Robust Optimization in Simulation: Taguchi and Krige Combined

Gabriella Dellino
IMT Institute for Advanced Studies, 55100 Lucca, Italy, g.dellino@imtlucca.it

Jack P. C. Kleijnen
Department of Information Management and CentER, Tilburg School of Economics and Management, Tilburg University, 5000 LE Tilburg, The Netherlands, kleijnen@uvt.nl

Carlo Meloni
Department of Electrical Engineering and Electronics, Polytechnic University of Bari, 70125 Bari, Italy, meloni@deemail.poliba.it

Optimization of simulated systems is the goal of many methods, but most methods assume known environments. We, however, develop a "robust" methodology that accounts for uncertain environments. Our methodology uses Taguchi's view of the uncertain world but replaces his statistical techniques by design and analysis of simulation experiments based on Kriging (Gaussian process model); moreover, we use bootstrapping to quantify the variability in the estimated Kriging metamodels. In addition, we combine Kriging with nonlinear programming, and we estimate the Pareto frontier. We illustrate the resulting methodology through economic order quantity (EOQ) inventory models. Our results suggest that robust optimization requires order quantities that differ from the classic EOQ. We also compare our results with results we previously obtained using response surface methodology instead of Kriging.

Key words: statistics, design of experiments; inventory production, simulation; decision analysis: risk
History: Accepted by Marvin Nakayama, Area Editor for Simulation; received October 2009; revised July 2010, January 2011; accepted March 2011. Published online in *Articles in Advance* July 21, 2011.

1. Introduction

In practice, some inputs of a given simulation model are uncertain so the optimum solution that is derived—ignoring these uncertainties—may be wrong. Strategic decision making in such an uncertain world may use *Taguchi*'s approach, originally developed to help Toyota design "robust" cars, i.e., cars that perform reasonably well in many circumstances; see Beyer and Sendhoff (2007), Kleijnen (2008), Park et al. (2006), Taguchi (1987), and Wu and Hamada (2000).

We use Taguchi's view of the world but not his statistical methods. Taguchians use rather restrictive designs and analysis models, so we propose to adopt a design and analysis of simulation experiments (DASE) approach based on *Kriging*, which is also called a Gaussian process model (Kleijnen 2008). DASE including Kriging treats the simulation model as a black box; i.e., only the input/output (I/O) of the simulation model is observed. (Black-box methods have wider applicability but lower efficiency than white-box methods such as perturbation analysis and the score function.) This Kriging applied to simulation or "computer" models is explained by Sacks et al. (1989) and Santner et al. (2003). Kriging gives metamodels, which are also called response surfaces, surrogates, emulators, auxiliary models, and repro-models, to name a few; see Barton and Meckesheimer (2006). (We provide so many synonyms because simulation is used in many disciplines, each with its own terminology.) These metamodels run much faster than the underlying—and possibly computationally expensive—simulation models; e.g., a simulation run on a high-performance computer in the aerospace-engineering case study reported by Oberguggenberger et al. (2009) takes 32 hours. Branke et al. (2001) adopt regression models combined with evolutionary algorithms in noisy optimization, obtaining good performance in terms of both computer time savings and approximation quality. More recently, Dellino et al. (2007) discuss the design optimization of an automotive component through Kriging metamodels and an evolutionary algorithm, and they observe that the optimization process assisted by surrogates is three orders of magnitude faster than the one based on plain simulation.

Within the Taguchian framework, to solve the simulation-optimization problem, we combine Kriging metamodeling with *nonlinear programming* (NLP). In our NLP approach, we select one of the multiple simulation outputs as the goal or objective, and

the remaining outputs must satisfy given constraints or threshold values provided by the decision makers. This combination of Kriging and NLP gives an estimate of the robust solution of the simulation optimization problem.

Next, in the NLP model, we change specific threshold values for the constrained simulation outputs to estimate the *Pareto frontier*. Finally, the robust optimization literature ignores the analysis of the statistical variation in the estimated Kriging model. For this analysis we use *bootstrapping*. We also compare our results with those of a previous paper in which we used *response surface methodology* (RSM) to optimize the simulated system; see Dellino et al. (2010).

We use MATLAB software for the various components of our heuristic (i.e., design of experiments, metamodel fitting, metamodel-based optimization, and statistical analysis), because this software is well documented and is used by many simulationists for the postprocessing of their simulation I/O data. In general, the simulation model itself is programmed in either special simulation software (such as Arena, ProModel, or AnyLogic) or a general-purpose language (such as Fortran or C++); here, we program our simulation models in Arena. Even though MATLAB is known to be slow (in particular, in optimization, which is only one of the steps in our procedure), we program our methodology in MATLAB for the following reasons: (i) In "expensive" simulation, the postprocessing is (by definition) negligible compared with the computer time needed to run the underlying simulation model (which may take days or weeks to execute a single run); therefore, MATLAB does not significantly affect the performance of our framework. (ii) The speed of MATLAB code can be enhanced (if necessary) through interfacing (e.g., using `mex` files) with an external optimization solver. However, the choice for such a solver depends on the specific application, and it is beyond the scope of this research; we refer to Neumaier et al. (2005) for a comparison of global optimization solvers that perform a complete search. (iii) MATLAB has many built-in functions used in the various steps of our procedures, thus quickening research activities.

Our methodology may also be applied to cases affected by so-called *implementation errors*, which are the differences between the (nominal or expected) values chosen for the design parameters and the values in practice; see Stinstra and den Hertog (2008). Inspired by Bertsimas et al. (2010), we illustrate the application of our methodology to a problem that includes such implementation errors; see the Online Supplement (available at http://dx.doi.org/10.1287/ijoc.1110.0465).

The remainder of this paper is organized as follows. Section 2 presents Taguchi's worldview. Section 3 summarizes RSM for robust optimization. Section 4 discusses Kriging basics in §4.1 (readers familiar with Kriging may skip this subsection) and the use of Kriging for robust optimization including bootstrapping in §4.2. Section 5 illustrates the new methodology through the classic economic order quantity (EOQ) simulation model (which has a known I/O function and is a building block for more complicated and realistic supply-chain simulation models). Section 6 presents our conclusions and possible topics for future research. An extensive list of references follows. The Online Supplement provides more details, including additional computational experiments and supplementary references.

2. Taguchi's Worldview

Taguchi (1987) distinguishes between two types of factors (or inputs or variables): (i) *decision* or *control* factors, which managers can control and which we denote by d_j ($j = 1, \ldots, k$) (e.g., in inventory management, the order quantity is supposed to be controllable); and (ii) *environmental* or *noise* factors, which are beyond management's control and which we denote by e_g ($g = 1, \ldots, c$) (e.g., the demand rate in inventory management is an environmental factor). The latter type of factor is uncertain, so ignoring this lack of control may make the classic optimum solution wrong. Taguchi assumes a single output—say, w—and focuses on the ratio of the mean and the variance of this output.

We adopt Taguchi's view but not his statistical methods, which have been criticized by many statisticians; see Nair (1992) and the Online Supplement for further details. Instead of these methods, we use Kriging and include designs such as Latin hypercube sampling (LHS). Our reason for selecting Kriging is that the experimental area in computer simulation experiments may be much larger than in physical (real-life) experiments, so a low-order polynomial used by Taguchians may be an inadequate approximation (nonvalid metamodel). Our main reason for choosing a non-Taguchian design is that simulation experiments enable the exploration of many more factors, factor levels, and combinations of factor levels than do physical experiments. For further discussion of various metamodels and designs in simulation, we refer the reader to Kleijnen (2008) and Kleijnen et al. (2005).

In our approach we keep the main Taguchian robustness concept that prescribes to optimize a specific performance measure while minimizing its possible variations. However, instead of Taguchi's scalar loss function, we use nonlinear programming (see (1) below) in which one of these characteristics—namely, the mean of the primary simulation output,

$E(w)$—is the goal function to be minimized, and the other characteristics—namely, the standard deviation of the goal output, s_w—must meet given constraints (Lehman et al. 2004 also minimize the mean while satisfying a constraint on the variance; they use a Bayesian approach):

$$\min \; E(w)$$
$$\text{s.t.} \; s_w \leq T. \tag{1}$$

Because we assume costly simulations, we use a metamodel-assisted optimization; i.e., we replace $E(w)$ and s_w by their (Kriging) approximations. Next we change the threshold T in the constraint in (1) and estimate the Pareto-optimal efficiency frontier—called the *Pareto frontier* for short—where we consider the mean and standard deviation as criteria for finding a trade-off. This is a classical approach for solving optimization problems with multiple criteria; see Miettinen (1999). For further discussion of robust optimization, we also refer the reader to Beyer and Sendhoff (2007), Jin and Sendhoff (2003), Myers et al. (2009), Park et al. (2006), and Wu et al. (2009).

3. RSM and Robust Optimization

In this section we summarize Dellino et al. (2010), who use RSM for robust simulation optimization; this RSM guides our Kriging approach. Like Myers et al. (2009), Dellino et al. (2010) use the following type of RSM metamodel, which consists of an incomplete second-order polynomial approximation of the I/O function implied by the simulation model:

$$y = \beta_0 + \boldsymbol{\beta}'\mathbf{d} + \mathbf{d}'\mathbf{B}\mathbf{d} + \boldsymbol{\gamma}'\mathbf{e} + \mathbf{d}'\boldsymbol{\Delta}e + \epsilon, \tag{2}$$

where y denotes the regression predictor of the expected (mean) simulation output $E(w)$, ϵ denotes the residual with $E(\epsilon) = 0$ if this metamodel has no lack of fit (i.e., if this metamodel is valid) and with constant variance σ_ϵ^2; $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ denotes the first-order effects of the decision variables $\mathbf{d} = (d_1, \ldots, d_k)'$; $\mathbf{B}$ denotes the $k \times k$ symmetric matrix with the purely quadratic effects $\beta_{j;j}$ on the main diagonal and half the interaction effects $\beta_{j;j'}/2$ off the diagonal; $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_c)'$ denotes the first-order effects of the environmental factors $\mathbf{e} = (e_1, \ldots, e_c)'$; and $\boldsymbol{\Delta} = (\delta_{j;g})$ denotes the "decision-by-environmental" two-factor interactions.

