17.71</td>
<td>8.01</td>
<td>120.62</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>42</td>
<td>841</td>
<td>1764</td>
<td>12.16</td>
<td>8.95</td>
<td>125.62</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>30</td>
<td>52</td>
<td>900</td>
<td>2704</td>
<td>16.98</td>
<td>7.46</td>
<td>121.96</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>31</td>
<td>62</td>
<td>961</td>
<td>3844</td>
<td>20.79</td>
<td>6.06</td>
<td>120.23</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>32</td>
<td>72</td>
<td>1024</td>
<td>5184</td>
<td>27.38</td>
<td>4.03</td>
<td>124.79</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>33</td>
<td>66</td>
<td>1089</td>
<td>4356</td>
<td>24.08</td>
<td>4.18</td>
<td>119.68</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>34</td>
<td>44</td>
<td>1156</td>
<td>1936</td>
<td>14.16</td>
<td>6.68</td>
<td>126.22</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>35</td>
<td>54</td>
<td>1225</td>
<td>2916</td>
<td>20.41</td>
<td>3.86</td>
<td>125.45</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>36</td>
<td>64</td>
<td>1296</td>
<td>4096</td>
<td>24.63</td>
<td>3.55</td>
<td>128.19</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>37</td>
<td>74</td>
<td>1369</td>
<td>5476</td>
<td>27.38</td>
<td>1.98</td>
<td>125.95</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>38</td>
<td>76</td>
<td>1444</td>
<td>5776</td>
<td>31.7</td>
<td>2.35</td>
<td>126.97</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>39</td>
<td>46</td>
<td>1521</td>
<td>2116</td>
<td>15.95</td>
<td>4.91</td>
<td>127.47</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
A column showing the Boolean OR for these results. A 1 indicates constraint violation, whereas a 0 indicates feasibility of the trial point. The presence of so few 0s should not be disturbing, as the Kriging procedure outlined above will generally produce a host of feasible prediction points in the prediction grid.

When a purely response surface approach is taken with the experimental results shown in Table 2, one obtains the surface graphs shown in Figure 4. MS-Excel Solver\textsuperscript{18} was applied to the second-order response surfaces produced by the LHD, with the following results:

\[ s = 26.34, \quad S = 63.48, \quad f_0(x) = $120.03, \quad f_1(x) = $14.17, \quad \text{and} \quad f_2(x) = $8.00 \]

Simulating the 2\textsuperscript{2} design points surrounding this predicted optimum, we find that none of these four points is feasible. Hence, the classical RSM approach fails to produce a satisfactory solution to this problem when applied with a LHD.

Figure 5 shows the surface graphs for the Kriging predictions of (a) holding cost, (b) shortage cost, and (c) total cost for the (s, S) inventory model for a 20-by-20 Kriging prediction grid. The formation of the binary constraint-violation vector \( v_z \) in accordance with the 10-step procedure outlined above produced a predicted optimum solution at \( s = 27, \quad S = 60, \quad f_0(x) = $118.56, \quad f_1(x) = $19.30, \quad \text{and} \quad f_2(x) = $7.94 \). Hence, the shortage cost constraint is nearly binding.

3. Multiple-objective optimization

The multiple-objective optimization problem can be written as follows:

\[
\max (\min) \quad G(f(x), j = 1, \ldots, m) \tag{16}
\]

subject to \( a \leq x \leq c \), where \( x \) is the \( n \)-vector of input factors \( x_i, \ i = 1, \ldots, n, \ a \) and \( c \) are lower bounds and upper bounds, respectively, on \( x \), and the function \( G(f(x)) \) is a policy for prioritizing \( f_j(x), \ j = 1, \ldots, m. \)
The following sections compare the classical RSM approach with a Kriging metamodeling technique for their effectiveness in deriving an optimal solution to the problem posed in (16).

3.1. A response surface approach to multi-objective optimization

To illustrate the regression metamodeling procedure using the CCD approach, an Arena model of the \((s, S)\) inventory system was simulated. The multi-objective optimization formulation of this system is to find the values of the reorder point \(s\), or \(x_1\), and an order-up-to level \(S\), or \(x_2\), so as to minimize total cost \(f_0(x)\), while maximizing customer satisfaction \(f_1(x)\), subject to the non-negativity conditions \(x_1 \geq 0\) and \(x_2 \geq 0\) and the constraint \(x_1 > x_2\) (or \(S > s\)). Total cost is the sum of holding cost, ordering cost, and shortage cost, while customer satisfaction is the proportion of customers whose demand is met immediately without resorting to backorders. The Arena-based \((s, S)\) model was simulated for \(r = 5\) replications at each of the \(k = 13\) design points in the CCD, yielding the results shown in Table 3. The response surfaces for total cost \(f_0(x)\) and \(f_1(x)\) are shown in Figure 6. It should be pointed out that the Minitab\(^{21}\) surface graph function utilizes the raw data and does not necessarily produce very regular-looking surface graphs.

