

Analysis of Computer Experiments with Multiple Noise Sources

Christian Dehlendorff,^{*,†} Murat Kulahci and Klaus K. Andersen

In this paper we present a modeling framework for analyzing computer models with two types of variations. The paper is based on a case study of an orthopedic surgical unit, which has both controllable and uncontrollable factors. Our results show that this structure of variation can be modeled effectively with linear mixed effects models and generalized additive models. Copyright © 2009 John Wiley & Sons, Ltd.

Keywords: computer experiments; discrete event simulation; response surface

1. Introduction

Computer and simulation experiments are becoming the preferred method for analyzing systems for which physical experimentation is usually not feasible. Computer experiments are based on computer codes for which a given set of inputs generates the output(s) frequently in a deterministic manner^{1, 2}. Therefore in the analysis of computer experiments, interpolation models such as Kriging are used to guarantee the zero prediction error at the data points^{3–5}. In some applications, however, the outcome is stochastic. In stochastic simulation models, for example, a seed controls a random number stream and changing the seed results in different outcomes. There are also applications where the factors can be separated into two groups as ‘controllable’ and ‘uncontrollable’ based on their characteristics in the physical system. The uncontrollable factors could, for example, be the customer arrival rate in a grocery store or the room temperature in a laboratory and the controllable factors could, for example, be the number of checkout counters. Since the uncontrollable factors cannot be controlled in the actual physical system, their input values in the simulation model have to be varied. These uncontrollable factors are different from the controllable factors and thus need to be treated differently in the analysis as well as when designing the experiments. The analysis of the uncontrollable factors is the primary focus of this paper.

Kleijnen^{3, 5} considers simulation models as a special class of computer models, which typically include one or more stochastic elements. The sources of variation are the seed controlling the random number generator and the set of uncontrollable factors that are included in the computer model to account for the environmental variations of the underlying physical system. The variation in the output from varying the seed in a simulation model originates from the embedded stochastic components such as queues, arrival processes and procedures and can be considered to correspond to the experimental error in a physical experiment. The second type of variation in simulation models comes from the changes in the uncontrollable factors. To mimic the uncertainty from the environmental factors in the physical system, the settings of the uncontrollable factors are varied in the simulation model (see Section 5). The variation associated with the uncontrollable factors is important for robustness⁶, since the results from a simulation model generally need to be reliable under different environmental settings in the actual physical system.

The functional relationship between the uncontrollable factors and the outcome is often left unspecified and considered merely as a source of variation. Kleijnen⁵ suggests for robustness analysis to summarize the mean and variance for each controllable factor settings and model them by two separate second-order polynomials. Another approach is to model the functional relationship between the outcome and the uncontrollable factors. This may unveil the important uncontrollable factors. But more importantly it may unveil important interactions between controllable and uncontrollable factors, which may then be used to set the system in a more robust operating mode.

In this paper the sources of variation are quantified by means of a linear mixed effects (LME) model to separate the variation into a component corresponding to changes in the uncontrollable factor settings and a component corresponding to changes in the seed. Additionally, a generalized additive model (GAM) is used to model the functional relationship between the outcome and the uncontrollable factors, which replaces the variance components in the LME model.

Technical University of Denmark—DTU Informatics, Richard Petersens Plads, Kgs. Lyngby DK-2800, Denmark

*Correspondence to: Christian Dehlendorff, Technical University of Denmark—DTU Informatics, Richard Petersens Plads, Kgs. Lyngby DK-2800, Denmark.

†E-mail: cd@imm.dtu.dk

2. The case study

The computer model considered in this paper is a discrete event simulation model of an orthopedic surgical unit. The model simulates the individual patient's flow through the unit (illustrated in Figure 1) and has been developed in collaboration with the medical staff at the Gentofte University Hospital in Copenhagen. The unit undertakes both acute and elective (planned) surgery and performs more than 4600 operations a year. The patients come from several wards throughout the hospital, but the main sources of incoming patients are the four orthopedic wards and the emergency care unit.

