No. 2009–03

METHODOLOGY FOR DETERMINING THE ACCEPTABILITY OF GIVEN DESIGNS IN UNCERTAIN ENVIRONMENTS

By Jack P.C. Kleijnen, Henri Pierreval, Jin Zhang

January 2009

ISSN 0924-7815

Methodology for determining the acceptability of given designs in uncertain environments

Jack P.C. Kleijnen a), Henri Pierreval b) and Jin Zhang c)

a) Department of Information Management, Tilburg University, Postbox 90153, 5000 LE Tilburg, Netherlands, e-mail: kleijnen@uvt.nl, http://center.uvt.nl/staff/kleijnen/; phone 31-13-466-2029; fax: 31-13-466-3069
b) LIMOS (Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes)UMR CNRS 6158
IFMA BP 265, F-63175 Aubiere Cedex France
+33 (0)4 73 28 81 06 or +33 (0)4 73 28 81 01, email : henri.pierreval@ifma.fr
c) Department of Information Management, Tilburg University, Postbox 90153, 5000 LE Tilburg, Netherlands, email : jin.zhang@ifma.fr

JEL: C0, C1, C9

Version: 7 January 2009

Abstract

Managers wish to verify that a particular engineering design meets their requirements. This design's future environment will differ from the environment assumed during the design. Therefore it is crucial to determine which variations in the environment may make this design unacceptable. The proposed methodology estimates which uncertain environmental parameters are important (so managers can become pro-active) and which parameter combinations (scenarios) make the design unacceptable. The methodology combines simulation, bootstrapping, and metamodeling. The methodology is illustrated through a simulated manufacturing system, including fourteen uncertain parameters of the input distributions for the various arrival and service times. These parameters are investigated through sixteen scenarios, selected through a two-level fractional-factorial design. The resulting simulation Input/Output (I/O) data are analyzed through a first-order polynomial metamodel and bootstrapping. A second experiment gives some outputs that are indeed unacceptable. Polynomials fitted to the I/O data estimate the border line (frontier) between acceptable and unacceptable environments.

Key words: Uncertainty modeling; Risk analysis; Robustness and sensitivity analysis; Simulation; Bootstrap

---

1 Introduction

In the design of complex systems (e.g., manufacturing plants, supply chains, hospitals, harbours), it is common that a team of designers works on defining the specifications of the future system. Once these specifications are provided, managers are often interested in verifying—through a simulation analysis—that these specifications are acceptable with regard to the performance required for this system (e.g., annual production quantities, minimum waiting times in the emergency room of a hospital). In practice, this type of study is common, but the literature neglects the question of the acceptability of a proposed design. Answering this question raises several theoretical issues.

The simulation literature includes many publications on finding the most important design factors (so-called 'sensitivity analysis') and finding the optimal design; see Kleijnen (2008). However, determining whether the proposed design is acceptable is a different type of question. Now the crucial question becomes: will the proposed design meet the decision makers' performance requirements? This question raises methodological issues caused by two types of randomness (also see the general discussion of these types by De Rocquigny et al. (2008, pp. 199-211)):

(1) *Aleatory* uncertainty caused by the inherent randomness of the system (e.g., arrival and processing times in the example summarized in Table 1 below).
(2) *Epistemic* uncertainty caused by the imperfect knowledge of the designers about the statistical distributions that cause the aleatory uncertainty (e.g., the parameter values of the exponential and normal distributions assumed in Table 1 may be unrealistic). In the design stage, many parameters are not known so they may turn out to differ from their assumed values, see Pierreval and Durieux (2007). Therefore, it is prudent to consider these parameter uncertainties and determine under which conditions the given design will still be acceptable.

A related yet different issue is *robustness*. The literature on simulation and especially Mathematical Programming presents methods that account for the uncertainty in the input data when searching for 'optimal' solutions; see Kleijnen (2008) and Dellino, et al. (2008). These approaches aim at deriving 'robust' solutions that are less sensitive than classic optimal solutions to uncertainties. We, however, try to identify the conditions under which the proposed design still satisfies the managerial requirements when the environment becomes uncertain. This problem is of major practical relevance, because organizations must operate in environments that have become volatile so the future has become very uncertain.

To solve this problem, we propose a *methodology* that combines random (stochastic) simulation models, metamodels (namely first-order polynomials) that approximate the underlying simulation models, Design Of Experiments (DOE), and bootstrapping (a statistical computer technique). A statistical complication is that we consider simulation responses that are not averages but quantiles (or percentiles, as some people call them).

We organize our article as follows. Section 2 formalizes the problem. Section 3 details our methodology. Section 4 presents a manufacturing example taken from Feyzioglu et al. (2005), which includes aleatory uncertainty. Section 5 gives a 2^{14-10} design and fits a first-order polynomial regression metamodel to the data of a pilot experiment; because the classic regression assumptions do not hold, bootstrapping and cross-validation are applied. Section 6 presents a second experiment, which gives some simulation responses that are acceptable and some that are not. Section 7 summarizes our conclusions and topics for future research. (Hasty readers may skip paragraphs that start with the word 'Note'.)

2 Problem formulation

In most publications, the simulation analysts are interested in studying decision variables that characterize the possible design solutions. In manufacturing systems, typical examples include the number of operators, number of pallets, etc. In our study, the design has already been suggested by the designer team, so we emphasize those parameters that are used in the simulation model when evaluating system performance but are actually uncertain. We therefore consider these parameters as variables, and study the effects of their changes. Let $\mathbf{x} = (x_1, ..., x_i, \ldots, x_k)'$ be the k-dimensional vector of variables that may change due to epistemic uncertainty. Examples are the breakdown rate of a machine, and the mean customer demand for a particular product; we shall present other examples in the manufacturing simulation of Section 4. We assume that each x_i may have values within an interval $I_i = [L_i; H_i]$ (in robustness analysis, x_i is sampled from an interval with possibly $L_i = -\infty$ or $H_i = \infty$; we, however, assume finite intervals).

Obviously, these uncertain variables x_i can have an important effect on system performance, and hence on the acceptability of the future system. Following Feyzioglu et al. (2005), we assume that the production requirements are expressed as the probability of the future system producing at least a given number of products b_j (to meet forecasted demand). If the system produces (say) s types of products (or product families), then the acceptability of the system can be quantified as its capability of producing more than b_j products of type j ($j = 1, \ldots, s$), where b_j represents the threshold for product

j given by management (using the demand forecast). The random number of manufactured products of type j (say) q_j may be expressed as a function of the uncertain variables $\mathbf{x}$, so $q_j = f_j(\mathbf{x})$. Because of the system's aleatory and epistemic uncertainties, the requirement for each product will be satisfied only with a probability that in general is lower than 100%; of course, managers would like that this requirement be satisfied with a probability that is sufficiently high. Let $q_{j;p}$ denote the p^{th} *quantile* (e.g., the 5% quantile) of the j^{th} output (say) q_j:

$$P(q_j < q_{j;p}) = p_q \ (j = 1, \ldots, s); \tag{1}$$

i.e., there is only a p_q chance that output j is lower than $q_{j;p}$. Hence, the *managerial requirement* can be expressed as

$$q_{1;p} \geq b_1, \ldots, q_{s;p} \geq b_s \tag{2}$$

where $\mathbf{b} = (b_1, \ldots, b_s)'$is given by management. If all s constraints hold, then we say that the system gives acceptable output. Consequently, determining the acceptability of a given design solution consists in determining which combinations of x_i lead to an $\mathbf{x}$ that meet the threshold values in (2). Assuming that these combinations form a closed set, we wish to find the frontier (say) $G(\mathbf{x})$ that separates acceptable and unacceptable solutions; (2) implies that acceptable solutions $\mathbf{x}$ satisfy the constraints $f(\mathbf{x}) \geq \mathbf{b}$. We shall see that in our example, our estimate of this frontier is a hyperplane in the k-dimensional space formed by x_i $(i = 1, \ldots, k)$.