To examine whether the assumed metamodel (2) is an *adequate approximation*, Dellino et al. (2010) use leave-one-out cross-validation. In brief, this cross-validation eliminates one I/O combination, recomputes the regression parameter estimates, and compares the estimated regression output with the simulation output eliminated.

Using the symbol $E(\mathbf{e}) = \boldsymbol{\mu}_\mathbf{e}$, it is easy to derive that (2) implies

$$E(y) = \beta_0 + \boldsymbol{\beta}'\mathbf{d} + \mathbf{d}'\mathbf{B}\mathbf{d} + \boldsymbol{\gamma}'\boldsymbol{\mu}_\mathbf{e} + \mathbf{d}'\boldsymbol{\Delta}\boldsymbol{\mu}_\mathbf{e}. \tag{3}$$

In denoting the variances and covariances of $\mathbf{e}$ by $\text{cov}(\mathbf{e}) = \boldsymbol{\Omega}_\mathbf{e}$, (2) implies

$$\text{var}(y) = (\boldsymbol{\gamma}' + \mathbf{d}'\boldsymbol{\Delta})\boldsymbol{\Omega}_\mathbf{e}(\boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{d}) + \sigma_\epsilon^2 = \mathbf{l}'\boldsymbol{\Omega}_\mathbf{e}\mathbf{l} + \sigma_\epsilon^2, \tag{4}$$

where $\mathbf{l} = \boldsymbol{\gamma} + \boldsymbol{\Delta}'\mathbf{d}$. The mean and variance in (3) and (4) may be estimated through plugging the ordinary least squares (OLS) estimates of the regression coefficients in (2) into the right-hand sides of (3) and (4), respectively. (del Castillo 2007, pp. 250–253, shows how the resulting bias in the estimated variance might be eliminated; however, Dellino et al. 2010 do not use this bias correction.)

Dellino et al. (2010) minimize the estimated mean—say, $\hat{\bar{y}}$—resulting from plugging the OLS estimates $\hat{\beta}_0$ through $\hat{\boldsymbol{\Delta}}$ into (3) while keeping the estimated standard deviation $\hat{\sigma}_y$ resulting from (4) below a given threshold T (see (1)). They use the standard deviation instead of the variance because the standard deviation has the same scale as the mean. They solve this constrained minimization problem that is nonlinear in the decision variables $\mathbf{d}$ (see (3) and (4)); this gives the estimated robust decision variables $\hat{\mathbf{d}}^+$. Next, the authors vary T, which may give different solutions $\hat{\mathbf{d}}^+$ with corresponding $\hat{\bar{y}}^+$ and $\hat{\sigma}_y^+$. They then collect the pairs $(\hat{\bar{y}}^+, \hat{\sigma}_y^+)$ to estimate the Pareto frontier. Finally, they estimate the variability of this frontier through parametric bootstrapping of the OLS estimates that gave $\hat{\bar{y}}^+$ and $\hat{\sigma}_y^+$ (we shall return to bootstrapping in §4.2). They illustrate their methodology through some of the EOQ models that we shall use and extend in §5.

4. Kriging and Robust Optimization

In this section we first summarize the basics of Kriging; we then discuss how to use Kriging for robust optimization.

4.1. Kriging Basics

We base this subsection on Kleijnen (2008, 2010). Typically, Kriging models are fitted to data that are obtained for larger global experimental areas than the small local areas used in low-order polynomial regression metamodels such as (2). Originally, Kriging was developed in geostatistics—also known as spatial statistics—by the South African mining engineer Danie Krige; see the classic geostatistics textbook by Cressie (1993). Later on, Kriging was often applied to the I/O data of deterministic simulation models in engineering; see the classic article by Sacks et al. (1989). Recently, Kriging has also been applied to random (stochastic) simulation in operations research and management science; see Ankenman et al. (2010).

Kriging uses the *linear* predictor

$$y = \boldsymbol{\lambda}' \mathbf{w}, \tag{5}$$

where—unlike the regression coefficients in (2)—the weights $\boldsymbol{\lambda}$ are not constants but decrease with the *distance* between the "new" input combination to be predicted and the "old" input combinations that have already been simulated and that resulted in the simulation outputs $\mathbf{w}$.

Our heuristic uses the simplest type of Kriging called *ordinary* Kriging, which assumes

$$w = \mu + \delta, \tag{6}$$

where w is the simulation output (which depends on the input combination), μ is the simulation output averaged over the whole experimental area, and δ is the additive noise that forms a stationary covariance process (so its covariances decrease with the distances between the simulation input combinations) with zero mean. Note that Ankenman et al. (2010) call δ the "extrinsic noise." They add another term to (6) that they call "intrinsic noise," which is caused by the pseudo-random numbers (PRNs) that are used in random simulation, so the same input combination still shows intrinsic noise (this intrinsic noise has variances that may vary with the input combination and is correlated if common PRNs are used to simulate various input combinations). The Online Supplement discusses a more general approach for building Kriging metamodels, namely, *universal* Kriging. Notice that our framework allows the replacement of ordinary Kriging by universal Kriging with limited programming efforts.

If (6) holds, then the *optimal* weights in (5) can be proven to be

$$\boldsymbol{\lambda}_o = \boldsymbol{\Gamma}^{-1} \left[\boldsymbol{\gamma} + \mathbf{1} \frac{1 - \mathbf{1}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}}{\mathbf{1}'\boldsymbol{\Gamma}^{-1}\mathbf{1}} \right], \tag{7}$$

where $\boldsymbol{\Gamma} = \{\text{cov}(w_i, w_{i'})\}$ $(i, i' = 1, \ldots, n)$ is the $n \times n$ matrix with the covariances among the n old outputs, and $\boldsymbol{\gamma} = \{\text{cov}(w_i, w_0)\}$ is the n-dimensional vector with the covariances between the old outputs w_i and w_0, the output of the combination to be predicted that may be either new or old. Obviously, $\boldsymbol{\gamma}$ varies with w_0, so $\boldsymbol{\lambda}_o$ in (7) varies with w_0.

The covariances $\boldsymbol{\Gamma}$ and $\boldsymbol{\gamma}$ in (7) are often based on the *Gaussian correlation function*

$$\exp\left(-\sum_{j=1}^{k} \theta_j h_j^2 \right) = \prod_{j=1}^{k} \exp(-\theta_j h_j^2), \tag{8}$$

where h_j denotes the Euclidian distance between input j of the new and the old combinations, and θ_j denotes the importance of input j (the higher θ_j is,

the less effect input j has). Different models for the correlation function are also possible, as discussed in the Online Supplement.

Substituting the correlation function (8) into (7) implies that the weights are relatively high for inputs close to the input to be predicted. Moreover, the optimal weights (7) imply that for an old input, the predictor equals the observed simulation output at that input (all weights are zero except the weight of the observed output); i.e., the Kriging predictor is an *exact interpolator*. (RSM uses OLS, which minimizes the sum of squared residuals, so it is not an exact interpolator; Kriging accounting for the intrinsic noise in random simulation is also not an exact interpolator.)

Thus, the optimal weights (7) depend on the correlation function (8), but this correlation function is unknown. Consequently, the parameter values θ_j in (8) must be estimated. The standard Kriging software and literature use maximum likelihood estimators (MLEs) assuming the noise δ in (6) is (multivariate) Normally (Gaussian) distributed. We estimate the correlation functions and corresponding optimal weights through DACE, a MATLAB toolbox that is free of charge and well documented by Lophaven et al. (2002).

To get the I/O simulation data to which the Kriging model is fitted, simulation analysts often use LHS, assuming that a valid metamodel is more complicated than a low-order polynomial (which is assumed in RSM). LHS does not assume a specific metamodel or I/O function; instead, it tries to fill the design space formed by the simulation inputs. That is, LHS is a space-filling design (references and websites for other space-filling designs are given in Kleijnen 2008, pp. 127–130). We further discuss our design of experiments (DOE) through the EOQ examples in §5.

4.2. Two Kriging Approaches to Robust Simulation Optimization

To solve robust simulation-optimization problems, we propose the following two approaches using Kriging metamodels, sketched in Figure 1.

1. Inspired by Dellino et al. (2010), we fit *two* Kriging metamodels, namely, one model for the mean and one for the standard deviation—both estimated from the *simulation*'s I/O data.

2. Inspired by Lee and Park (2006), we fit a *single* Kriging metamodel to a relatively small number, say, n, of combinations of the decision variables $\mathbf{d}$ and the environmental variables $\mathbf{e}$. Next, we use this metamodel to compute the *Kriging predictions* for the simulation output w for $N \gg n$ combinations of $\mathbf{d}$ and $\mathbf{e}$ accounting for the distribution of $\mathbf{e}$.