2.1. Input factors

The simulation model has several noise sources; these can be separated into noise caused by variations in the uncontrollable factors and noise caused by variation in the seed. The seed controls the random number stream embedded in the simulation model and thus variations influence the embedded queues and processes and mimic the experimental error in a physical experiment. The uncontrollable factors are, for example, the arrival rate of acute patients and the cleaning time of the operating rooms (ORs). Moreover, a set of controllable factors, for example the number of ORs and the number of surgeons, influences the performance of the model. The factors in the model are summarized in Table I, which shows that the majority of the uncontrollable factors are related to resources being shared with other segments of the department and other departments of the hospital and hence might be occupied for other tasks. The outcomes from the simulation model are waiting times, patient throughput (the total number of patients treated) and the amount of overtime used on elective surgery. The simulation model is implemented in Extend⁷ and controlled from a Microsoft Excel spreadsheet with a Visual Basic for applications script.

2.2. Performance measures

As performance measures, we consider the total throughput (TT), the percentage of elective patients treated outside regular hours (EOU) and the extent of long waiting times. Often the longest waiting times are the most important ones since from the patient's perspective they are the most bothersome⁸. The waiting time distribution for the case study is highly right-skewed with a minimum of 0 min, a mean of 28 min, a 95% quantile of 51 min and a maximum of 140 min.

We suggest measuring the extent of long waiting times by the conditional value at risk (CVaR) measure⁹. The measure originates from finance as an extension of value at risk (VaR)^{10–12}. Both VaR and CVaR quantify a distribution of losses, for example, of a portfolio

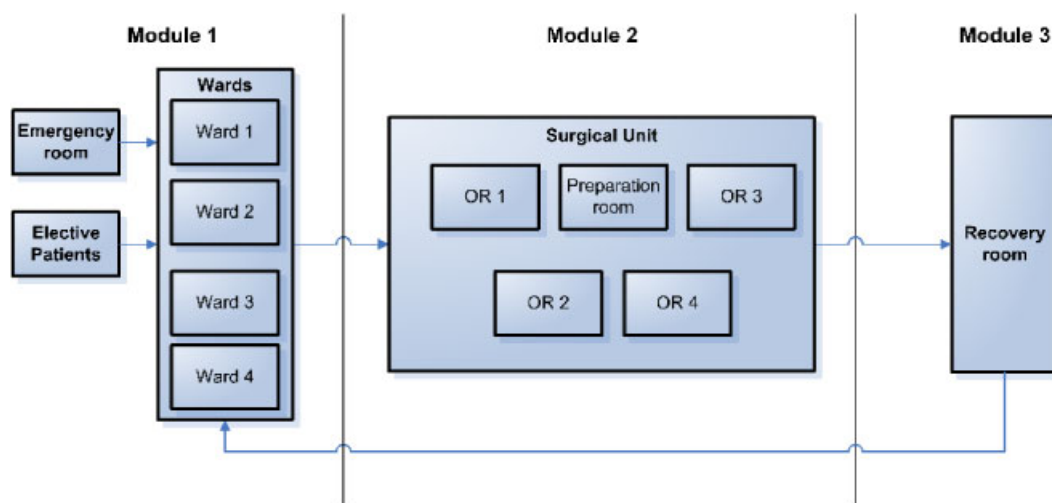


Figure 1. Basic layout of surgical unit. The patient flow is from left to right. This figure is available in colour online at www.interscience.wiley.com/journal/qre

Table I. Controllable and uncontrollable factors used in the simulation model

Controllable factors	Uncontrollable factors
Porters	Porters occupied
Elective patients	Surgeon occupied
ORs	OR cleaning time
Recovery beds	Recovery bed occupied
Cleaning teams	Cleaning teams occupied
Anesthesiologists	Anesthesiologist occupied
Operating days	Length of procedures
Acute intake	Acute arrival rate

of assets in a single statistic. For the set of waiting times $T_x = \{t_{x1}, \dots, t_{xN}\}$ from the x th run, $\text{CVaR}_\alpha(T_x)$ is defined as the expected value of the α -tail distribution of T_x ¹⁰, i.e.