In summary, our problem is to find the frontier $G(\mathbf{x})$ between acceptable and unacceptable design solutions. We select a manufacturing system as an example, but our methodology also applies to other areas, such as the design of a new hospital with uncertain expected values for the number of patients, their examination times, etc. In the next section, we present our methodology for estimating the frontier G.

3 Methodology

To solve the problem formulated in the previous section, we start with simulation to estimate the performance of the given design for various environments. Next we compare the simulation outputs with the thresholds b_j in (2), to determine whether these outputs satisfy the performance requirements. If some environments do not satisfy these requirements, then we try to estimate the frontier between acceptable and unacceptable environments.

To estimate the I/O behavior of this simulation model, we select simulation

experiments through the statistical methods of DOE: see Kleijnen (2008). To estimate the acceptability frontier, we use regression metamodels estimated from these simulation I/O data. Because Feyzioglu et al. (2005) express the company's requirements though quantiles instead of expected values of the outputs, we use bootstrapping to validate the regression metamodels and to determine Confidence Intervals (CIs) for the regression parameters. The details of the methodology now follow.

3.1 Simulation estimate of quantile

To estimate quantiles, we select the number of replicated simulation runs (say) m, as follows. To simplify our notation, we suppress the product index j in (1). We sort the m replicated simulation responses from low to high, which gives the 'order statistics' $q_{(1)}, \ldots, q_{(m)}$ (so $q_{(1)} = \min_r q_{(r)}$ and $q_{(m)} = \max_r q_{(r)}$ with $r = 1, \ldots, m$; see Kleijnen (1987, p. 38). A classic point estimate $\widehat{q_p}$ of the quantile q_p is

$$\widehat{q_p} = q_{(\lceil p_q m \rceil)} \tag{3}$$

where $\lceil\ \rceil$ denotes the integer resulting from rounding upwards (e.g., $\lceil 2.1 \rceil = \lceil 2.9 \rceil = 3$; see Chen (2008). A $(1 - \alpha)$ CI follows from Conover (1999):

$$P(q_{(L)} \leq q_p \leq q_{(H)}) = 1 - \alpha \tag{4}$$

with the 'Low' value L and the 'High' value H given by

$$L = \lceil p_q m - z_{1-\alpha/2} \sqrt{p_q \times (1 - p_q) \times m} \rfloor,\ H = \lceil p_q m + z_{1-\alpha/2} \sqrt{p_q \times (1 - p_q) \times m} \rfloor \tag{5}$$

where $z_{1-\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal variable z (analogous to q_p).

3.2 Metamodel estimated from simulation I/O data

To approximate the I/O behavior of the underlying simulation model, we use a metamodel (also called response surface, emulator, etc.). We start with a first-order polynomial to predict the quantile as a function of the uncertain environmental factors x_i $(i = 1, \ldots, k)$:

$$y_j = \beta_{0;j} + \beta_{1;j} x_1 + \ldots + \beta_{k;j} x_k + \varepsilon_j \tag{6}$$

where y_j denotes the metamodel's predictor of the simulated quantile $q_{j;p}$, $\beta_{0;j}$ denotes its intercept ('grand mean', 'overall mean'), $\beta_{i;j}$ denotes its first-order ('main') effect of input (or environmental variable), and ε_j denotes its

'residual'. It is important to distinguish between the 'original' and the 'standardized' (coded) inputs; the latter are scaled such that they vary between -1 and +1. This standardization implies that the effects β_k measure the *relative* importance of the inputs; see Kleijnen (2008, p. 31). Because we focus on the relative importance of the various inputs, we standardize the inputs.

To estimate the effects $\beta_j = (\beta_{0;j}, \beta_{1;j}, \ldots, \beta_{k;j})'$ in (6), we simulate a number of *scenarios*—combinations of values for the environmental variables. To select these scenarios, we use a *Resolution-III* (R-3) $n \times k$ design matrix (say) $\mathbf{D}$, which (by definition) enables the computation of the estimated regression parameters $\widehat{\beta_j} = (\widehat{\beta_{0;j}}, \widehat{\beta_{1;j}}, \ldots, \widehat{\beta_{k;j}})'$; see Kleijnen (2008). R-3 designs are often used as *screening* designs in the literature; i.e., these designs estimate which factors are really important (if there were hundreds of factors, then other design types would be needed; see Kleijnen 2008). After this screening or pilot stage, we shall examine the effect of these important environmental factors on the acceptability of the given combination of decision factors.

To further reduce the variances of the estimated effects, we use Common Random Numbers (CRN) when simulating the same replicate r for each scenario; we use a different sequence of Pseudo-Random Numbers (PRNs) for a different replicate r' ($r, r' = 1, \ldots, m; r \neq r'$; see Section 3.1).

We estimate the regression parameters β_j from the simulation I/O data $(\mathbf{X}, \widehat{\mathbf{q}})$ where $\mathbf{X} = (\mathbf{1}, \mathbf{D})$denotes the $n \times (1 + k)$ matrix of regression variables based on the R-3 design and $\widehat{\mathbf{q}_j}$ the ns estimated quantiles defined in (3). The *Least Squares* (LS) estimate is

$$\widehat{\beta_j} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{q}_j} \ (j = 1, \ldots, s). \tag{7}$$

To *test* these effects, we need their CIs. Actually, LS assumes *white noise*; i.e., the residual ε_j in (6) is Normally Independently Identically Distributed (NIID) (moreover, ε_j and $\varepsilon_{j'}$ are also correlated, but LS is not affected by this characteristic; see Rao 1959). Unfortunately, these assumptions do not hold in practice:

(1) The estimated quantiles $\widehat{\mathbf{q}_j}$ are only asymptotically normally distributed; see Chen (2008).
(2) To reduce the variance of the estimated factor effects $\widehat{\beta_j}$, we use CRN and assume that CRN creates positive correlations among the simulation outputs of different scenarios; i.e., the outputs are not independent.
(3) The estimated quantiles have variances (say) σ_i^2 that vary with the scenarios (so $\sigma_i^2 \neq \sigma^2; i = 1, \ldots, n$).

If the white-noise assumption does not hold, the OLS estimator is still unbiased but their CIs should not use the classic t statistic. We solve this problem

through bootstrapping.

Note: A complication (not discussed in Kleijnen 2008) is that we estimate a quantile—not a mean. We have m IID observations $q_1, \ldots, q_m$, which give the point estimate $\widehat{q} = q_{(\lceil p_q m \rceil)}$ (see equation 3) for the p_q quantile; for the mean the point estimate would be $\overline{q} = \sum_i q_i/m$. The latter estimator has estimated variance $s^2(\overline{q}) = \sum_i (q_i - \overline{q})^2/\{(m-1)m\}$. We can define an average quantile estimator, partitioning the m observations q_i into groups of equal size; e.g., partition the m observations on q into four groups of size $m/4$. From these groups we can compute the average $\overline{\widehat{q}} = \sum_l \widehat{q}_l/4$ where $\widehat{q}_l$ is based on the $m/4$ observations in group l, and $s^2(\overline{\widehat{q}}) = \sum_l (\widehat{q}_l - \overline{\widehat{q}})^2/\{(4-1)4\}$. However, the variance of $\widehat{q}_l$ does not equal the variance of $\widehat{q}$.