In the first approach, which we refer to as one-layer Kriging metamodeling (1L-KM), we select the input combinations for the simulation model through

Robust Optimization Combining Taguchi and Kriging
Input: design matrix $\mathbf{X}_{n_\mathbf{d} \times n_\mathbf{e}}$ in the decision-by-environmental space $(\mathbf{d}, \mathbf{e})$; probability distribution of environmental factors;
begin
    for each design point (d_i, e_j) in $\mathbf{X}$, $i = 1, \ldots, n_\mathbf{d}$, $j = 1, \ldots, n_\mathbf{e}$
       1. Run the simulation model, obtaining $\mathbf{W} = w(d_i, e_j)$
    2. **switch** Kriging approach
      A. **Case** 1L-KM
        A.1 Estimate the average response $\bar{w}_i$ and its standard deviation s_i $(i = 1, \ldots, n_\mathbf{d})$
        A.2. Fit one metamodel for the average response, $K_1(\bar{w} \mid \mathbf{d})$, and one metamodel for its standard deviation, $K_2(s_w \mid \mathbf{d})$
        A.3. Validate the metamodels, using leave-one-out cross-validation
      B. **Case** 2L-KM
        B.1 Fit a Kriging metamodel for the output, $K(w \mid \mathbf{d}, \mathbf{e})$
        B.2 Validate the metamodel, using leave-one-out cross-validation
        B.3 Generate a new design matrix $\mathbf{X}^l_{N_\mathbf{d} \times N_\mathbf{e}}$ in the decision-by-environmental space
        B.4 Estimate the average predicted response $\hat{\bar{y}}_i$ and its standard deviation $s_{\hat{y}}$ $(i = 1, \ldots, n_\mathbf{d})$
        B.5 Fit one Kriging metamodel for the mean of the predicted response, $K_1(\hat{\bar{y}} \mid \mathbf{d})$, and one for its standard deviation, $K_2(s_{\hat{y}} \mid \mathbf{d})$
        B.6 Validate the metamodels, using leave-one-out cross-validation
    3. Estimate the Pareto frontier, iteratively solving the problem in (1) with varying threshold values T
    4. Apply B times distribution-free bootstrap to the simulation output data w, obtaining w_b^* $(b = 1, \ldots, B)$
    5. **switch** Kriging approach
      Case 1L-KM
        For each $b = 1, \ldots, B$, repeat Steps A.1–A.3 replacing w with w^*
      Case 2L-KM
        For each $b = 1, \ldots, B$, repeat Steps B.1–B.6 replacing w with w^*
    6. Derive confidence regions using the bootstrapped metamodels
end

Figure 1   **Summary of the Two Kriging Approaches**

a *crossed* (combined) design for the decision and environmental factors (as is also traditional in Taguchian design); i.e., we combine the, say, $n_\mathbf{d}$, combinations of the decision variables $\mathbf{d}$ with the $n_\mathbf{e}$ combinations of the environmental variables $\mathbf{e}$. These $n_\mathbf{d}$ combinations are *space filling*, so we can avoid extrapolation when using the Kriging metamodels to obtain predictions; Kriging is known to give bad extrapolators, as observed by van Beers and Kleijnen (2004). The $n_\mathbf{e}$ combinations are *sampled* from their input distribution; we use LHS for this sampling. Simulating these $n_\mathbf{d} \times n_\mathbf{e}$ combinations gives the outputs $w_{i,j}$ with $i = 1, \ldots, n_\mathbf{d}$ and $j = 1, \ldots, n_\mathbf{e}$. These I/O data enable the following estimators of the $n_\mathbf{d}$ conditional means and variances:

$$\bar{w}_i = \frac{\sum_{j=1}^{n_\mathbf{e}} w_{i,j}}{n_\mathbf{e}} \quad (i = 1, \ldots, n_\mathbf{d}), \qquad (9)$$

$$s_i^2 = \frac{\sum_{j=1}^{n_\mathbf{e}} (w_{i,j} - \bar{w}_i)^2}{n_\mathbf{e} - 1} \quad (i = 1, \ldots, n_\mathbf{d}). \qquad (10)$$

These two estimators are *unbiased* because they do not assume any metamodel; metamodels are only approximations, so they may have a lack of fit.

Note that Dellino et al. (2010) mention that they use a crossed design, even though RSM does not require such a design. Alternatives for a crossed design are the split-plot design presented by Dehlendorff et al. (2011) or simultaneous perturbation stochastic approximation (SPSA) recently described by Miranda and del Castillo (2011). Furthermore, note that the variability of the estimators is much smaller for the mean than it is for the variance, e.g., under the normality assumption $\text{var}(\bar{w}) = \sigma^2/n_\mathbf{e}$ and $\text{var}(s^2) = 2(n_\mathbf{e} - 1)\sigma^4/n_\mathbf{e}^2$; this problem is also studied by Koch et al. (1998).

In the second approach, referred to as two-layer Kriging metamodeling (2L-KM), we select a relatively small number of input combinations for the simulation model, say, n, using a space-filling design for the $k + c$ input factors (k and c still denote the number of decision and environmental factors, respectively; refer to the first paragraph of §2); i.e., the environmental factors are not yet sampled from their distribution. Then, we use the I/O simulation data to fit a Kriging metamodel for the output w using (5). Finally, for the larger design with N combinations, we use a space-filling design for the decision factors but LHS for the environmental factors accounting for their distribution. We do not simulate the N combinations of this large design, but we do compute the Kriging predictors for the output; then we derive the conditional means and standard deviations using equations analogue to (9) and (10); i.e., in the right-hand sides of (9) and (10), we replace $n_\mathbf{e}$ and $n_\mathbf{d}$ by $N_\mathbf{e}$ and $N_\mathbf{d}$ (the large-sample analogues of the small-sample $n_\mathbf{e}$ and $n_\mathbf{d}$) and w by $\hat{y}$, where $\hat{y}$ denotes the Kriging predictor computed through (5). We use these predictions to fit two Kriging metamodels for the mean and standard deviation of the output.

We shall further explain both approaches through some EOQ examples in the next section. Our methodology assumes that in practice the simulation model

is expensive, although we shall illustrate the two approaches through this inexpensive EOQ simulation model.

Both approaches use their estimated Kriging metamodels for the mean and standard deviation to estimate the robust optimum that minimizes the mean while satisfying a constraint on the standard deviation, according to (1). Varying the value of the right-hand side for that constraint gives the estimated Pareto frontier (see the last paragraph of §3).

This Pareto frontier is built on estimates of the mean and standard deviation of the output. Obviously, estimates are random, so we wish to quantify the variability in the estimated mean and standard deviation. Bootstrapping is a versatile statistical technique; see Efron and Tibshirani (1993) and the many references in Kleijnen (2008, pp. 84–87). Special bootstrap procedures are currently available in many statistical software packages, including the `BOOT` macro in SAS and the `bootstrap` command in S-Plus (see Novikov and Oberman 2007). Nevertheless, we easily implemented these procedures in MATLAB, which we have already used to process the data so far. Whereas Dellino et al. (2010) apply parametric bootstrapping, we apply *nonparametric* or *distribution-free bootstrapping*; i.e., we resample—with replacement—the original simulation observations (which may have any kind of distribution, not necessarily Gaussian). Moreover, bootstrapping (both parametric and nonparametric) assumes that the original observations are identically and independently distributed. Because we cross the design for the decision variables and the environmental variables in our robust optimization procedure, the $n_\mathbf{d}$ observations on the output for a given combination of the environmental factors are not independent (this dependence may be compared with the dependence created by the use of common random numbers in stochastic simulation, which is investigated by Chen et al. 2010). We therefore resample the $n_\mathbf{e}$ *vectors* $\mathbf{w}_j$ $(j = 1, \ldots, n_\mathbf{e})$—with replacement. This resampling gives the $n_\mathbf{e}$ bootstrapped observations $\mathbf{w}_j^* = (w_{1,j}^*, \ldots, w_{n_\mathbf{d},j}^*)$; the asterisk is the usual symbol for bootstrapped values. (Simar and Wilson 1998 also use distribution-free bootstrapping, albeit in the context of data envelopment analysis instead of Pareto frontiers.) Analogously to (9) and (10), we estimate the $n_\mathbf{d}$ bootstrapped conditional means and variances as follows:

$$\bar{w}_i^* = \frac{\sum_{j=1}^{n_\mathbf{e}} w_{i,j}^*}{n_\mathbf{e}} \quad (i = 1, \ldots, n_\mathbf{d}), \qquad (11)$$

$$s_i^{2*} = \frac{\sum_{j=1}^{n_\mathbf{e}} (w_{i,j}^* - \bar{w}_i^*)^2}{n_\mathbf{e} - 1} \quad (i = 1, \ldots, n_\mathbf{d}). \qquad (12)$$

We apply Kriging to the estimates computed through (11) and (12), respectively.

To reduce the sampling error when bootstrapping, we repeat this sampling, say, B times; B is called the bootstrap sample size. This sample size gives the B bootstrapped conditional averages and variances $\bar{w}_{i;b}^*$ and $s_{i;b}^{2*}$ $(b = 1, \ldots, B)$; see (11) and (12). Next, we apply Kriging to $\bar{w}_{i;b}^*$ and $s_{i;b}^{2*}$. For each optimal solution $\hat{\mathbf{d}}^+$ belonging to the "original" (i.e., nonbootstrapped) Pareto frontier, we compute the predictions for the average and standard deviations of the output over the B bootstrapped metamodels, obtaining, say, $\bar{w}_{i;b}^{+*}$ and $s_{i;b}^{+*}$. The B bootstrap observations enable us to compute a *confidence region* for the mean and standard deviation of the output; i.e., we obtain simultaneous confidence intervals—called a confidence region—for these two outputs. These confidence intervals allow us to account for management's risk attitude associated with the threshold value. More specifically, we compute the following distribution-free bootstrapped confidence interval (see Efron and Tibshirani 1993, pp. 170–174) as follows:

$$\left[\hat{\bar{y}}_{(\lfloor B(\alpha/2)/2 \rfloor)}^{+*}, \, \hat{\bar{y}}_{(\lceil B(1-(\alpha/2))/2 \rceil)}^{+*} \right], \qquad (13)$$

where $\hat{\bar{y}}_{(\cdot)}^{+*}$ denotes the bootstrapped mean output that the Kriging model predicts for the estimated Pareto-optimal decision variable $\hat{\mathbf{d}}^+$, the subscript ( ) denotes the order statistic (i.e., the B bootstrapped observations are sorted from smallest to largest), $\lfloor \rfloor$ denotes the floor function (which gives the integer part), $\lceil \rceil$ denotes the ceiling function (rounding upwards), and $\alpha/2$ gives a two-sided confidence interval; Bonferroni's inequality implies that the Type I error rate for the interval per output is divided by the number of outputs (which is two, namely, the mean and standard deviation).