$$\text{CVaR}_\alpha(T_x) = \frac{\left(\frac{i_\alpha}{N} - \alpha\right) t_{xi_\alpha} + \sum_{i=i_\alpha+1}^N \frac{t_{xi}}{N}}{1 - \alpha} \quad (1)$$

with $t_{x1} \leq t_{x2} \leq \dots \leq t_{xN}$, i_α is the index satisfying $i_\alpha/N \geq \alpha > (i_\alpha - 1)/N$, t_{i_α} is the α -quantile and in finance denoted as the value at risk (VaR). CVaR can be seen as a compromise between the average waiting time ($\alpha = 0$) and the maximum waiting time ($1 - 1/N < \alpha < 1$), where α reflects the weight put on the longest waiting times in the sample: A high α implies fewer waiting times used in the statistic and hence more weight on the longest waiting times. In the following $\alpha = 0.95$ is used so that CVaR is the average of the 5% longest waiting times.

The two other outcomes, TT and EOUT, are quality measures. They are required to fulfill the following quality constraints:

1. At least the same number of patients treated compared to the reference setting;
2. The percentage of elective patients treated outside regular hours compared with the reference setting may not increase;

where the reference setting corresponds to the current setting (see Section 5), i.e. corresponding to the performance under the current resource allocation at the department. The requirements are constraints that ensure that a performance improvement in terms of CVaR is not obtained by treating fewer patients or generating more overtime by treating more patients outside regular hours. In this study, we focus on estimating the size of the variations in CVaR and from which sources they arise. Moreover, we want to analyze the possibility of lowering CVaR while fulfilling the quality requirements.

3. Modeling framework

As mentioned earlier, the output from the simulation model is stochastic with two types of noise coming from the uncontrollable factors and the seed controlling the random number stream. The Kriging framework often used in the analysis of computer experiments is seen not to be well suited in our case, since the output is non-deterministic. There are further complications, since in our case study most controllable factors are discrete and thus interpolation is not necessarily appropriate. The presence of uncontrollable factors implies that the factors fall in two groups with different interpretations. In this study the focus is on the uncontrollable factors and we treat the controllable factor settings as a single factor. As initial model a linear model is considered as

$$y(x_{ci}, x_{ej}, s_k) = \beta_i + \varepsilon_{jk} \quad (2)$$

where β_i is the effect of controllable setting x_{ci} and $\varepsilon_{jk} \sim N(0, \sigma_\varepsilon^2)$ the residual variation. x_{ci} is the i th controllable factor setting, x_{ej} the j th environmental factor setting and s_k the seed in the k th replicate. The model has parameters for each controllable factor setting and a single error term covering the variation due to both the seed and the uncontrollable factor setting.

The linear model does not estimate the variations related to the uncontrollable factors and the seed separately. To target both types of variations explicitly a LME model is proposed¹³. The LME is formulated such that it quantifies the two sources of variation by estimating the variance component for each in the following model:

$$y(x_{ci}, x_{ej}, s_k) = \beta_i + E_j + S_k \quad (3)$$

β_i is the effect of controllable setting i , $E_j \sim N(0, \sigma_E^2)$ is the random effect of the j th uncontrollable factor setting and $S_k \sim N(0, \sigma_S^2)$ is the variation corresponding to the seed. The model is estimated by restricted maximum likelihood estimation (REML) as described in Venables and Ripley¹⁴.

The LME model quantifies the variation corresponding to varying the settings of the uncontrollable factors in a single term. It is estimated in the variance component σ_E^2 . An alternative approach is to model the functional relationship between y and each of the uncontrollable factors. These functional relationships can, for example, be estimated using a GAM¹⁵. In this modeling framework, the effects of the uncontrollable factors are modeled as non-parametric smooth additive functions and the resulting model is given as

$$y(x_{ci}, x_{ej}, s_k) = \beta_i + \sum_{l=1}^m f_l(x_{ej}^l) + S_k \quad (4)$$

with x_{ej}^l being the j th setting for the l th uncontrollable factor and $S_k \sim N(0, \sigma_S^2)$ the residual or seed term. f_l is a spline-based smooth function with the smoothness determined by a penalty term. By estimating the functional relationship between the uncontrollable factors and the outcome, the uncontrollable factors that are needed to be tightly controlled may be identified. But more importantly, interactions between controllable and uncontrollable factors may also be estimated. The estimation of the β 's and the smooth functions can, for example, be done with the R-code provided by Wood^{16, 17}.