The quadratic regression models for total cost and customer satisfaction are, respectively

\[
\begin{align*}
    f_0(s, S) &= 145 - 0.310s - 0.507S + 0.00928s^2 \\
    &+ 0.00439S^2 - 0.00318sS \\
    f_1(s, S) &= -0.154 + 0.0270s + 0.0159S \\
    &- 0.000188s^2 - 0.000064S^2 - 0.00015sS
\end{align*}
\]  

(17a) (17b)

Tables 4 and 5 show the analysis of variance (ANOVA) tables for models (17a) and (17b), respectively, produced by Minitab.\(^{22}\) In Table 4 we can see that the \(R^2\)-value and the \(p\)-value for the total cost regression (17a) are weak (70.0% and 0.077, respectively). In Table 5, however, the \(R^2\) and \(p\)-values for customer satisfaction regression (17b) are quite strong (98.2% and 0.000, respectively). The issue of a ‘weak’ \(p\)-value,
that is, $0.05 < p < 0.1$, is addressed by either lengthening the time duration of the simulation run or increasing the number of replications $r$, or both. In this instance, we increased the number of replications to $r = 100$, and achieved $p$-values less than 0.001 for both response surfaces. For illustrative purposes, however, we shall use the response surface models in (17a) and (17b).

Table 7 shows results for a selected set of prediction points that have values for customer satisfaction exceeding 1.0 for the RSM approach, which are clearly the result of prediction errors. A comparison with the same predictions from Kriging shows that the corresponding Kriging predictions were more reliable. When these ‘troubling’ prediction points are simulated, there is strong agreement between the Kriging predictions and the simulation results.

Figure 3 is a spreadsheet representation of a LHD for an experiment involving 20 design points for the two input factors $s$ and $S$. The recommended number of design points in the LHD is $10n$, where $n$ is the number of input factors. The column of 1s at the right-hand side of the spreadsheet and the row of 1s at its bottom demonstrate the essential space-filling character of the LHD – that is, there must be exactly one 1 in every row and exactly one 1 in every column of the LHD. Harking back to Figure 1, we see that the CCD does not possess this feature and hence cannot be applied with Kriging metamodeling.

Table 6 gives the simulation results for the 20 design points in the LHD. Figure 7 shows surface plots of total cost and customer satisfaction, respectively. The Kriging features of the Matlab$^{16}$ DACE toolbox$^{20}$ are applied to the data $x$ and $y$. The optimal solution ($x^*$$^8$ is

---

**Table 3.** Simulation results for a central-composite experimental design with the $(s, S)$ inventory system

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>S</th>
<th>$s^2$</th>
<th>$S^2$</th>
<th>$sS$</th>
<th>Total cost $$$</th>
<th>Customer satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>119.78</td>
<td>0.944</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>50</td>
<td>1156</td>
<td>2500</td>
<td>1700</td>
<td>125.24</td>
<td>0.948</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>100</td>
<td>1156</td>
<td>10,000</td>
<td>3400</td>
<td>128.74</td>
<td>0.992</td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>196</td>
<td>2500</td>
<td>700</td>
<td>1400</td>
<td>130.34</td>
<td>0.725</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>196</td>
<td>10,000</td>
<td>1400</td>
<td>2640</td>
<td>131.55</td>
<td>0.969</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>40</td>
<td>576</td>
<td>1600</td>
<td>2640</td>
<td>131.55</td>
<td>0.755</td>
</tr>
<tr>
<td>7</td>
<td>24</td>
<td>110</td>
<td>12,100</td>
<td>2640</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>75</td>
<td>100</td>
<td>5625</td>
<td>750</td>
<td>129.23</td>
<td>0.82</td>
</tr>
<tr>
<td>9</td>
<td>38</td>
<td>75</td>
<td>1444</td>
<td>5625</td>
<td>2850</td>
<td>123.37</td>
<td>0.986</td>
</tr>
<tr>
<td>10</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>124.61</td>
<td>0.953</td>
</tr>
<tr>
<td>11</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>126.54</td>
<td>0.935</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>124.48</td>
<td>0.952</td>
</tr>
<tr>
<td>13</td>
<td>24</td>
<td>75</td>
<td>576</td>
<td>5625</td>
<td>1800</td>
<td>124.01</td>
<td>0.955</td>
</tr>
</tbody>
</table>

---

**Figure 6.** Surface plots for total cost and customer satisfaction for the $(s, S)$ inventory system as estimated by a central-composite design (CCD).
 solutions for which total cost is less than $120, but with customer satisfaction around 0.96. When the two constraints are tightened, the set of circles shows solutions for which total cost is between $120 and $121, with customer satisfaction exceeding 0.98. This Pareto optimal approach is ideal for examining Kriging results. This successive tightening of the constraint bounding values $d_j$, $j = 1, \ldots, m$ is often a very effective strategy in multiple-response optimization, and is certainly shown to be effective in this instance.