The following confidence interval for the standard deviation of the output is the analogue of (13):

$$\left[\hat{s}_{(\lfloor B(\alpha/2)/2 \rfloor)}^{+*}, \, \hat{s}_{(\lceil B(1-(\alpha/2))/2 \rceil)}^{+*} \right]. \qquad (14)$$

We shall further detail our procedure through some EOQ examples in the next section.

5. EOQ Inventory Simulation

Like Dellino et al. (2010), we apply our methodology to the classic EOQ inventory model for which Zipkin (2000, pp. 30–39) uses the following symbols and assumptions: (i) The demand rate, say, a units per time unit, is known and constant. (ii) The order quantity is Q units. (iii) Total costs consist of setup cost per order K, cost per unit purchased or produced c, and holding cost per inventory unit per time unit h. Management's goal is to minimize the total costs per time unit C over an infinite time horizon.

It is easy to derive that this problem has the following *true I/O function*, which we shall use to check our simulation results:

$$C = \frac{aK}{Q} + ac + \frac{hQ}{2}. \tag{15}$$

This function implies that the EOQ is

$$Q_o = \sqrt{\frac{2aK}{h}}, \tag{16}$$

and the corresponding minimal cost is

$$C_o = C(Q_o) = \sqrt{2aKh} + ac. \tag{17}$$

In our example we use the parameter values of an example from a classic operations research textbook (Hillier and Lieberman 2001, pp. 936–937 and 942–943): $a = 8{,}000$, $K = 12{,}000$, $c = 10$, and $h = 0.3$. Substituting these values into (16) and (17) gives $Q_o = 25{,}298$ and $C_o = 87{,}589$.

To study robust simulation optimization, we follow Dellino et al. (2010) and consider a variant of the classic EOQ model that assumes an unknown demand rate. The robustness of the EOQ model is also examined by Yu (1997), who uses other criteria and methods (e.g., he uses two minmax criteria and analytical methods instead of simulation). Borgonovo (2010, p. 127) gives references to the literature on EOQ "in the presence of imprecision in the parameter estimates" (and then proceeds to sensitivity analysis, not robustness analysis). First, however, we consider classic optimization.

5.1. Classic Simulation Optimization of the EOQ Model

In this subsection we use classic optimization; i.e., we ignore the uncertainty of the environment. As in Dellino et al. (2010), we use the following steps in our simulation experiment.

We select an experimental area for Q, namely, the interval $[15{,}000; 45{,}000]$ (remember that EOQ equals $25{,}298$). Furthermore, we pick five equally spaced points in this interval, including the extreme points, $15{,}000$ and $45{,}000$; see the first row in Table 1. Running the simulation model with these five input values gives the total costs $C(Q_i) = C_i$ $(i = 1, \ldots, 5)$; see the second row in Table 1. Based on these I/O data, we estimate the *Kriging* metamodel; see Figure 2, which also displays the true I/O function derived
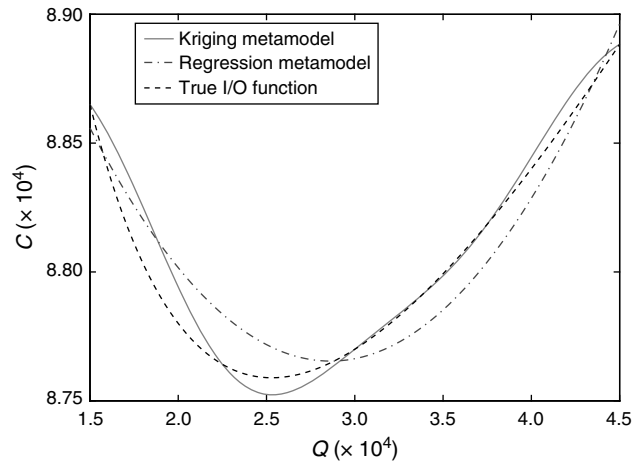

Figure 2 Kriging and RSM Metamodels and the True I/O Function of the Classic EOQ Model

through (15) and the second-degree polynomial RSM metamodel used by Dellino et al. (2010).

To validate this Kriging metamodel, we use *cross-validation*, as Dellino et al. (2010) do to validate their RSM metamodel. This gives Figure 3, which shows the scatterplots for the Kriging and RSM metamodels. Because scatterplots may use scales that are misleading, Table 2 gives the relative prediction errors $\hat{y}_{-i}/C_i$, where the subscript $-i$ means that I/O combination i is eliminated in the cross-validation of the metamodel. This validation shows that in this example Kriging does not give a better approximation than does the second-order polynomial. Our explanation uses the Taylor-series argument; i.e., the EOQ model has a simple, smooth I/O function that is well approximated by a second-order polynomial in our relatively small experimental area. However, van Beers and Kleijnen (2003) give examples in which Kriging does give better predictions than do regression metamodels. Moreover, the next paragraph will give estimated optimal order quantities and corresponding costs that
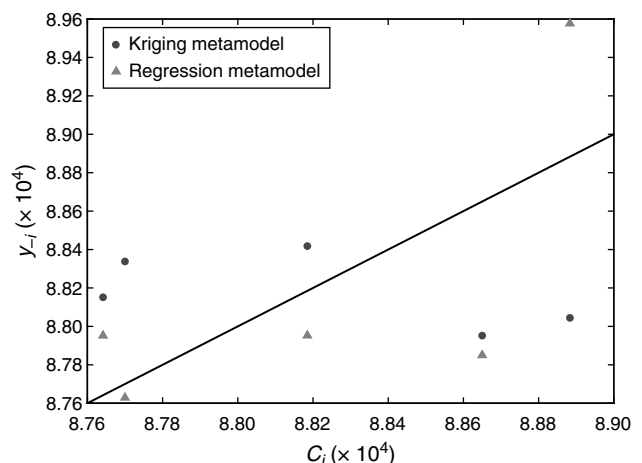
Table 1 I/O Data of the Classic EOQ Simulation

i	1	2	3	4	5
Q_i	15,000	22,500	30,000	37,500	45,000
C_i	88,650.00	87,641.66	87,700.00	88,185.00	88,883.34

Figure 3 Scatterplots for the Cross-Validation of the Kriging and RSM Metamodels of the Classic EOQ Model

Table 2   Cross-Validation of Kriging and RSM Metamodels for EOQ Cost

	Kriging		Regression	
i	$\hat{y}_{-i}$	$\hat{y}_{-i}/C_i$	$\hat{y}_{-i}$	$\hat{y}_{-i}/C_i$
1	87,951.83	0.9921	87,849.94	0.9910
2	88,151.34	1.0058	87,952.11	1.0035
3	88,337.88	1.0073	87,628.92	0.9992
4	88,417.91	1.0026	87,951.95	0.9974
5	88,044.17	0.9906	87,576.98	1.0078

are closer to the true optimal values when using Kriging instead of a second-order polynomial. (Of course, the true optimum is known only for simple academic models such as the EOQ model; cross-validation can be applied to any metamodel.)

To compute the *estimated optimum*, say, $\hat{Q}_o$, we apply MATLAB's function fmincon to the Kriging metamodel for the EOQ cost (we could have used some other solver, including an external global optimizer, instead of a local optimizer such as fmincon). This gives $\hat{Q}_o = 25{,}337.31$ and the estimated minimal cost $\hat{C}_o = 87{,}523.30$. To verify these estimated optimal values, we use the true values following from (16) and (17). This gives $\hat{Q}_o/Q_o = 25{,}337.31/25{,}298 = 1.0016$ and $\hat{C}_o/C_o = 87{,}523.30/87{,}589 = 0.9992$, so we conclude that the estimated optimal cost and order quantity virtually equal the true optimal values. For the second-order polynomial metamodel, Dellino et al. (2010) give the following results: $\hat{Q}_o = 28{,}636$ and $\hat{C}_o = 887{,}654$; thus $\hat{Q}_o/Q_o = 1.13$ and $\hat{C}_o/C_o = 1.001$, which are slightly less accurate estimated optimal values compared with our Kriging estimates.

Note that we also experiment with a *smaller* experimental area, i.e., a smaller Q interval. This interval gives an even more accurate Kriging metamodel; the resulting estimated optimum is only 0.32% above the true EOQ, and the corresponding cost virtually equals the true cost.

5.2. Robust Optimization with Uncertain Demand Rate

Now we follow Dellino et al. (2010), and assume that a (demand per time unit) is an *unknown* constant; i.e., a has a Normal distribution with mean μ_a and standard deviation σ_a: $a \sim N(\mu_a, \sigma_a)$. In our experiments we assume that $\mu_a = 8{,}000$ (the "base" value used in §5.1) and $\sigma_a = 0.10\mu_a$ (uncertainty about the true input parameter). This standard deviation can give negative values for a, so we resample until we get nonnegative values only; this adjustment of the Normal distribution is ignored in our further analysis. We apply the two Kriging approaches sketched in §4.2.