To estimate the variances of $\widehat{\beta_j}$ (LS estimates of factor effects; see equation 7), we apply *bootstrapping*. General bootstrap principles are presented by Kleijnen (2008), including references. First we explain bootstrapping for obtaining a CI for the estimated *quantile* (next, we shall explain bootstrapping for the resulting estimated factor effects). Because we use CRN, we resample the m n-dimensional vectors—each vector has n correlated simulation outputs $q_{r;i}$ ($r = 1, \ldots, m$; $i = 1, \ldots, n$); i.e., we have m IID multivariate observations $\mathbf{q}_r$. We formalize our bootstrap through the following pseudo-code:

(1) Initialize the replicate number: $r = 1$.
(2) Resample—with replacement—a replicate number r^* from the uniform distribution defined on the integers $1, \ldots, m$; i.e., the uniform density function is $p(r^*) = 1/m$ with $r^* = 1, \ldots, m$.
(3) Replace the r^{th} 'original' vector $\mathbf{q}_r$ by $\mathbf{q}_r^* = \mathbf{q}_{r*}$.
(4) If $r \leq m$ then $r = r + 1$ and return to Step 2; else proceed to the next step.
(5) Sort the bootstrapped q_r^* per scenario i ($i = 1, \ldots, n$), which gives $q_{(r)}^*$.
(6) Compute the bootstrapped p_q quantile $\widehat{q_i^*} = q_{(\lceil p_q m \rceil)}^*$ per scenario; see (3).
(7) Repeat Steps 1 through 6 (say) $B = 1000$ times (we select a high B value, because $\widehat{q_i^*}$ may 'very well' be the same value in some bootstrap samples).

To obtain CIs for the estimated *factor effects*, we proceed as follows. We use $\widehat{\mathbf{q}_b^*}$ (n-dimensional vector with the bootstrapped quantiles $\widehat{q_i^*}$ in bootstrap sample b with $b = 1, \ldots, B$), to calculate the bootstrapped factor effects:

$$\widehat{\beta_b}^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{q}_b^*} \ (b = 1, \ldots, B). \tag{8}$$

To estimate a (say) 95% CI per factor effect i, we sort the B values of $\widehat{\beta}_{i;b}^*$ and find the order statistics $\widehat{\beta}_{(\lceil 0.025B \rceil);i}^*$ and $\widehat{\beta}_{(\lceil 0.975B \rceil);i}^*$, which are the lower and upper bound for the CI.

3.3 Validation of metamodel and testing of factors effects

Before we use the CIs derived in Section 3.2 to test the individual estimated factor effects, we *validate* the estimated metamodel as a whole. This metamodel is based on (6):

$$\widehat{y_j} = \widehat{\beta_{0;j}} + \widehat{\beta_{1;j}} x_1 + \ldots + \widehat{\beta_{k;j}} x_k. \tag{9}$$

There are several ways to check whether this fitted metamodel is an adequate approximation of the true I/O function implied by the underlying simulation model; see Kleijnen (2008, pp. 54-63, 97-99). We apply measures that are also used by Feyzioglu et al. (2005, p. 257, 260). We do not apply the classic F lack-of-fit test because that test assumes white noise, which does not hold in our simulation experiment.

First, we compute the *Absolute Relative Error (ARE)*:

$$ARE_{j;i} = \frac{|\widehat{y_{j;i}} - q_{j;i}|}{q_{j;i}} \; (j = 1, \ldots, s; i = 1, \ldots, n). \tag{10}$$

Per product type j, we may compute the ARE averaged over the n scenarios, the maximum ARE, and the number of scenarios with AREs exceeding a given threshold (say) 0.10.

Next, we perform *leave-one-out cross-validation*: For each product j we delete I/O combination i from the complete set of n combinations, which gives the I/O data set $(\mathbf{X}_{-i}, \widehat{\mathbf{q}_{j;-i}})$. We recompute the original OLS estimator defined in (7):

$$\widehat{\beta_{j;-i}} = (\mathbf{X}'_{-i}\mathbf{X}_{-i})^{-1}\mathbf{X}'_{-i}\widehat{\mathbf{q}_{j;-i}}. \tag{11}$$

We use this recomputed estimator $\widehat{\beta_{j;-i}}$ to compute the regression predictor for the scenario of the deleted combination: $\widehat{y_{j;-i}} = \mathbf{x}'_i\widehat{\beta_{j;-i}}$. We repeat this procedure, until we have processed all n scenarios. Analogous to (10, we may then compute

$$ARE_{j;-i} = \frac{|\widehat{y_{j;-i}} - q_{j;i}|}{q_{j;i}}. \tag{12}$$

We may also determine *scatterplots* with the pairs $(\widehat{y_{j;-i}}, q_{j;i})$. Scatterplots, however, may have misleading scales, so we also compute the *normalized prediction errors*:

$$z_{j;i} = \frac{q_{j;i} - \widehat{y}_{j;-i}}{\sqrt{\widehat{var(q_{j;i})} + \widehat{var(\widehat{y}_{j;-i})}}} \tag{13}$$

where $\widehat{var(q_{j;i})}$ is estimated through the bootstrapped estimator $\widehat{var(q^*_{j;i})}$, and

$\widehat{var(\widehat{y}_{j;-i})} = \mathbf{x}_i' \widehat{cov(\widehat{\beta}_{j;-i})} \mathbf{x}_i$ where (11) implies

$$\widehat{cov(\widehat{\beta}_{j;-i})} = (\mathbf{X}_{-i}'\mathbf{X}_{-i})^{-1}\mathbf{X}_{-i}' \widehat{\mathbf{cov}(\widehat{\mathbf{q}^*_{j;-i}})} \mathbf{X}_{-i}(\mathbf{X}_{-i}'\mathbf{X}_{-i})^{-1} \tag{14}$$

where the matrix $\widehat{\mathbf{cov}(\widehat{\mathbf{q}^*_{j;-i}})}$ results from eliminating row i and column i from the $n \times n$ matrix $\widehat{\mathbf{cov}(\widehat{\mathbf{q}^*_j})}$, which has the elements

$$\widehat{cov(\widehat{q^*_{j;i}}, \widehat{q^*_{j;i'}})} = \frac{\sum_{b=1}^{B} (\widehat{q^*_{j;i;b}} - \overline{\widehat{q^*_{j;i}}})(\widehat{q^*_{j;i';b}} - \overline{\widehat{q^*_{j;i'}}})}{B-1}. \tag{15}$$

Note: The correlation coefficients $\rho_{j;i;i'} = \widehat{cov(\widehat{q^*_{j;i}}, \widehat{q^*_{j;i'}})} / \sqrt{\widehat{var(\widehat{q_{j;i}})} \widehat{var(\widehat{q_{j;i'}})}}$ quantify the effects of CRN. Furthermore, Kleijnen (2008, p. 61) uses t_{m-1} instead of z, but we study a quantile instead of an average. Finally, we cannot use the shortcut in Kleijnen (2008, p.61)—which uses $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$—because it assumes white noise.

Because (13) gives n values for each product, we reject the regression metamodel for product j if

$$\max_i |z_{j;i}| > z_{1-[\alpha/(2n)]} \tag{16}$$

where the right-hand side follows from *Bonferroni's inequality*, which implies that the classic type-I error rate (in this case $\alpha/2$) is replaced by the same value divided by the number of tests (namely, n).