4. Conclusions

This study has shown that Kriging metamodeling has the potential to identify superior solutions to those obtained by classical response surface approaches when faced with a multiple-objective optimization environment. This superiority is due both to the space-filling character of LHDs, as opposed to the relative sparseness of design points in a CCD, as well as to the more accurate predictions offered by the Kriging approach. One downside to Kriging metamodeling, however, is that it only works when there are a relatively large number of design points in the LHD. Kriging worked well in this study when we used 10n
Table 5. Minitab-based Regression Analysis: Customer Satisfaction versus s and S

The regression equation is

\[
\text{Customer Satisfaction} = -0.154 + 0.0270 \, s + 0.0159 \, S - 0.000188 \, s^2 - 0.000064 \, S^2 + 0.000151 \, s \times S
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.15440</td>
<td>0.08478</td>
<td>-1.82</td>
<td>0.111</td>
</tr>
<tr>
<td>s</td>
<td>0.027004</td>
<td>0.003675</td>
<td>7.35</td>
<td>0.000</td>
</tr>
<tr>
<td>S</td>
<td>0.015886</td>
<td>0.001605</td>
<td>9.90</td>
<td>0.000</td>
</tr>
<tr>
<td>S^2</td>
<td>-0.00018770</td>
<td>0.00005878</td>
<td>-3.19</td>
<td>0.015</td>
</tr>
<tr>
<td>S^2</td>
<td>-0.00006351</td>
<td>0.00000940</td>
<td>-6.75</td>
<td>0.000</td>
</tr>
<tr>
<td>s*S</td>
<td>-0.00015125</td>
<td>0.00003056</td>
<td>-4.95</td>
<td>0.002</td>
</tr>
</tbody>
</table>

\[ S = 0.0152807 \] \( R^2 = 98.2\% \) \( R^2(\text{adj}) = 96.9\% \)

Analysis of Variance

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>5</td>
<td>0.089621</td>
<td>0.017924</td>
<td>76.76</td>
<td>0.000</td>
</tr>
<tr>
<td>Residual Error</td>
<td>7</td>
<td>0.001634</td>
<td>0.000233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>0.091255</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Seq SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>1</td>
<td>0.035028</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
<td>0.036883</td>
</tr>
<tr>
<td>S^2</td>
<td>1</td>
<td>0.001343</td>
</tr>
<tr>
<td>S^2</td>
<td>1</td>
<td>0.010648</td>
</tr>
<tr>
<td>s*S</td>
<td>1</td>
<td>0.005719</td>
</tr>
</tbody>
</table>

Unusual Observations

<table>
<thead>
<tr>
<th>Obs</th>
<th>s</th>
<th>Customer Satisfaction</th>
<th>Fit</th>
<th>SE Fit</th>
<th>Residual</th>
<th>St Resid</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>34</td>
<td>0.94752</td>
<td>0.92513</td>
<td>0.01213</td>
<td>0.02239</td>
<td>2.41R</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>0.75502</td>
<td>0.77419</td>
<td>0.01203</td>
<td>-0.01917</td>
<td>-2.04R</td>
</tr>
</tbody>
</table>

R denotes an observation with a large standardized residual.