5.2.1. 1L-KM Approach: Kriging Models for Mean and Standard Deviation Estimated from Simulation I/O Data. To select a suitable sample size for the

environmental factor a in our simulation, we choose $n_a = 100$, which gives quite accurate estimates of the sample mean and sample standard deviation of the simulation output (moreover, $n_a = 100$ gives "enough" data to bootstrap later on). For the environmental factors design, we use LHS; it splits the range of possible a values ($0 < a < \infty$) into n_a equally likely subranges. We use lhsnorm from the MATLAB Statistics Toolbox (Mathworks Inc. 2005) to select these values from $N(\mu_a, \sigma_a)$. For the decision variable Q, we select $n_Q = 10$ equally spaced values within $[15{,}000; 45{,}000]$, which is the range selected in §5.1. We cross these two designs for a and Q, respectively, which gives 100×10 combinations of the two factors. Running the simulation model for these 1,000 input combinations, together with (9) and (10), gives the estimated mean and standard deviation of the cost C conditional on the n_Q values of Q:

$$\bar{C}_i = \frac{\sum_{j=1}^{n_a} C_{i,j}}{n_a} \quad (i = 1, \ldots, n_Q), \qquad (18)$$

$$s_i = \left[\frac{\sum_{j=1}^{n_a} (C_{i,j} - \bar{C}_i)^2}{n_a - 1} \right]^{1/2} \quad (i = 1, \ldots, n_Q). \qquad (19)$$

The latter estimator is biased because $E(\sqrt{s^2}) = E(s) \neq \sqrt{E(s^2)} = \sqrt{\sigma^2} = \sigma$; we ignore this bias.

Using (18) and (19), we fit one Kriging metamodel for the estimated mean cost—which gives $\hat{\bar{C}}$—and one Kriging metamodel for the estimated standard deviation of cost—which gives $\hat{s}$. Obviously, each of these metamodels is based on $n_Q = 10$ observations (in the terminology of §4.1, there are 10 "old" observations). The two Kriging metamodels are shown in Figures 4 and 5, which also display the true cost function

$$E(\bar{C}) = \left(\frac{K}{Q} + c \right) \mu_a + \frac{hQ}{2} \qquad (20)$$

and the true standard deviation function

$$\sigma_C = c\sigma_a + \frac{K\sigma_a}{Q}, \qquad (21)$$

which are easy to derive from (15). Figure 4 resembles the classic EOQ graph (see the dashed curve in Figure 2), which assumes a known demand rate. Figure 5 illustrates that the standard deviation decreases as the order quantity increases; i.e., the increased order quantity provides a buffer against unexpected variations in the demand rate.

We validate these two metamodels—without using the true I/O functions (20) and (21)—through leave-one-out cross-validation. This validation gives the scatterplots in Figures 6 and 7. Given these two figures, we accept the two Kriging metamodels.

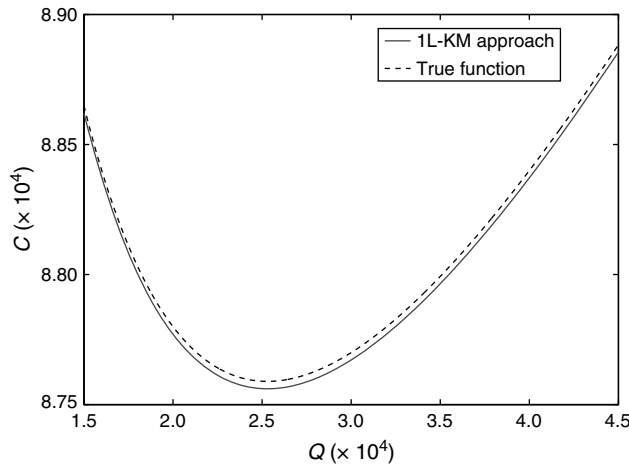Based on these Kriging metamodels, we try to find the order quantity that minimizes the mean cost while

Figure 4 Kriging Metamodel for Mean Cost in the 1L-KM Approach and the True Mean Cost

Figure 6 Scatterplot of Kriging Metamodel for Mean Cost in the 1L-KM Approach

ensuring that the standard deviation does not exceed the given threshold T. We again solve this constrained optimization problem through MATLAB's function fmincon. Next we vary this threshold and find the set of optimal solutions that estimates the Pareto frontier; see Figure 8, which also shows the true Pareto frontier derived from (20) and (21).

Note that we select a range for the threshold T that differs from the range in Dellino et al. (2010). Selecting the same range would have resulted in an unconstrained optimization problem, so the Pareto frontier would have been a single point.

5.2.2. 2L-KM Approach: Kriging Models for Mean and Standard Deviation Estimated from Intermediate Kriging Predictions. We select the same number of input combinations for the simulation model as we did for the 1L-KM approach; i.e., we select $n_a \times n_Q = 100 \times 10 = 1{,}000$ input combinations. To select these 1,000 values, we use a space-filling design in these two factors (in the 1L-KM approach, we use a

space-filling design only for Q). To avoid extrapolation when using the metamodel, we select $\max a_j = \mu_a + 3\sigma_a$ and $\min a_j = \max(\mu_a - 3\sigma_a, \epsilon)$, where ϵ is a small positive number. After running the simulation for these 1,000 input combinations, the resulting I/O data give a Kriging metamodel for the cost $\hat{C}$ as a function of the demand rate a and the order quantity Q. This metamodel is used in the following procedure:

Step 1. Use LHS to sample $N_a \gg n_a$ values from the distribution of the environmental variable a, and use a space filling design to select $N_Q \gg n_Q$ values for the decision variable Q. We select $N_a = 200$ and $N_Q = 30$. Note that exceeding the upper bound for a is rare; if this event occurs, we simply take a new sample.

Step 2. Combine the values of Step 1 into $N_a \times N_Q$ input combinations.

Step 3. Compute the Kriging predictions $\hat{C}_{i,j}$ ($i = 1, \ldots, N_Q$; $j = 1, \ldots, N_a$) for the combinations of Step 2 using the Kriging metamodel estimated from
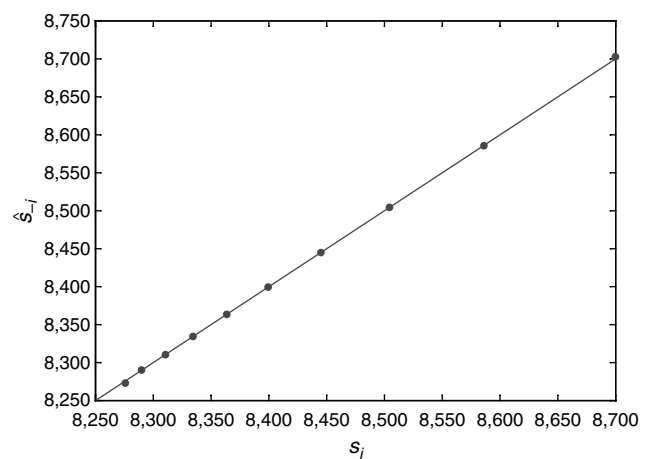

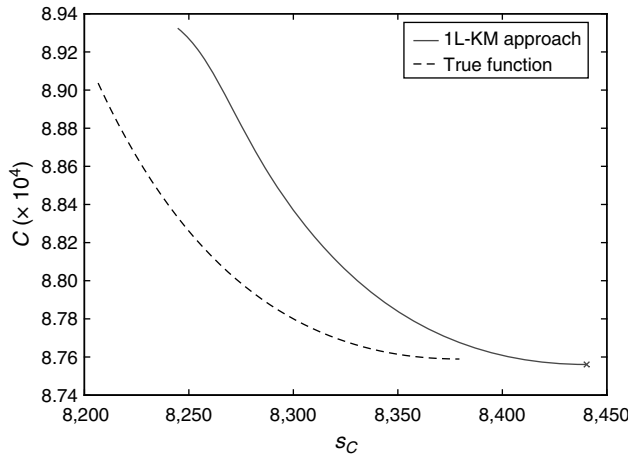**Figure 5** Kriging Metamodel for the Cost's Standard Deviation in the 1L-KM Approach and the True Standard Deviation

Figure 7 Scatterplot of Kriging Metamodel for Cost's Standard Deviation in the 1L-KM Approach

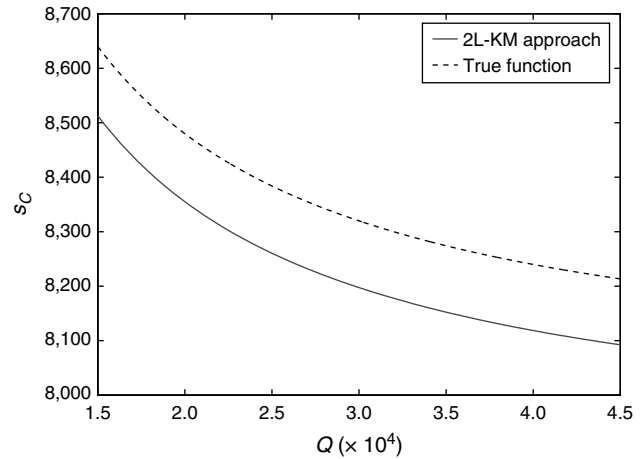**Figure 8** **Estimated Pareto Frontier in the 1L-KM Approach and True Pareto Frontier for EOQ Example**

the smaller experiment with the simulation model with $n_a \times n_Q$ input combinations.

Step 4. Use these predictions $\hat{C}_{i,j}$ to estimate the N_Q conditional means and standard deviations of the cost C, which are analogous to (18) and (19) but use metamodel predictions instead of simulated values as follows:

$$\hat{C}_i = \frac{\sum_{j=1}^{N_a} \hat{C}_{i,j}}{N_a} \quad (i = 1, \dots, N_Q), \tag{22}$$

$$\hat{\sigma}_i = \frac{\sum_{j=1}^{N_a} (\hat{C}_{i,j} - \hat{C}_i)^2}{N_a - 1} \quad (i = 1, \dots, N_Q). \tag{23}$$

Step 5. Fit one Kriging metamodel to the N_Q estimated means resulting from (22), and fit another Kriging metamodel to the N_Q estimated standard deviations resulting from (23).