After we have validated the s first-order polynomials, we test the individual factor effects. As point estimators we use the OLS estimates using all n scenarios. To obtain CIs for these effects, we should not use the classic formulas that assume white noise, namely $cov(\widehat{\beta_j}) = (\mathbf{X}'\mathbf{X})^{-1} var(q_j)$. Instead we use the bootstrapped quantiles $\widehat{q^*_{j;i}}$ that we have already generated to obtain $\widehat{cov(\widehat{q^*_{j;i}}, \widehat{q^*_{j;i'}})}$; see again (15). These bootstrapped quantiles $\widehat{\mathbf{q}^*_j}$ give the bootstrapped factor effects using (8). We reject the null-hypothesis of an unimportant effect for a given input and product if its CI interval $[\widehat{\beta}^*_{(\lceil 0.025B \rceil)}, \widehat{\beta}^*_{(0.975B)}]$ does not contain the value 0; this gives the (say) h significant factors among the k potential effects per product.

3.4 Estimate acceptable factors combinations

Once the fitted metamodels are validated, substituting these metamodels into (2) gives

$$\widehat{\beta_{0;j}} + \widehat{\beta_{1;j}} x_1 + \ldots + \widehat{\beta_{h;j}} x_h \geq b_j \ (j = 1, ..., s). \tag{17}$$

Replacing the $\geq$ sign by the $=$ sign in (17) gives the estimated frontier per product j. It is easy to give a geometric representation of the acceptability frontier if we focus on the two most important factors (highest $\left|\widehat{\beta_{h;j}}\right|$).

4 Example: a manufacturing system

4.1 The Feyzioglu et al. (2005) example

Feyzioglu et al. (2005) consider a company that wished to build a manufacturing system so they asked some engineers to make a design for their factory. Now that the factory is designed, the company realizes that the future may differ from the future that was assumed during the design. Therefore the company now asks whether the design will turn out to be 'acceptable'. Feyzioglu et al. (2005) focus on the 0.05 quantiles $q_{j;0.05}$ (called the '0.05-percentiles' by Feyzioglu et al. 2005, p. 260), and assume that the production requirements are

$$q_{1;0.05} \geq 15000 \text{ and } q_{2;0.05} \geq 17000 \tag{18}$$

where 15000 and 17000 are the threshold values given by the company.

In this example, the factory consists of four workstations. Each workstation has identical machines. In our article, we assume that the number of machines per workstation is given by the 'optimal' solution derived by Feyzioglu et al. (2005). (Because we re-programmed the simulation model, our model and Feyzioglu et al.'s model may differ; anyhow, we use their solution as the solution that management wishes us to evaluate.)

The factory produces two types of product, called prod1 and prod2. These product types have different routings. If we denote WorkStation i by WS_i with $i = 1, \ldots, 4$, then prod1 has routing WS_1, WS_3, WS_4 and prod 2 has WS_2, WS_3, WS_4. There may be a queue—denoted by Q—after WS_1, WS_2, and WS_3; also see Figure 2 in Feyzioglu et al. (2005, p. 259). Denoting the exponential distribution with mean λ by $E(\lambda)$ and the normal distribution with mean μ and standard deviation σ by $N(\mu, \sigma)$ gives the *base scenario* in Table 1, reproduced from Feyzioglu et al. (2005, p. 259).

The outputs of the simulated factory are the two production volumes produced during a thirty-day period, after a warm-up period of 10 days. These responses (say) q_1 and q_2 must satisfy given 'threshold' values. These values are minimum values for the 5% quantiles of the distributions of q_1 and q_2 respectively; these thresholds are 15000 and 17000; see (18).

Product	Interarrival time a	Processing time p			
		WS_1	WS_2	WS_3	WS_4
prod1	$E(2)$	$N(5, 0.1)$	-	$N(3, 0.05)$	$N(2, 0.2)$
prod2	$E(1.4)$	-	$N(5.5, 0.1)$	$N(3, 0.05)$	$N(2.5, 0.1)$

Table 1
Base scenario: interarrival and processing time distributions (parameters in minutes)

4.2 Example revisited

In practice, the distributions in Table 1 are *uncertain*; e.g., the interarrival times of product 1 may indeed be exponentially distributed but with a parameter not equal to 2. In this article, we therefore focus on these epistemically uncertain factors; i.e., if the distributions in Table 1 change, will the 'optimal' design still satisfy the requirements in (18)? We allow all fourteen parameters (λ_1, λ_2, $\mu_{1;1}$, ..., $\sigma_{2,4}$.) to differ from the base-scenario values in Table 1.

Like Feyzioglu et al., we estimate the production volumes q_j through discrete-event simulation. We program the simulation model in Arena (whereas they use Promodel 4.2); see Kelton et al. (2007). Feyzioglu et al. use 25 *replicates* to obtain the estimates (say) $\widehat{q_j}$ of the two quantiles $q_{j;0.05}$ defined in (1), where we simplify the notation, suppressing the subscript 0.05 in $\widehat{q_j}$. We, however, derive the number of replicates using (3) through (5); e.g., $m = 100$ gives $L = 1$ and H=10, $m = 25$ gives $L = 0$ and $H = 4$ (so the lower bound of the CI is not defined), and $m = 80$ makes L becomes1. We decide to select $m = 100$, because this choice gives a reasonably accurate estimate for the 'base scenario' defined by Table 1 (other scenarios defined below may require other m values, but we ignore this complication).

5 Pilot experiment

5.1 Design and metamodel

We select a 2^{14-10} design, which implies 16 scenarios that enable estimation of the fourteen effects β in (6). (This design for fourteen noise factors happens to have the same size as the 2^4 design for the four decision factors in Feyzioglu et al., p. 260.) We use the following 'generators', where (e.g.) $\mathbf{5 = 1.2}$ denotes $x_{i;5} = x_{i;1}x_{i;2}$ ($i = 1, \ldots, n = 2^{14-10}$): $\mathbf{5 = 1.2}$, $\mathbf{6 = 1.3}$, $\mathbf{7 = 1.4}$, $\mathbf{8 = 2.3}$, $\mathbf{9 = 2.4}$, $\mathbf{10 = 3.4}$, $\mathbf{11 = 1.2.3}$, $\mathbf{12 = 1.2.4}$, $13 = \mathbf{1.3.4}$, and $\mathbf{14 = 2.3.4}$. (We could use a different design; e.g., we could replace $\mathbf{14 = 2.3.4}$ by $\mathbf{14 = 1.2.3.4}$; i.e., there are alternative 2^{14-10} designs that all require 16 runs.) These generators imply

that the 16×14 design matrix $\mathbf{D}$ has a first row $\mathbf{d}_1'$ with only plus ones (so $d_{1;j} = 1$) and a last row $\mathbf{d}_{16}' = (-1, -1, -1, -1, 1, 1, 1, 1, 1, 1, -1, -1, -1, -1)$'. In the first-order polynomial metamodel, the matrix of explanatory regression variables $\mathbf{X}$ adds a column with 16 ones to $\mathbf{D}$ so $\mathbf{X} = (\mathbf{1}, \mathbf{D})$.

In our pilot experiment we choose as lower and upper factor values 90% and 110% of the values in Table 1; e.g., $x_{16;1}$ = -1 means that the original variable λ_1 has the value $0.90 \times 2 = 1.8$.

Note: A smaller change (say, 5%) has a higher probability of resulting in scenarios with acceptable simulation responses, because they differ less from the Pareto optimal solution derived by Feyzioglu et al.; actually, we are interested in finding the border line between acceptable and unacceptable scenarios so we like to see both types in our simulation. A statistical argument in favor of the 10% instead of the 5% change is that the larger change decreases the variances of the estimated factor effects; see Kleijnen (2008, p. 28). Moreover, a larger change decreases the probability of extrapolation when estimating the acceptability frontier; extrapolation increases the prediction error; see Kleijnen (2008, p.110). The mathematical Taylor series argument, however, favors small changes in the factors, because it increases the probability of a valid first-order polynomial approximation. In the design that follows after the pilot design we shall return to this issue.