The fraction of runs not fulfilling the quality requirements can also be analyzed within the GAM framework. For the output y_q , $q \in \{\text{CVaR}, \text{TT}, \text{EOU}\}$ and the quality requirement c_q , the outcome is binary, $I(y_q < c_q)$ (1 if fulfilled and 0 if not). A GAM with a binomial distribution family is considered with the linear predictor given as

$$E \left[\log \left(\frac{p}{1-p} \right)_{ij} \right] = \beta_i + \sum_{l=1}^m f_l(x_{e_l}^l) \quad (5)$$

where p is the fraction of runs not meeting the requirements for a given controllable factor setting.

The advantage of using the GAM framework is the employment of the smooth functions, which for example implies that a potential complex effect of an uncontrollable factor can be easily presented graphically. Moreover, the GAM does not impose a parametric form on the functional relationship except for the spline-based functions and the additivity, which implies that the data decide the model. Another advantage is that the controllable factor effects can be corrected for the effect of the uncontrollable factors. The disadvantage of the GAM framework is the additivity assumption, which in this paper implies that only marginal effects are considered. It is possible to expand the GAM to include functions of more than one variable and interactions with, e.g., controllable factors, which could potentially lead to rather complex models.

4. Example

To illustrate our modeling framework presented in Section 3, we consider a simple queuing-system operating in one of two modes: M/M/1 (single server) or M/M/2 (2 servers working in parallel). The M/M/1 (M/M/2) queue consists of a single arrival process with Poisson arrivals and one (two) server(s) with exponential service times. The arrival rate, the service rate and the number of servers are denoted by λ , μ and m , respectively. In our single server system the service time is defined to be approximately half as long as the service time of the servers in the two server system, which corresponds to the server utilization, $\rho = \lambda / (m\mu)$, being constant for fixed λ .

We consider the expected waiting time in the queue, W_q , as the performance parameter of the system. The expected waiting time is known to be

$$W_q = \begin{cases} \frac{\lambda}{\mu(\mu - \lambda)} = \frac{\rho^2}{\lambda(1 - \rho)}, & m = 1 \\ \frac{\lambda^2}{\mu(4\mu^2 - \lambda^2)} = \frac{2\rho^3}{\lambda(1 - \rho^2)}, & m = 2 \end{cases} \quad (6)$$

Instead of considering μ directly, we use ρ . On log-scale the expected waiting time in the queue is given as

$$\log(W_q) = \begin{cases} -\log(\lambda) + 2\log(\rho) - \log(1 - \rho), & m = 1 \\ -\log(\lambda) + \log(2) + 3\log(\rho) - \log(1 + \rho) - \log(1 - \rho), & m = 2 \end{cases} \quad (7)$$

The advantage of considering the expected waiting time on log-scale is that it provides a more interpretable model that separates λ from ρ . Another advantage is that it gives a more symmetric distribution of the output, which would be the argument for transforming the data if the true model was not known. In the following we set $LW_q = \log(W_q)$ for ease of notation. We treat m as a controllable factor, and λ and ρ as uncontrollable factors, since it is deemed possible to control the number of servers but not the average arrival nor the service rates. The difference in the waiting time for $m = 2$ vs $m = 1$ is $LW_q(2) - LW_q(1) = \log(\rho) - \log(1 + \rho) + \log(2)$.

4.1. Design

A simulation model that can operate as both an M/M/1 and an M/M/2 queue is implemented in Extend⁷. Each run of the simulation model is run for 20 000 min where the first 10 000 min are used as warm up period to ensure that the waiting time is stabilized. Moreover the seed controlling the random number generator is changed before each run, which makes the simulation model stochastic.

Two experimental plans are constructed; one for each setting of m . Each experimental plan consists of a uniform design with 2 factors (λ , ρ) and 100 runs. We use uniform designs since they are robust against model misspecification and do not rely on a certain model structure¹⁸. The uncontrollable factor region is given as the rectangle spanned by the intervals $\lambda \in [0.67, 1]$ and $\rho \in [0.48, 0.72]$ corresponding to varying the uncontrollable factors 20% around their average values. The simulation model takes $\mu = \lambda / (m\rho)$ as input value, but the design and analysis are done for ρ . To estimate the variation related to the random seed, five replications are taken for each combination of m , λ and ρ , which in total gives 1000 runs.