Figures 9 and 10 display the two Kriging models that result from Step 5 and the true functions.
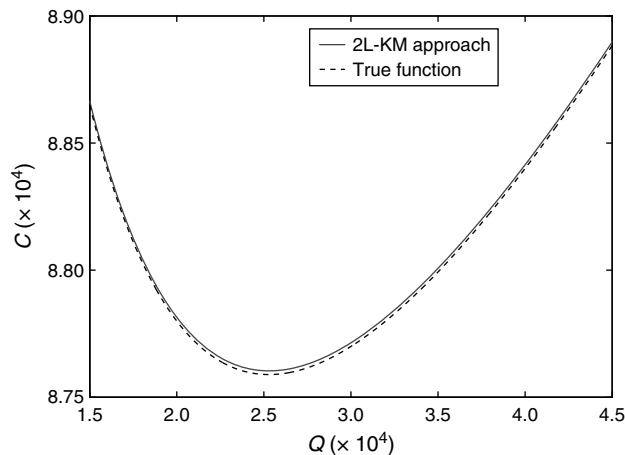

Figure 9 **Kriging Metamodel for Mean Costs Based on Kriging Predictions in the 2L-KM Approach, and the True Mean Cost**
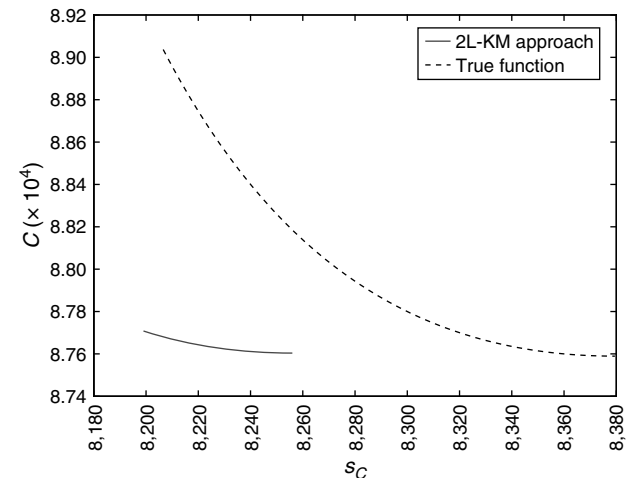

Figure 10 **Kriging Metamodel for the Cost's Standard Deviation Based on Kriging Predictions in the 2L-KM Approach, and the True Function**

Cross-validation gives scatterplots for these two metamodels, which look very good: all points are near the 45° line, so we do not display these two figures but refer to Figures 6 and 7 (which give the scatterplots for the 1L-KM approach). Cross-validation also gives values such as $\hat{C}_{(-1)}/\hat{C}_1 = 0.999999332$ and $\hat{s}_{(-30)}/\hat{\sigma}_{30} = 1.0000002496$, which imply very small relative prediction errors. Thus we accept these two Kriging metamodels as adequate approximations.

We solve the constrained optimization problem, again using fmincon. Next we vary the threshold T, albeit over a range that differs from the previous range, to get interesting results—namely, neither unconstrained nor infeasible results. The resulting Pareto frontier and the true frontier are displayed in Figure 11.

A first and intuitive way to compare the two Kriging approaches uses their relative costs and benefits. A benefit may be the accuracy of the approach;

Figure 11 **Estimated Pareto Frontier in the 2L-KM Approach and True Pareto Frontier for EOQ Example**

e.g., how close the estimated frontier is to the true frontier derived from (20) and (21). A comparison of Figures 8 and 11 shows that the accuracy on the Pareto frontier provided by the 1L-KM approach is higher than that of the 2L-KM approach. A cost component is the computation time needed by the two approaches: because they use two designs of the same size, their computational costs are roughly the same, assuming that the time required for fitting the metamodels and deriving predictions from them is negligible compared with that of running the simulation model. Assessing the advantages of one robust simulation optimization approach over the other is not straightforward. There are several issues mainly related to possible variability in the metamodels, which could yield misleading interpretations of the obtained Pareto-optimal solutions. Therefore, in the next subsection, we compare the two approaches through their confidence regions for the mean and standard deviation.

5.2.3. Bootstrapped Confidence Regions. As we discussed in §§5.2.1 and 5.2.2, we estimated the Pareto frontier through *random* simulation outputs $C_{i,j}$ (whether we use the 1L-KM or the 2L-KM approach) as a consequence of the uncertainty in the demand rate. Therefore, it becomes important to perform further analyses of the statistical variation of this frontier to better understand the performance of the two approaches proposed. To estimate this variability, we use *distribution-free bootstrapping*, described in (11) and (12). This bootstrap gives $\bar{C}_b^*$ and s_b^* ($b = 1, \ldots, B$), respectively, which gives bootstrapped Kriging metamodels. The results for 1L-KM are given in Figure 12, where the vertical line will be explained below.

Figure 12 shows that the bootstrapped curves envelop both the original and true curves. Dellino et al. (2010) use bootstrapping to derive a bundle of estimated Pareto curves, but we think that the following analysis is more relevant. In particular, we focus on the original (non-bootstrapped) Pareto frontier and study the variability of the optimal solutions $\hat{Q}^+$ using the bootstrapped pairs of metamodels. Such an analysis considers the implications of a decision (in terms of $\hat{Q}^+$) based on this Pareto frontier as a result of uncertainty. In fact, given the original Pareto frontier, management selects its preferred combination of the mean and standard deviation of the inventory cost; e.g., $\hat{\bar{C}} = 87,560.84$ and $\hat{s} = 8,440.33$, which corresponds with the "cross" in Figure 8 (near the end of the Pareto frontier estimated through 1L-KM). Making the standard deviation not exceed its threshold (namely, $T = 8,600$) implies a specific order quantity—namely, $\hat{Q}^+ = 25,304.19$; see the "cross" in Figure 5. This order quantity may actually give a mean and a standard deviation that differ from the
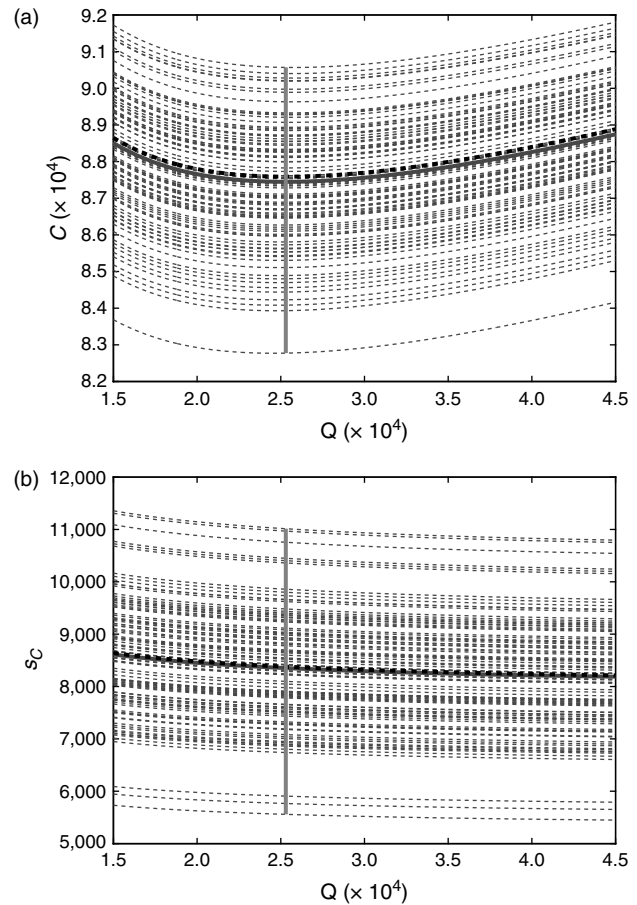

Figure 12 **(a) Bootstrapped Estimated Costs, Original Metamodel in the 1L-KM (Heavy Curve) and True Cost (Dotted Curve); (b) Bootstrapped Estimated Standard Deviations of Cost, Original Metamodel in the 1L-KM (Heavy Curve) and True Standard Deviation (Dotted Curve)**

"original" values in Figure 12; see the vertical lines in Figure 12, panels (a) and (b), at $Q = \hat{Q}^+ = 25,304.19$.

The solution $\hat{Q}^+$ selected by the management is subject to variability, and we need to take this into account when using the Pareto frontier as a "decisional tool"; we adopt bootstrapping to study this variability. Thus this estimated Pareto-optimal order quantity $\hat{Q}^+$ corresponds with B bootstrapped values for the mean and standard deviation, respectively. From these B values, we estimate a *confidence region* for the mean and standard deviation of the cost given $\hat{Q}^+$, applying (13) and (14). Figure 13 displays the rectangular *confidence regions* for two points on the original estimated Pareto curve. Figure 13(a) corresponds with the relatively small threshold value $T = 8,300$, so $\hat{Q}^+ = 40,495$ for the 1L-KM approach and $\hat{Q}^+ = 28,035.38$ for the 2L-KM approach. Figure 13(b) corresponds with the larger threshold value $T = 8,600$, so we obtain the smaller $\hat{Q}^+ = 25,304.19$ for the 1L-KM approach and $\hat{Q}^+ = 25,727.24$ for the 2L-KM approach. These two threshold values may reflect
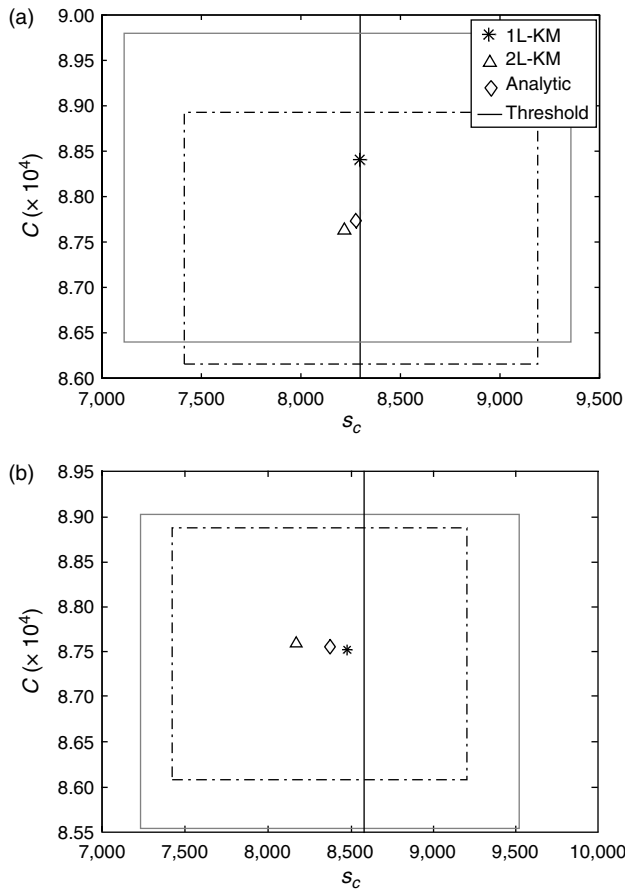
Figure 13 **Confidence Regions for σ_C (on x Axis) and $E(C)$ (on y Axis) Based on Bootstrapped Kriging in 1L-KM (Solid Rectangle) and 2L-KM (Dash-Dotted Rectangle) at (a) $T = 8,300$ and (b) $T = 8,600$ (See the Vertical Line)**