Simulation of these 16 scenarios—each replicated $m = 100$ times—gives $\widehat{q}_j$, the estimated 5% quantiles for the two products defined in (3), and the 95% CI defined in (4) where $L = 1$ and H =10 so this CI is $(q_{(10)} - q_{(1)})$; see Table 2. We analyze this table as follows.

First we compare $\widehat{q}_1$ and $\widehat{q}_2$ (simulation estimates of two quantiles) with their thresholds 15000 and 17000; see (18). Only scenario 2 gives an unacceptable result ($14724 < 15000$). Next we examine the CIs. Only scenario 2 gives a significantly low output. Even if we are risk-averse pessimists, we conclude that the other 15 scenarios do not give significantly low outputs: $\widehat{q}_1 > 15000$ in 15 scenarios and $\widehat{q}_2 >$17000 in all 16 scenarios.

Note: The CIs $(q_{(10)} - q_{(1)})$ are very short. The cause is the low variability in the simulation responses (because of the long simulation runs); e.g., the base scenario gives simulation outputs for product 1 that vary only between 17,272 and 17,286**.** Nevertheless, we need as many as 100 responses to estimate a small quantile like a 5% quantile.

To examine whether the *variability* of the simulation output varies with the scenarios, we compute the ratio (say) r_j of the maximum and minimum CI

Scenario	$\widehat{q_1}$	confidence interval	$\widehat{q_2}$	confidence interval
1	15705	(15704, 15706)	21415	(21413, 21417)
2	14724	(14723, 14725)	24541	(24538, 24541)
3	19194	(19181, 19196)	21416	(21415, 21417)
4	18558	(18553, 18559)	25308	(25306, 25308)
5	15706	(15703, 15707)	21417	(21414, 21417)
6	15705	(15702, 15706)	26176	(26173, 26178)
7	19196	(19189, 19197)	21417	(21415, 21419)
8	19195	(19189, 19196)	26175	(26172, 26177)
9	15704	(15701, 15705)	21416	(21415, 21418)
10	15705	(15703, 15705)	26178	(26175, 26178)
11	19193	(19191, 19195)	21418	(21417, 21418)
12	17995	(17993, 17996)	24540	(24536, 24541)
13	15706	(15705, 15707)	21416	(21414, 21417)
14	15705	(15704, 15706)	26176	(26174, 26178)
15	18558	(18557, 18559)	20705	(20704, 20706)
16	16611	(16610, 16612)	20705	(20704, 20706)

Table 2
Estimated quantiles and their CIs from 16 scenarios

lengths for product j:

$$r_j = \frac{\max_i[q_{(10);i;j} - q_{(1);i;j}]}{\min_i[q_{(10);i;j} - q_{(1);i;j}]} \ (j = 1, 2). \tag{19}$$

If q were a sample average (not an order statistic), then r^2 would be the ratio of two sample variances and we could apply the test mentioned in Kleijnen (2008, p. 88). We do not know a statistical test for (19) (we could derive a bootstrap test). Anyhow, the observed r_j values 7.5 and 5 suggest that the assumption of constant variances in classic regression analysis may not hold; we shall return to this issue.

It is easy to check that the 2^{14-10} design gives an orthogonal $\mathbf{X}$, so (7) simplifies to

$$\widehat{\beta_{h;j}} = \frac{\sum_{i=1}^{16} x_{i;h}\widehat{q_{i;j}}}{16} \ (h = 0, 1, \ldots, 14) \ (j = 1, 2). \tag{20}$$

This gives the LS estimates displayed in Table 3. We interpret this table, assuming that the first-order polynomial is an adequate approximation (valid

LS estimate	product 1	product 2
$\widehat{\beta_{0;j}}$	17072.50	23151.19
$\widehat{\beta_{1;j}}$	175.38	331.94
$\widehat{\beta_{2;j}}$	−24.75	127.81
$\widehat{\beta_{3;j}}$	−1490.00	440.69
$\widehat{\beta_{4;j}}$	297.75	−1823.69
$\widehat{\beta_{5;j}}$	−227.38	−440.94
$\widehat{\beta_{6;j}}$	−297.88	−536.56
$\widehat{\beta_{7;j}}$	−95.38	−243.19
$\widehat{\beta_{8;j}}$	−147.75	−332.19
$\widehat{\beta_{9;j}}$	54.00	−39.06
$\widehat{\beta_{10;j}}$	−175.00	−352.20
$\widehat{\beta_{11;j}}$	104.88	236.06
$\widehat{\beta_{12;j}}$	147.88	351.44
$\widehat{\beta_{13;j}}$	218.13	447.81
$\widehat{\beta_{14;j}}$	68.25	242.94

Table 3
LS estimates of effects in first-order polynomial, for products 1 and 2

metamodel) and the estimated factor effects are significant (we shall return to these assumptions). We focus on product 1, to save space.

(1) The estimated effect of factor 1 (λ_1) has a plus sign, whereas we expect a minus sign because higher mean interarrival time implies fewer products arriving into the system so the output decreases. However, workstations 1 and 3 may reach saturation points, i.e. the number of products arriving exceeds the workstations' capacities.
(2) Factor 3 ($\mu_{1;1}$) is the most important factor ($\widehat{\beta_{3;1}} = -1490$). Its minus sign implies that a higher mean processing time at work station 1 decreases the output of product 1.
(3) Factor 9 ($\sigma_{1;1}$)'s plus sign means that a higher standard deviation increases the output number, which is hard to explain.
(4) Factors 5 ($\mu_{1;3}$), 7 ($\mu_{1;4}$), 11 ($\sigma_{1;3}$) and 13($\sigma_{1;4}$) have the same signs as factors 3 and 9, because their influences are similar.
(5) Factor 4 ($\mu_{2;2}$) and factor 10 ($\sigma_{2;2}$) do not direct influence the output of product 1, but a higher processing time at workstation 2 for product 2 obviously gives fewer arrivals of product 2 before workstation 3. Because workstation 3 serves FIFO, fewer products 2 in its queue implies smaller

waiting time for product 1.

(6) Factors 6 ($\mu_{2;3}$) and 8 ($\mu_{2;4}$) have the same influence on product 1; their minus signs are explained by the fact that products 1 and 2 share workstations 3 and 4. If the processing times for product 2 increase at workstations 3 or 4, then the output of product 1 reduces.

Some factors may have non-significant effects; e.g. $\widehat{\beta_{2;1}} = -24.75$; its sign may actually be positive.

After bootstrapping as described in Subsection 3.2, we get $\widehat{var(\widehat{q^*_{i;j}})}$ displayed in Table 4. We see that these variances are small—which agrees with our comment on the lengths of the distribution-free CI in Table 2. We shall use these variances below.

Note: This table further show that $\max_i(\widehat{var(\widehat{q^*_{i;1}})}) / \min_i(\widehat{var(\widehat{q^*_{i;1}})} = 7.739/1.233 = 6.3$, whereas the ratio of the ranges of the distribution-free CIs was 7.5; see (19). Product 2 gives $\max_i(\widehat{var(\widehat{q^*_{i;2}})}) / \min_i(\widehat{var(\widehat{q^*_{i;2}})} = 5.312/1.659 = 3.2$, whereas the distribution-free CIs gives 5.

5.2 Validation of the fitted polynomial metamodel

We apply measures that are also used by Feyzioglu et al. (2005, p.257, 260). First, we compute the classic *coefficient of determination* R^2 and the adjusted coefficient $R^2_{adjusted}$. We find that these two measures are 0.9998 and 0.997 for product 1, and 0.9956 and 0.934 for product 2. We find these values excellent (statistical tests of these values are rather complicated; see Kleijnen and Deflandre 2006).