4.2. Results

The LM, LME and GAM models defined in Section 3 are used to model the LW_q values obtained from the simulation model. The parametric part of the models is given as

$$LW_q = \beta_0 + \beta_1 I(m = 2) \quad (8)$$

Table II. Summary for modeling LW_q -results from queuing system				
Model	σ_E	σ_S	β_0 (SD)	β_1 (SD)
LM		0.48	0.11(0.02)	-0.30(0.03)
LME	0.47	0.08	0.11(0.05)	-0.30(0.07)
GAM		0.08	0.11(0.01)	-0.30(0.01)

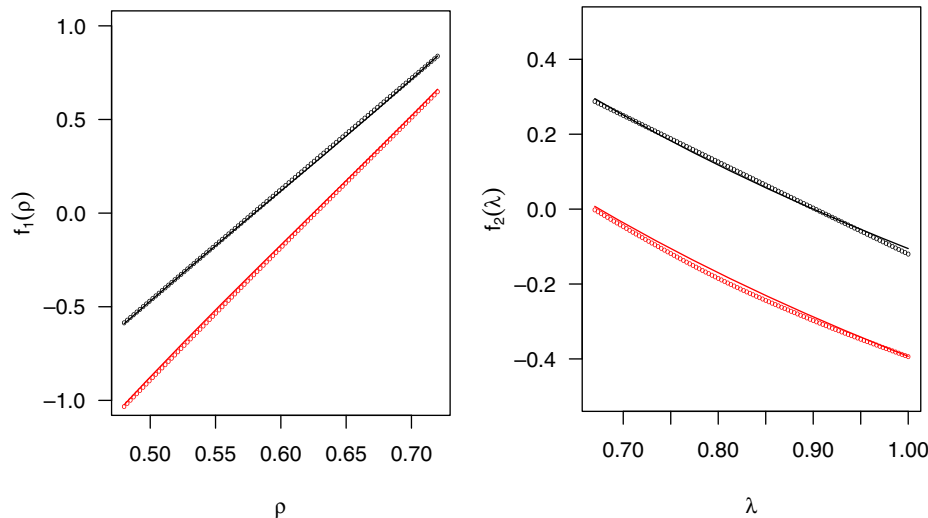


Figure 2. Estimated partial effects of ρ and λ on LW_q . Lines indicated with 'o' are estimated partial effects of ρ and λ on LW_q , solid lines are the theoretical partial effects. For both ρ and λ , the top curves correspond to $m=1$ and the bottom curves to $m=2$. This figure is available in colour online at www.interscience.wiley.com/journal/qre

where $I()$ is the indicator function. ρ and λ are included in the GAM model on their original scale with a smoother for each m , yielding the following combined model:

$$LW_q = \beta_0 + \beta_1 I(m=2) + f_1(\lambda) I(m=1) + f_2(\lambda) I(m=2) + f_3(\rho) I(m=1) + f_4(\rho) I(m=2) \quad (9)$$

where the smooth functions are expected to be $f_1(\lambda) = f_2(\lambda) = -\log(\lambda)$, $f_3(\rho) = 2\log(\rho) - \log(1-\rho)$ and $f_4(\rho) = \log(2) + 3\log(\rho) - \log(1+\rho) - \log(1-\rho)$. In the LME model each combination of m , ρ and λ corresponds to one level of E_j .

Table II summarizes the parameters of the models for LW_q . The estimates for the LME model show that the residual variation in the LM model for LW_q mostly consists of variation caused by varying the uncontrollable factors. The residual variation in the LM-model is split into a main component corresponding to the variation related to the uncontrollable factors and a minor component corresponding to the variation in the seed in the LME model. The residual variance in the GAM is seen to be the same as in the LME model, which indicates that the variation related to the uncontrollable factors is modeled adequately by the smooth function. Moreover, it is seen that the sum of the variance component in the LME models is comparable with the total variation in the linear model.