Notes. $*$ and $\triangle$ denote the "true" solutions based on (20) and (21), in 1L-KM and 2L-KM, respectively: (a) $(\hat{\bar{C}}_{1L\text{-}KM} = 88,416.55, \hat{s}_{1L\text{-}KM} = 8,296.97)$ at $\hat{Q}^+_{1L\text{-}KM} = 40,495$, and $(\hat{\bar{C}}_{2L\text{-}KM} = 87,643.98, \hat{s}_{2L\text{-}KM} = 8,219.55)$ at $\hat{Q}^+_{2L\text{-}KM} = 28,035.38$; and (b) $(\hat{\bar{C}}_{1L\text{-}KM} = 87,560.84, \hat{s}_{1L\text{-}KM} = 8,440.33)$ at $\hat{Q}^+_{1L\text{-}KM} = 25,304.19$, and $(\hat{\bar{C}}_{2L\text{-}KM} = 87,605.02, \hat{s}_{2L\text{-}KM} = 8,249.82)$ at $\hat{Q}^+_{2L\text{-}KM} = 25,727.24$.

risk-averse and *risk-seeking* management. In this example, 2L-KM gives a smaller confidence region and still covers the true point: the Kriging metamodel estimated from relatively few (possibly expensive) simulation observations is so accurate that the relatively large sample based on this metamodel gives accurate predictions. We further notice that the high accuracy in the metamodel of total cost obtained in the first layer of the 2L-KM approach implies that predictions computed from this metamodel can be treated as if they were simulated data. From this point on, the two approaches perform the same computations. Therefore, the 2L-KM approach can provide better results than can the 1L-KM approach at the same computational costs: in fact, we derive the metamodels of mean and standard deviation in 2L-KM using an arbitrarily large set of design points without increasing

the computational costs as would be required by 1L-KM, whose accuracy can be increased by running additional expensive simulations.

The confidence interval for the standard deviation shows that the original estimated Pareto-optimal order quantity may still give a standard deviation that violates the threshold because of the metamodels' variability (see Figure 12(b)). In fact, we could also use the confidence region derived to estimate the degree of infeasibility of a solution (i.e., the probability that an optimal solution obtained for a given T becomes infeasible because of uncertainty) as the fraction of the overall area that is beyond (i.e., to the right of) the threshold value T used in the optimization problem. Confronted with this possibility, management may want to change the order quantity such that the probability of violating the threshold becomes acceptable; e.g., management may switch from the relatively small threshold (and corresponding high Q) in Figure 13(a) to the higher threshold (and smaller Q) in Figure 13(b). The formalization of the problem of choosing among random outputs is beyond this paper—a classic reference is Keeney and Raiffa (1976). We also refer to the first future-research issue described in §6.

5.3. Robust Optimization with Uncertain Demand Rate and Cost Coefficients

Inspired by Borgonovo and Peccati (2007), we extend our robust formulation of the EOQ model such that it accounts for fixed but uncertain cost parameters. Thus the number of environmental factors increases from one (demand rate) to three (holding and setup costs, as well as demand rate). Both Kriging approaches give confidence regions that cover the true point, but the 2L-KM approach again gives a smaller confidence region. We provide further details and computational results in the Online Supplement.

6. Conclusions and Future Research

Robust optimization of simulated systems may use Taguchi's worldview, which distinguishes between decision variables that may be optimized and environmental variables that remain uncertain but should be taken into account when optimizing. Taguchi's statistical techniques, however, may be replaced by Kriging metamodels (instead of low-order polynomials) and their space-filling designs such as LHS (instead of Taguchian or RSM designs). We developed two Kriging variants—namely, the 1L-KM approach that fits one Kriging metamodel for the mean and one for the standard deviation estimated from the simulation data, and the 2L-KM approach that fits a single Kriging metamodel to a relatively small simulation sample and uses this metamodel to compute Kriging predictions for a larger sample. In the EOQ examples,

2L-KM gives more accurate predictions than 1L-KM, given the same computational cost measured by the number of expensive simulation runs. Moreover, we point out that the only way to improve the metamodels' accuracy for 1L-KM is to run additional expensive simulations, which increases the computational cost of this approach. On the contrary, we can arbitrarily increase the accuracy of the metamodels derived through 2L-KM with negligible costs because we do not need to use simulations anymore; however, this is possible only if the metamodel obtained in the first layer is sufficiently accurate to provide "reliable" predictions. The whole Kriging-based methodology for robust optimization may be further enhanced by distribution-free bootstrapping, which quantifies the variability of the estimated robust optimization results so that it may help management make the final compromise decision. The confidence regions derived through bootstrapping for both approaches confirm that the 2L-KM approach performs better than does the 1L-KM approach in the EOQ examples. Besides the EOQ examples, we tested our methodology on another example, which is discussed in detail in the Online Supplement; this example includes more decision and environmental factors and also accounts for implementation errors. The results confirm the behavior of the two proposed approaches.

Future research may address the following issues. Instead of minimizing the mean under a variance constraint, we may minimize a specific quantile of the simulation output (see Batur and Choobineh 2010, Bekki et al. 2009, Kleijnen et al. 2011) or minimize the conditional value at risk (CVaR) (see Chen et al. 2009, Dehlendorff et al. 2010, García-González et al. 2007, Natarajan et al. 2009); the mean–variance trade-off is also criticized by Yin et al. (2009). Besides CVaR, other risk measures are surveyed by Sordo (2009), including the "expected shortfall at level p," which is popular in the actuarial literature. Our Kriging-based method may be compared with alternative approaches based on either different metamodeling techniques (e.g., the so-called universal Kriging discussed in the Kriging literature and applied in classic and robust optimization in engineering, RSM (low-order polynomial linear regression) models proposed by Dellino et al. (2010) for robust optimization, and generalized linear models proposed by Lee and Nelder 2003) or different optimization solvers (e.g., evolutionary algorithms proposed by Branke et al. 2001 for robust optimization of noisy systems and by Jin and Sendhoff 2003 for robust optimization with a multiobjective formulation). Investigating the performance of alternative DOEs might also be of interest; see Dehlendorff et al. (2011). Based on the promising results obtained from the bootstrap procedure, we plan to further investigate

and refine this analysis. In our next study, we adjust our methodology to accommodate *random* simulation models, which imply aleatory uncertainty as well as epistemic uncertainty; these two types of uncertainty are discussed by de Rocquigny et al. (2008) and Helton (2009). Examples of random simulation are (s, S) models, with either explicit out-of-stock costs resulting in scalar output or a service constraint resulting in vector output (the difference $S - s$ in these models is often based on the EOQ model discussed in this paper). Finally, we plan to apply our methodology to complex *supply chain* models; see also Cannella and Ciancimino (2010), Narasimhan and Talluri (2009), Rao and Goldsby (2009), and Shukla et al. (2010).

Electronic Companion

An electronic companion to this paper is available as part of the online version at http://dx.doi.org/10.1287/ijoc.1110.0465.

Acknowledgments

References

Ankenman, B., B. L. Nelson, J. Staum. 2010. Stochastic kriging for simulation metamodeling. *Oper. Res.* **58**(2) 371–382.

Barton, R. R., M. Meckesheimer. 2006. Metamodel-based simulation optimization. S. G. Henderson, B. L. Nelson, eds. *Handbooks in Operations Research and Management Science*, Vol. 13. Elsevier, Amsterdam, 535–574.

Batur, D., F. Choobineh. 2010. A quantile-based approach to system selection. *Eur. J. Oper. Res.* **202**(3) 764–772.

Bekki, J. M., J. W. Fowler, G. T. Mackulak, M. Kulahci. 2009. Simulation-based cycle-time quantile estimation in manufacturing settings employing non-FIFO dispatching policies. *J. Simulation* **3**(2) 69–83.

Bertsimas, D., O. Nohadani, K. M. Teo. 2010. Robust optimization for unconstrained simulation-based problems. *Oper. Res.* **58**(1) 161–178.

Beyer, H.-G., B. Sendhoff. 2007. Robust optimization—A comprehensive survey. *Comput. Methods Appl. Mech. Engrg.* **196**(33–34) 3190–3218.

Borgonovo, E. 2010. Sensitivity analysis with finite changes: An application to modified EOQ models. *Eur. J. Oper. Res.* **200**(1) 127–138.

Borgonovo, E., L. Peccati. 2007. Global sensitivity analysis in inventory management. *Internat. J. Production Econom.* **108**(1–2) 302–313.