Next we compute the $ARE_{j;i}$ defined in (10) For product 1 our experiment gives an average ARE of 0.0015 and a maximum of 0.0017, so no ARE exceeds the threshold of 0.10. For product 2 these values are 0.006 and 0.007, so again no ARE exceeds the threshold. Besides these $ARE_{j;i}$ for the 16 'extreme' scenarios $i = 1, \ldots, 16$, we compute the ARE for the base scenario (say) $ARE_{j;0}$, which gives 0.0117 and 0.0173. Altogether we find the AREs excellent.

We also make a classic *scatterplot* with the 16 quantiles estimated from the simulated outputs versus the quantiles predicted through the fitted first-order polynomial (9) for product 1; see Figure 1. For product 2 we obtain a similar scatterplot; see the online appendix.

Finally, we compute the *normalized prediction errors* $z_{i;j}$ defined in (13). We find that the corresponding test statistic $\max_{i;j} |z_{i;j}|$ defined in (16) is ex-

Scenario	$\widehat{var(\widehat{q^*_{i;1}})}$	$\widehat{var(\widehat{q^*_{i;2}})}$
1	3.626	2.945
2	1.834	3.019
3	3.393	2.510
4	1.986	5.312
5	1.516	2.218
6	2.185	4.046
7	1.947	1.659
8	2.466	4.159
9	4.306	3.863
10	2.959	4.602
11	7.739	2.307
12	4.701	2.972
13	2.606	3.340
14	2.960	3.885
15	1.233	2.927
16	2.255	2.927

Table 4
Bootstrapped variances of estimated outputs for sixteen scenarios

tremely significant, namely -192 in scenario 16 for product 1 and -1347 in scenario 7 for product 2. Our explanation is that the simulation outputs show very little variability, as we have already pointed out**.** This result illustrates that statistically significant results are not always important; see Kleijnen (2008, p.31) for a general discussion. In our case the normalized prediction errors are very significant, but the AREs are acceptable for our purpose, namely validating first-order polynomials and identifying the important factors in these polynomials. (Nevertheless, the test in (16) may be useful in other simulation models with more variability.)

Note; The online appendix also gives $\widehat{\beta_{j;-i}}$ where $\widehat{\beta_{j;-i}}$ denotes the vector with the 15 estimated effects (including the intercept) for $i = 0, 1, \ldots, 16$ where $i = 0$ means that no combination is deleted; j corresponds with product 1 and product 2 respectively. Furthermore, this appendix gives the n predictions $\widehat{q_{j;-i}}$ $(i = 1, ..., n)$ based on $\widehat{\beta_{j;-i}}$, and the corresponding ARE based on (10). Finally, this appendix gives *scatterplots*.

Once, the metamodel-as-a-whole is validated, we find the significant factors

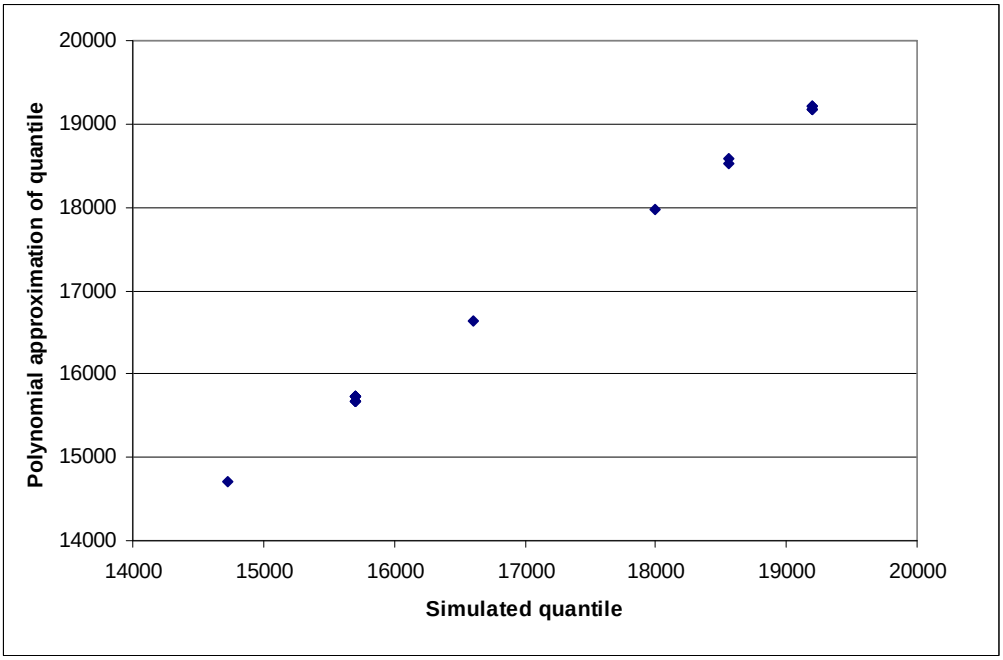

Fig. 1. Quantiles simulated versus quantiles predicted through first-order polynomial for product 1

through the CI based on the bootstrapped effects (8). This gives Table 5, which implies that all effects are significantly different from zero (which can be explained by the low variability of the simulated quantiles).

Note: If we were interested in the mean of the simulation response instead of its quantile, then we would not need to bootstrap; i.e., Kleijnen (2008, p. 95) explains how to derive a CI from the m replicates assuming $m \geq 2$. Each replicate gives an unbiased estimator of the mean, whereas we need many replicates to get a single estimate of the quantile.

We conclude that first-order polynomials give adequate approximations, and that all fourteen factors are significant. These preliminary conclusions guide our next experiment.

6 Follow-up experiment

In our pilot experiment, we changed all factors by -10% or +10% of their base values and found that nearly all scenarios gave acceptable responses; see again Table 2. In our next experiment, we select the lower and upper values per factor such that we may expect that only (approximately) half of the simulation responses will be acceptable; i.e., because we wish to estimate the acceptability frontier, we do not wish (nearly) all our responses to be

LS estimate	$\widehat{\beta}^*_{(2.5);1}$	$\widehat{\beta}^*_{(97.5);1}$	$\widehat{\beta}^*_{(2.5);2}$	$\widehat{\beta}^*_{(97.5);2}$
0	17076.97	17083.50	23156.56	23162.81
1	174.13	175.31	331.94	332.31
2	25.06	26.13	127.31	127.81
3	−1491.06	−1490.56	440.63	441.31
4	297.81	298.25	−1824.44	−1823.66
5	−228.13	−227.50	−441.06	−440.63
6	−298.16	−297.81	−537.06	−536.50
7	−95.72	−95.31	−243.63	−243.13
8	−148.25	−147.81	−332.38	−331.94
9	54.56	55.63	−38.75	−38.50
10	−175.44	−175.00	−351.81	−351.50
11	104.69	105.63	235.94	236.44
12	147.75	148.06	351.75	352.13
13	218.13	218.38	447.63	448.19
14	68.00	68.31	243.13	243.44

Table 5
CIs for intercept and 14 factor effects, computed from 100 bootstrapped quantiles, for products 1 and 2

acceptable. The DOE literature does not pay much attention to the selection of an appropriate *area of experimentation* or area of interest; Zeigler, Praehofer, and Kim (2000) call this area the experimental frame.