The estimated partial effects of ρ and λ on LW_q are shown in Figure 2 with the corresponding theoretical partial effects superimposed. It can be seen that the effects of the uncontrollable factors are close to the theoretical values of the effects. For λ , some minor deviations from the expected functions are seen and the two estimated curves are not perfectly parallel. The smoothed and theoretical curves are tightly superimposed, since a simple simulation model is used and the outcome is additive. The difference between the estimated effect of λ for one and two servers is, however, insignificant, whereas the difference for ρ is highly significant. The model explains more than 98% of the variation in the data and the residual variation is seen to be 0.08^2 compared with the LW_q values varying from -1.82 to 1.62.

The estimates of the parameters do also coincide with the true values. Together the models provide insight on the properties of the two queuing system, by using no prior information. In the following section, we return to the case study given in Section 2 and apply the proposed approach to model the CVaR waiting times.

5. Case study continued

For the case study given in Section 2, the average computer time needed for simulating 6 months of operation (with 1 week of warm-up) is around 7 min. For each run the system's performance is summarized in a set of measures, e.g. the total patient throughput, the

percentage of elective patients treated outside regular hours and the CVaR waiting time. Two experimental designs are considered:

1. The current controllable factor setting corresponding to the current setup simulated with
 - (a) four hundred different uncontrollable factor settings chosen such that the ranges of the eight uncontrollable factors are uniformly covered,
 - (b) two repetitions with different seeds for each setting of the uncontrollable factors, i.e. a total of $N=800$ runs,
 - (c) the combined design is denoted by D_C .
2. Twenty new controllable factor settings, which were found in a pilot study with the objective of finding good settings in terms of reducing the predicted CVaR waiting time while maintaining the performance on TT and EOUT. Each setting is simulated under
 - (a) twenty different uncontrollable factor settings chosen from the 400 run uniform design with 8 factors considered in the reference design,
 - (b) five repetitions under different seeds for each uncontrollable and controllable factor combination, i.e. a total of $N=2000$ runs,
 - (c) the combined design is denoted by D_N .

The sub-designs (the designs for the uncontrollable factors used for a certain setting of the controllable factor) considered in D_N are generated such that all 400 settings are assigned to one controllable factor setting each. This is done by first constructing a uniform design with 400 runs, then assigning each run to a whole plot (a combination of the settings of the controllable factors) such that all runs are assigned and each whole plot has 20 runs. The uniformity of the design is measured by the wrap-around discrepancy as suggested by Fang *et al.*¹⁸. Likewise the optimal construction of the sub-designs is achieved through the assignment of runs that minimize the maximal value of the wrap around values of the sub-designs. The main benefit of the design is that more uncontrollable factor settings can be tried compared with a crossed design, which is often used in applications with controllable and uncontrollable factors. This gives a better coverage of the uncontrollable factor space. For more detail, see Dehlendorff *et al.*¹⁹.

The analysis here is focused on the output from D_N if not stated otherwise. The results from D_C serve as baseline. The outputs from both designs are shown in Figure 3 and are seen to be similar for TT and EOUT. The CVaR waiting times are seen to be shorter