Branke, J., C. Schmidt, H. Schmeck. 2001. Efficient fitness estimation in noisy environments. L. Spector, E. D. Goodman, A. Wu, W. B. Langdon, H.-M. Voight, M. Gen, S. Sen et al., eds. *Proc. Genetic Evol. Comput.*, Morgan Kaufmann, San Francisco, 243–250.

Cannella, S., E. Ciancimino. 2010. Up-to-date supply chain management: The coordinated (S, R) order-up-to. W. Dangelmaier, A. Blecken, R. Delius, S. Klöpfer, eds. *Adv. Manufacturing Sustainable Logist.: 8th Internat. Heinz Nixdorf Sympos.* Springer-Verlag, Berlin, 175–185.

Chen, X., B. Ankenman, B. L. Nelson. 2010. The effects of common random numbers on stochastic Kriging metamodels. *Indust. Engrg.* **1** 1–31.

Chen, Y., M. Xu, Z. G. Zhang. 2009. A risk-averse newsvendor model under CVaR criterion. *Oper. Res.* **57**(4) 1040–1044.

Cressie, N. A. C. 1993. *Statistics for Spatial Data*, revised ed. John Wiley & Sons, New York.

Dehlendorff, C., M. Kulahci, K. Andersen. 2011. Designing simulation experiments with controllable and uncontrollable factors for applications in healthcare. *J. Roy. Statist. Soc. Ser. C* **60**(1) 31–49.

Dehlendorff, C., M. Kulahci, S. Merser, K. Andersen. 2010. Conditional value at risk as a measure for waiting time in simulations of hospital units. *Quality Tech. Quant. Management* **7**(3) 321–336.

del Castillo, E. 2007. *Process Optimization: A Statistical Approach*. Springer, New York.

Dellino, G., J. P. C. Kleijnen, C. Meloni. 2010. Robust optimization in simulation: Taguchi and response surface methodology. *Internat. J. Production Econom.* **125**(1) 52–59.

Dellino, G., P. Lino, C. Meloni, A. Rizzo. 2007. Enhanced evolutionary algorithms for MDO: A control engineering perspective. C. Grosan, A. Abraham, H. Ishibuchi, eds. *Hybrid Evolutionary Algorithms.* Studies in Computational Intelligence, Vol. 75. Springer, Berlin, 41–80.

de Rocquigny, E., N. Devictor, S. Tarantola. 2008. Uncertainty settings and natures of uncertainty. E. de Rocquigny, N. Devictor, S. Tarantola, eds. *Uncertainty in Industrial Practice: A Guide to Quantitative Uncertainty Management.* John Wiley & Sons, Chichester, UK, 199–212.

Efron, B., R. J. Tibshirani. 1993. *An Introduction to the Bootstrap.* Chapman & Hall, New York.

García-González, J., E. Parrilla, A. Mateo. 2007. Risk-averse profit-based optimal scheduling of a hydro-chain in the day-ahead electricity market. *Eur. J. Oper. Res.* **181**(3) 1354–1369.

Helton, J. C. 2009. Conceptual and computational basis for the quantification of margins and uncertainty. Report SAND2009-3055, Sandia National Laboratories, Albuquerque, NM.

Hillier, F. S., G. J. Lieberman. 2001. *Introduction to Operations Research*, 7th ed. McGraw-Hill, New York.

Jin, Y., B. Sendhoff. 2003. Trade-off between performance and robustness: An evolutionary multiobjective approach. C. M. Fonseca, P. J. Fleming, E. Zitzler, K. Deb, L. Thiele, eds. *Evolutionary Multi-Criterion Optimization.* Lecture Notes in Computer Science, Vol. 2632. Springer, Berlin, 237–252.

Keeney, R. L., H. Raiffa. 1976. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs.* John Wiley & Sons, New York.

Kleijnen, J. P. C. 2008. *Design and Analysis of Simulation Experiments.* Springer, New York.

Kleijnen, J. P. C. 2010. Sensitivity analysis of simulation models. J. J. Cochran, L. A. Cox Jr., P. Keskinocak, J. P. Kharoufeh, J. C. Smith, eds. *Wiley Encyclopedia of Operations Research and Management Science.* John Wiley & Sons, New York, 1–10.

Kleijnen, J. P. C., H. Pierreval, J. Zhang. 2011. Methodology for determining the acceptability of given designs in uncertain environments. *Eur. J. Oper. Res.* **209**(2) 176–183.

Kleijnen, J. P. C., S. M. Sanchez, T. W. Lucas, T. M. Cioppa. 2005. State-of-the-art review: A user's guide to the brave new world of designing simulation experiments. *INFORMS J. Comput.* **17**(3) 263–289.

Koch, P. N., D. Mavris, J. K. Allen, F. Mistree. 1998. Modeling noise in approximation-based robust design: A comparison and critical discussion. *ASME Design Engrg. Tech. Conf., Atlanta*, ASME, New York, 1–15.

Lee, Y., J. A. Nelder. 2003. Robust design. *J. Quality Tech.* **35**(1) 2–12.

Lee, K.-H., G.-J. Park. 2006. A global robust optimization using Kriging based approximation model. *J. Japan Soc. Mech. Engrg.* **49**(3) 779–788.

Lehman, J. S., T. J. Santner, W. I. Notz. 2004. Designing computer experiments to determine robust control variables. *Statistica Sinica* **14**(2) 571–590.

Lophaven, S. N., H. B. Nielsen, J. Søndergaard. 2002. DACE: A MATLAB Kriging toolbox (version 2.0). Technical Report IMM-TR-2002-12, Technical University of Denmark, Lyngby.

MathWorks Inc. 2005. *Statistics Toolbox.* User's manual. MathWorks, Natick, MA.

Miettinen, K. M. 1999. *Nonlinear Multiobjective Optimization.* Kluwer Academic Publishers, Boston.

Miranda, A. K., E. del Castillo. 2011. Robust parameter design optimization of simulation experiments using stochastic perturbation methods. *J. Oper. Res. Soc.* **62**(1) 198–205.

Myers, R. H., D. C. Montgomery, C. M. Anderson-Cook. 2009. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 3rd ed. John Wiley & Sons, New York.

Nair, V. N., B. Abraham, J. Mackay, G. Box, R. N. Kacker, T. J. Lorenzen, J. M. Lucas et al. 1992. Taguchi's parameter design: A panel discussion. *Technometrics* **34**(2) 127–161.

Narasimhan, R., S. Talluri. 2009. Perspectives on risk management in supply chains. *J. Oper. Management* **27**(2) 114–118.

Natarajan, K., D. Pachamanova, M. Sim. 2009. Constructing risk measures from uncertainty sets. *Oper. Res.* **57**(5) 1129–1141.

Neumaier, A., O. Shcherbina, W. Huyer, T. Vinkó. 2005. A comparison of complete global optimization solvers. *Math. Programming Ser. B* **103**(2) 335–356.

Novikov, I., B. Oberman. 2007. Optimization of large simulations using statistical software. *Comput. Statist. Data Anal.* **51**(5) 2747–2752.

Oberguggenberger, M., J. King, B. Schmelzer. 2009. Classical and imprecise probability methods for sensitivity analysis in engineering: A case study. *Internat. J. Approximate Reasoning* **50**(4) 680–693.

Park, G.-J., T.-H. Lee, K. H. Lee, K.-H. Hwang. 2006. Robust design: An overview. *AIAA J.* **44**(1) 181–191.

Rao, S., T. J. Goldsby. 2009. Supply chain risks: A review and typology. *Internat. J. Logist. Management* **20**(1) 97–123.

Sacks, J., W. J. Welch, T. J. Mitchell, H. P. Wynn. 1989. Design and analysis of computer experiments. *Statist. Sci.* **4**(4) 409–435.

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

Shukla, S. K., M. K. Tiwari, H.-D. Wan, R. Shankar. 2010. Optimization of the supply chain network: Simulation, Taguchi, and psychoclonal algorithm embedded approach. *Comput. Indust. Engrg.* **58**(1) 29–39.

Simar, L., P. W. Wilson. 1998. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Sci.* **44**(1) 49–61.

Sordo, M. A. 2009. Comparing tail variabilities of risks by means of the excess wealth order. *Insurance: Math. Econom.* **45**(3) 466–469.

Stinstra, E., D. den Hertog. 2008. Robust optimization using computer experiments. *Eur. J. Oper. Res.* **191**(3) 816–837.

Taguchi, G. 1987. *System of Experimental Designs*, Vols. 1 and 2. UNIPUB/Krauss International, White Plains, New York.

van Beers, W. C. M., J. P. C. Kleijnen. 2003. Kriging for interpolation in random simulation. *J. Oper. Res. Soc.* **54**(3) 255–262.

van Beers, W. C. M., J. P. C. Kleijnen. 2004. Kriging interpolation in simulation: A survey. R. G. Ingalls, M. D. Rossetti, J. S. Smith, B. A. Peters, eds. *Proc. 2004 Winter Simulation Conf.*, IEEE, Washington, DC, 113–121.

Wu, C. F. J., M. Hamada. 2000. *Experiments: Planning, Analysis, and Parameter Design Optimization.* John Wiley & Sons, New York.

Wu, J., J. Li, S. Wang, T. C. E. Cheng. 2009. A note on mean-variance analysis of the newsvendor model with stockout cost. *Omega* **37**(3) 724–730.

Yin, Y., S. M. Madanat, X.-Y. Lu. 2009. Robust improvement schemes for road networks under demand uncertainty. *Eur. J. Oper. Res.* **198**(2) 470–479.

Yu, G. 1997. Robust economic order quantity models. *Eur. J. Oper. Res.* **100**(3) 482–493.

Zipkin, P. H. 2000. *Foundations of Inventory Management.* McGraw-Hill, New York.