Table 6. Customer satisfaction predictions erroneously exceeding 1.00

<table>
<thead>
<tr>
<th>s</th>
<th>S</th>
<th>Customer satisfaction (response surface)</th>
<th>Customer satisfaction (Kriging)</th>
<th>Customer satisfaction (simulated with Arena)</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>77</td>
<td>1.001</td>
<td>0.985</td>
<td>0.984</td>
</tr>
<tr>
<td>37</td>
<td>75</td>
<td>1.001</td>
<td>0.986</td>
<td>0.983</td>
</tr>
<tr>
<td>37</td>
<td>76</td>
<td>1.002</td>
<td>0.986</td>
<td>0.985</td>
</tr>
<tr>
<td>37</td>
<td>77</td>
<td>1.002</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td>37</td>
<td>78</td>
<td>1.003</td>
<td>0.987</td>
<td>0.986</td>
</tr>
<tr>
<td>38</td>
<td>76</td>
<td>1.003</td>
<td>0.985</td>
<td>0.985</td>
</tr>
<tr>
<td>38</td>
<td>77</td>
<td>1.004</td>
<td>0.985</td>
<td>0.987</td>
</tr>
<tr>
<td>38</td>
<td>78</td>
<td>1.004</td>
<td>0.986</td>
<td>0.988</td>
</tr>
<tr>
<td>38</td>
<td>79</td>
<td>1.004</td>
<td>0.985</td>
<td>0.988</td>
</tr>
<tr>
<td>39</td>
<td>75</td>
<td>1.004</td>
<td>0.986</td>
<td>0.991</td>
</tr>
<tr>
<td>39</td>
<td>76</td>
<td>1.004</td>
<td>0.986</td>
<td>0.988</td>
</tr>
<tr>
<td>39</td>
<td>77</td>
<td>1.004</td>
<td>0.986</td>
<td>0.986</td>
</tr>
<tr>
<td>39</td>
<td>78</td>
<td>1.005</td>
<td>0.987</td>
<td>0.987</td>
</tr>
<tr>
<td>40</td>
<td>75</td>
<td>1.005</td>
<td>0.984</td>
<td>0.987</td>
</tr>
<tr>
<td>40</td>
<td>76</td>
<td>1.005</td>
<td>0.983</td>
<td>0.993</td>
</tr>
<tr>
<td>40</td>
<td>77</td>
<td>1.005</td>
<td>0.984</td>
<td>0.990</td>
</tr>
</tbody>
</table>
design points. We tried, without success, a LHD of 10
design points to a simulation model of a cellular
manufacturing system having four input factors. The
four input factors were the queue capacities at each of
four machine cells and were restricted to integer values.
Yet, the finite queues would ideally only range in size
from 1 to 5 (the mean queue sizes for typical simulation
trials were in the range from 1.5 to 3), so even letting
$1 \leq x_i \leq 10$ was impractical. In this case, CCDs
permitted the values $1 \leq x_i \leq 5$ and gave excellent
results. A LHD over the same range would have only
had five design points, far fewer than the $10n = 40$ rec-
ommended for Kriging metamodeling.

The success of the Kriging metamodeling
approaches to the multiple-objective optimization
problem examined in this paper, especially with the
SOOP formulation (5), can be traced to the 10-step
procedure outlined above. The length of the Boolean

![Surface Plot of Total Cost vs bigS, s (LHD)](image1)
![Surface Plot of Customer Satisfaction vs bigS, s (LHD)](image2)

**Figure 7.** Surface plots of Kriging predictions for total cost and customer satisfaction based on a Latin Hypercube Design (LHD).

![Trade-off Curves](image3)

**Figure 8.** A Pareto chart of Kriging results for the $(s, S)$ inventory system.
constraint violation vector $v_2 (20^2 = 400 \text{ here})$ can be imposing, but one is easily able to identify the feasible prediction values $\mathbf{x}$ and evaluate the objective function values $f(x_0)$ at these points.

**Acknowledgement**

The authors express their sincere appreciation for the helpful suggestions of two anonymous referees. Incorporating their suggestions has greatly improved this paper.

**Funding**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**Conflict of interest statement**

None declared.

**References**


**Mehdi Zakerifar** is a PhD candidate in the Department of Industrial Engineering at the University of Louisville. He received his BS degree in Applied Mathematics from the University of Tehran (Tehran, Iran) in 2000, and his Master of Engineering Degree in Industrial Engineering from Sharif University of Technology (Tehran, Iran) in 2003. He has worked as an industrial engineer, systems analyst, and later network administrator, database manager and software developer. He also has worked as a teaching assistantship at the University of Louisville since 2007. His research interests include simulation modeling and optimization, decision analysis, multi-criteria, and applied optimization.

**William E Biles** is a Professor in the Department of Industrial Engineering at the University of Louisville. He received his BS in Chemical Engineering from Auburn University, his MSIE from the University of Alabama in Huntsville, and his PhD in Industrial Engineering from Virginia Tech. He spent two years in the US Army and seven years in industrial R&D at Union Carbide’s Advanced Materials Laboratory and at Thiokol Chemical Corporation before undertaking an academic career. He has held academic appointments at the University of Notre Dame, Penn State, Louisiana State University, and the University of Louisville. Dr Biles received the Distinguished Engineering Alumnus award from the University of Alabama in Huntsville in 2004. His most recent publication is ‘Environmentally Benign Manufacturing,’ Chapter 1 in the 2007 John Wiley handbook *Environmentally Conscious Manufacturing*.

**Gerald W Evans** is a Professor in the Department of Industrial Engineering at the University of Louisville. He received his BS degree in Mathematics, and MS and PhD degrees in Industrial Engineering, all from Purdue University. Before entering academia, he worked as an industrial engineer for Rock Island Arsenal, and as a senior research engineer for General Motors Research Laboratories. He has also served as a NASA/ASEE faculty fellow at NASA Langley Research Laboratory and at NASA Kennedy Space Center. Besides simulation modeling and analysis, his research interests include multi-objective optimization, decision analysis, and discrete optimization.