In the new experiment, the standardized factor $x = -1$ means that the original factor changes by +10% and $x = +1$ means that the original factor changes by +30%; e.g., a 10% increase of the mean interarrival time implies fewer arrivals, so the production volume decreases. We use the same 2^{14-10} design in the standardized factors (see Section 5.1), and the same number of replicates, namely 100. This results in Table 6, which shows that nine of the sixteen scenarios give unacceptably low responses for product 1 (see $\widehat{q}_1$ and the threshold 15000 in equation 18); no scenario gives unacceptable responses for product 2 (see $\widehat{q}_2$ and the threshold 17000). The predictions $\widehat{y_1}$ and $\widehat{y_2}$ based on first-order polynomials give AREs defined in (10), which result in maxima of 0.002 for both product 1 and product 2; $R^2_{adjusted}$ is 0.99 for both products. So we conclude that the two metamodels are adequate.

Given the adequacy of the fitted metamodel, we next display the OLS estimates of the factor effects in the second experiment; see Table 7.

Scenario	$\widehat{q_1}$	$\widehat{y_1}$	$\widehat{q_2}$	$\widehat{y_2}$
1	13289	13314.31	18123	18152.88
2	12721	12695.69	20502	20472.13
3	15705	15679.69	18123	18093.13
4	15425	15450.31	21034	21063.88
5	13289	13263.69	18123	18093.13
6	13289	13314.31	21418	21447.88
7	15705	15730.31	18123	18152.88
8	15704	15678.69	21418	21388.13
9	13289	13263.69	18123	18093.13
10	13290	13315.31	21417	21446.88
11	15705	15730.31	18123	18152.88
12	15033	15007.69	20502	20472.13
13	13290	13315.31	18123	18152.88
14	13288	13262.69	21417	21387.13
15	15425	15399.69	17798	17768.13
16	14056	14081.31	19167	19196.88

Table 6
Follow-up experiment: quantiles simulated and predicted through metamodel

Next we replace the quantiles $q_{1;0.05}$ and $q_{2;0.05}$ in the managerial production requirement (18) by their *metamodel* estimates $\widehat{y_j}$ based on the first-order polynomial (6) with the estimated factor effects of Table 7. This substitution enables us to estimate whether the thresholds are satisfied for future scenarios; i.e., two hyperplanes in the 14-dimensional input space form the *frontier* of the region of acceptable scenarios:

$$\begin{aligned} \widehat{\beta_{0;1}} + \widehat{\beta_{1;1}} x_1 + \ldots + \widehat{\beta_{14;1.}} x_{14} &= 15000 \\ \widehat{\beta_{0;2}} + \widehat{\beta_{1;2}} x_1 + \ldots + \widehat{\beta_{14;2.}} x_{14} &= 17000. \end{aligned} \tag{21}$$

We illustrate this 14-dimensional frontier as follows.

First we assume that all 14 factors are at their base values; i.e., $\mathbf{x}' = (x_1, \ldots, x_{14}) = \mathbf{0}' = (0, \ldots, 0)$. Then (21) implies

$$E(\widehat{y_1} \mid \mathbf{x} = \mathbf{0}) = \widehat{\beta_{0;1}} \text{ and } E(\widehat{y_2} \mid \mathbf{x} = \mathbf{0}) = \widehat{\beta_{0;2}}. \tag{22}$$

Initially we assume that a single factor (say) x_h deviates from its base value

LS estimate	product 1	product 2
$\widehat{\beta_{0;j}}$	14281.44	19470.88
$\widehat{\beta_{1;j}}$	109.44	137.13
$\widehat{\beta_{2;j}}$	25.69	22.50
$\widehat{\beta_{3;j}}$	−1063.31	184.88
$\widehat{\beta_{4;j}}$	180.69	−1388.50
$\widehat{\beta_{5;j}}$	−131.56	−185.00
$\widehat{\beta_{6;j}}$	−180.56	−251.38
$\widehat{\beta_{7;j}}$	−74.56	−96.50
$\widehat{\beta_{8;j}}$	−96.56	−137.00
$\widehat{\beta_{9;j}}$	9.19	18.13
$\widehat{\beta_{10;j}}$	−109.56	−144.25
$\widehat{\beta_{11;j}}$	60.44	70.50
$\widehat{\beta_{12;j}}$	96.69	144.38
$\widehat{\beta_{13;j}}$	145.44	210.75
$\widehat{\beta_{14;j}}$	61.44	96.38

Table 7
LS estimates of first-order polynomial for products 1 and 2 in follow-up experiment

while all other factors remain at their base values: $\mathbf{x}_{-h}= \mathbf{0}$. Then (21) implies

$$E(\widehat{y_1} \mid x_h, \mathbf{x}_{-h}= \mathbf{0}) = \widehat{\beta_{0;1}} + \widehat{\beta_{h;1}} x_h \text{ and } E(\widehat{y_2} \mid x_h, \mathbf{x}_{-h}= \mathbf{0}) = \widehat{\beta_{0;2}} + \widehat{\beta_{h;2}} x_h. \tag{23}$$

Then (22) and (23) give Figure 2, assuming that $\widehat{\beta_{h;1}} > \widehat{\beta_{h;2}} > 0$; the intersection of the two (first-order polynomial) response curves with the two corresponding thresholds gives the acceptable ('green') variations and the unacceptable ('red') variations in (coded) factor x_h; actually, the threshold for product 1 gives the truly acceptable variations, because product 2 gives more acceptable variations than product 1 but both thresholds must be satisfied. The acceptability of a change in x_h is also determined by the change in the other factors $x_{h'}$—even though the first-order polynomials imply that there are no factor interactions.

Next we assume that *two* factors (say) x_h and $x_{h'}$ deviate from their base values while all other factors remain at their base values; we denote this scenario by

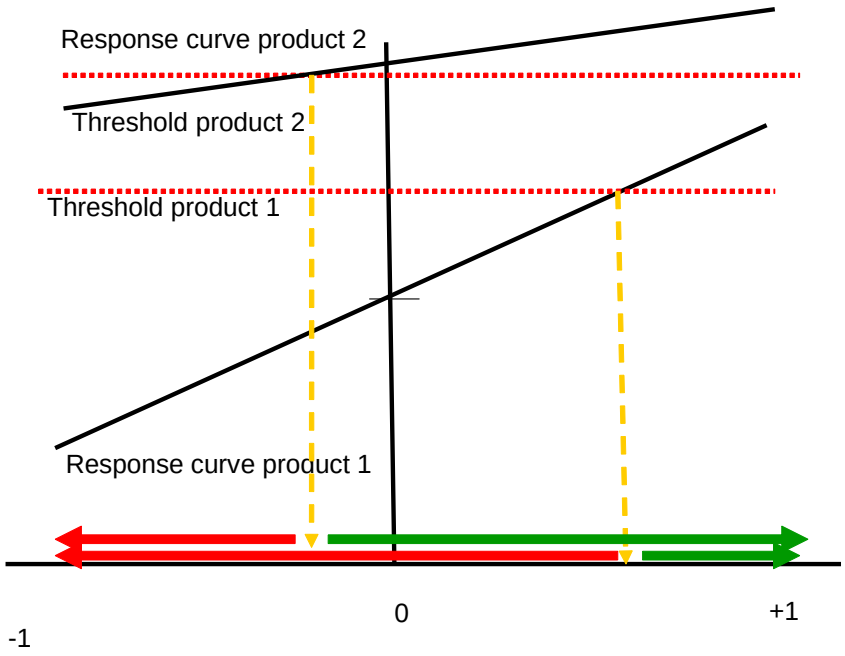

Fig. 2. Acceptable and unacceptable factor variations

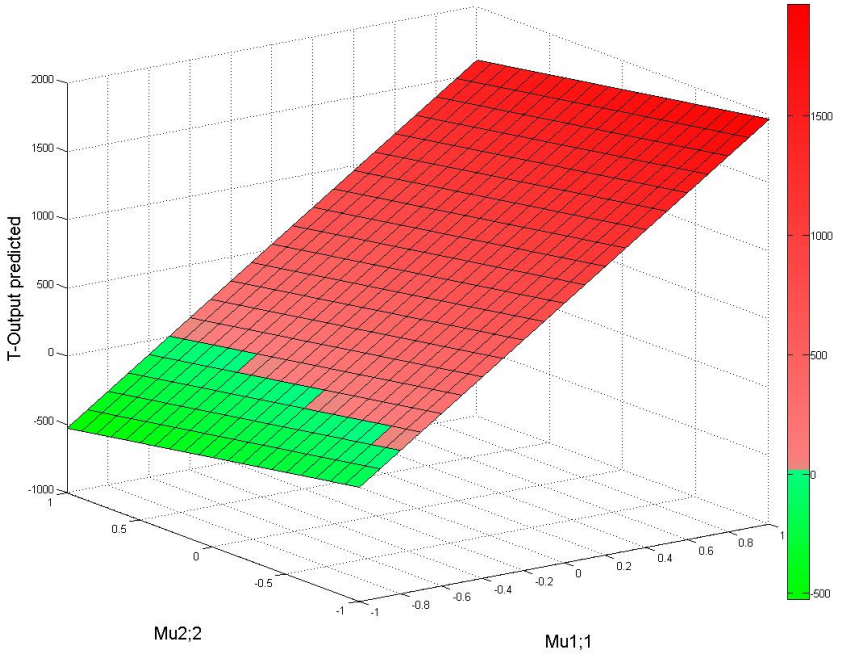

Fig. 3. Acceptability frontier for product 1, when factors 1 and 2 change

$x_h, x_{h'}, \mathbf{x}_{-(h;h')} = \mathbf{0}$. Then (21) implies

$$E(\widehat{y_1} \mid x_h, x_{h'}, \mathbf{x}_{-(h;h')} = \mathbf{0}) = \widehat{\beta_{0;1}} + \widehat{\beta_{h;1}} x_h + \widehat{\beta_{h';1}} x_{h'} \tag{24}$$
$$E(\widehat{y_2} \mid x_h, x_{h'}, \mathbf{x}_{-(h;h')} = \mathbf{0}) = \widehat{\beta_{0;2}} + \widehat{\beta_{h;2}} x_h + \widehat{\beta_{h';2}} x'_h.$$

To illustrate (24), we plot the effects of the two most important factors for product 1—namely x_1 corresponding with the original factor $\mu_{1;1}$ and x_2 or $\mu_{2;2}$ (see Table 7)—and the threshold 15000:

$$14281.44 - 1063.31x_1 + 180.6875x_2 = 15000.$$

This equation gives Figure 3, which shows that low values for $\mu_{1;1}$ give acceptable production volumes. For product 2, all combinations of its two most important factors—namely $\mu_{2;2}$ and $\mu_{2;3}$—give acceptable production volumes; we do not display this figure.

In practice, all factors may deviate from their base values, so the two preceding figures are simplifications (meant to illustrate the issue); we recommend the analytical representation (21) instead of the geometric representation.

Note: Our estimate of the frontier is very accurate, because the 100 IID simulation observations on the production volumes q_1 and q_2 show little variation so the bootstrapped quantiles $\widehat{q^*_{i;1}}$ and $\widehat{q^*_{i;2}}$ show little variation (see Table 4), which implies little variation in the bootstrapped factor effects $\widehat{\beta_j}^*$ (see Table 5). We do not present the plot with the bootstrapped planes because these planes seem to coincide—unless we zoom-in on the plot.

In conclusion, the simulated system is sensitive to the changes in the environmental factors. A 30% change of the interarrival rate for product 1 makes the system's performance unacceptable. The acceptability depends primarily on product 1, because the output volume for product 2 exceeds its threshold, for every simulated scenario. For product 1, the most important factors are the mean processing time of product 1 at workstation 1 and the mean processing time of product 2 at workstation 2; these two factors should be closely monitored.

7 Conclusions and future research

In the engineering design of various types of system, it is common that the future system specification is the result of a preliminary study. It is also common that a simulation study is required, to verify that the proposed specifications are indeed acceptable; i.e., the suggested system design satisfies the performance requirements. Unfortunately, most classical simulation studies carried out in such a context do not properly take account for the uncertainty of the data used in the simulation model during the design phase. Consequently, the conclusions about the ability of the system to satisfy the requirements may be wrong whenever the environmental conditions differ from those initially expected.

To avoid such dangerous situations, we consider two types of 'uncertainty' that are typically encountered in practice—namely aleatory and epistemic—and we formulate the problem of the acceptability of a given design accordingly. We show how this problem can be solved through a methodology that combines simulation, bootstrapping, and DOE.

Our example highlights that—though a design solution may be considered acceptable for the environment initially assumed by the decision makers and the analysts—specific changes in the uncertain input data may make this solution unacceptable. Knowledge of the conditions—in other words, the frontier—that make the solution unacceptable, provides precious information for the decisions makers. We therefore expect that our methodology will be beneficial in various real-world applications.

We based our DOE on first-order polynomial regression analysis, but other metamodeling approaches (e.g., Kriging) may also help to determine the acceptability regions; see Kleijnen (2008). To improve the identification of the frontier between acceptable and unacceptable zones, other relevant research approaches may be signal processing and reliability analysis.

References

Chen, E.J. (2008), Some procedures of selecting the best designs with respect to quantile. *Simulation*, 84, pp. 275-284

Conover, W.J. (1999), *Practical non-parametric statistics: third edition*. Wiley, New York

De Rocquigny, E., N. Devictor, and S. Tarantola (2008), Uncertainty settings and natures of uncertainty. In: *Uncertainty in industrial practice,* edited by E. de Rocquigny, N. Devictor, and S. Tarantola, Wiley, Chichester, England

Dellino, G., J.P.C. Kleijnen, and C. Meloni (2008), Robust optimization in simulation: Taguchi and Response Surface Methodology. Working Paper

Feyzioglu, O., H. Pierreval, and D. Deflandre (2005), A simulation-based optimization approach to size manufacturing systems. *International Journal of Production Research*, 43, no. 2, pp. 247-266

Kelton, W.D., R.P. Sadowski, and D.T. Sturrock (2007), *Simulation with Arena; fourth edition*. McGraw-Hill, Boston

Kleijnen, J.P.C. (1987), *Statistical tools for simulation practitioners*. Marcel Dekker, New York

Kleijnen, J.P.C. (2008), *Design and Analysis of Simulation Experiments*, Springer

Kleijnen, J.P.C. and D. Deflandre (2006), Validation of regression metamodels in simulation: Bootstrap approach. *European Journal of Operational Research*, 170, no. 1, pp. 120-131

Kleijnen, J.P.C. and C. R. Standridge, 1988, Experimental Design and Regression Analysis in Simulation: an FMS Case Study, *European Journal of Operational Research,* 33, 257-261

Madu, I E. and C.N. Madu (1999), Design optimization using signal-to-noise ratio. *Simulation Theory and Practice*, 7, pp. 349-372

Madu, C.N. and C.H. Kuei (1996), A metamodel of quality performances measures in an unreliable FMS, *International Journal in Computer Simulation*, 6, no. 3, 295-308

Pierreval, H. and S. Durieux (2007), Robust simulation with a base environmental scenario, *European Journal of Operational Research*, 182, 783–793

Rao, C.R. (1959), Some problems involving linear hypothesis in multivariate analysis. *Biometrika*, 46, pp. 49-58

Taguchi, G. (1987), *System of experimental designs, volumes 1 and 2.* UNIPUB/ Krauss International, White Plains, New York

Zeigler, B.P., H. Praehofer, T.G. Kim (2000), *Theory of modeling and simulation; second edition.* Academic Press, San Diego