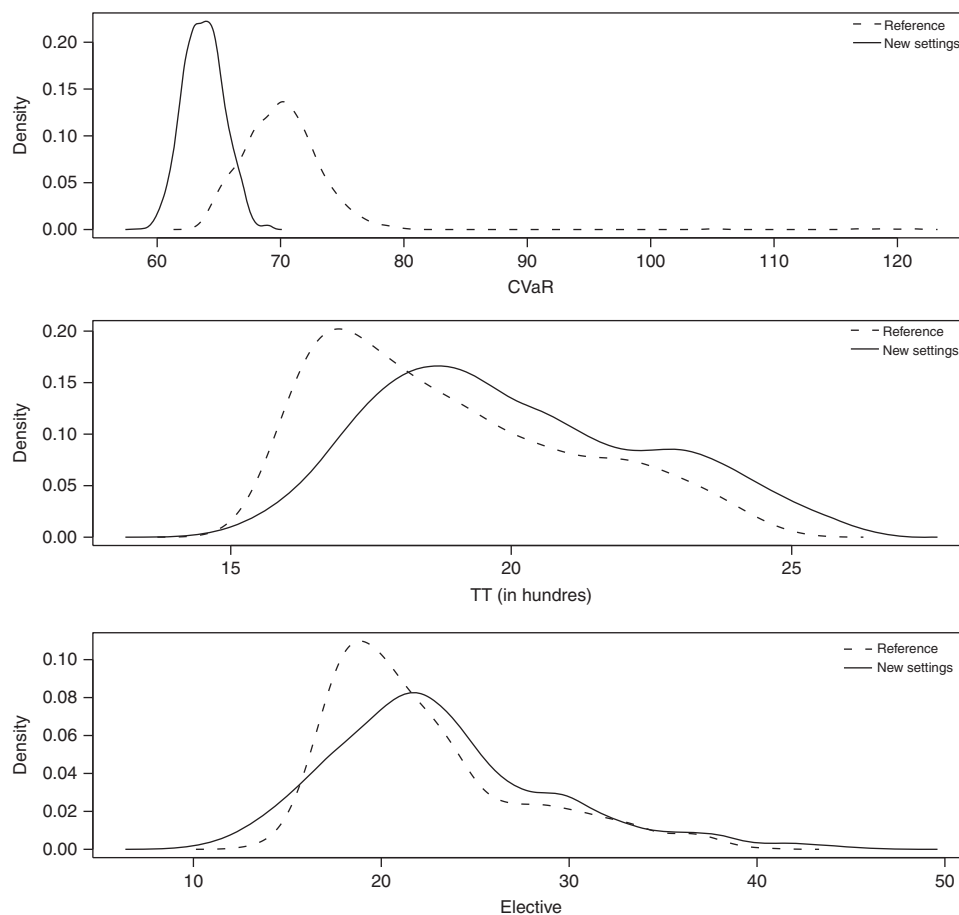


Figure 3. Estimated densities for CVaR (top), TT (middle) and EOUT (bottom) for reference design (dashed lines) and new settings (solid lines)

Table III. Estimates for models in Section 3. The variance components are summarized in σ_E and σ_S , the overall means in μ , and the adjusted R^2 in R_d^2 for the 20 suggested settings and the reference scenario for CVaR, EOUP and TT, respectively

		New settings			
Outcome	Model	σ_E	σ_S	R_a^2	μ
CVaR	LM*	—	1.63	0.02	63.77
	LME	1.17	1.16	—	
	GAM	—	1.15	0.51	
EOU [†]	LM	—	0.50	0.33	4.77
	LME	0.48	0.18	—	
	GAM	—	0.18	0.91	
TT	LM	—	223.50	0.18	2005.45
	LME	224.01	42.41	—	
	GAM	—	42.74	0.97	
Reference scenario					
CVaR	LM*	—	4.34	0 [‡]	70.23
	LME	2.19	3.74	—	
	GAM	—	3.69	0.28	
EOU [†]	LM	—	0.52	0 [†]	4.68
	LME	0.48	0.19	—	
	GAM	—	0.18	0.88	
TT	LM	—	226.80	0 [†]	1888.92
	LME	222.60	43.73	—	
	GAM	—	41.70	0.97	

*Linear regression model.

[†]The square root of the outcome is used.[‡]The linear model corresponds to the null-model since only one controllable factor setting is present.

for the new settings. Each run consists of approximately 2000 patients. Thus the CVaR waiting time becomes the average of the approximately 100 longest waiting times. The potential range for CVaR is from the 95% quantile (51 min) to the maximal waiting time (140 min). However, as the waiting time distribution is right skewed the range of the CVaR-values goes from 55 to 80 min with the exception of 4 runs.

5.1. Results

Figure 3 shows the CVaR waiting times for the 20 new settings and the current settings. It can be seen that the CVaR waiting times for the new settings are shorter compared with the current setup as expected from the pilot study. Furthermore, the coefficient of variation (CV) is lower for CVaR for the new settings (CV = 2.58%) compared with the reference (CV = 6.18%). The CVs for CVaR are also seen to be lower compared with TT (CV_{cur} = 12.01% and CV_{new} = 12.23%) and EOUP (CV_{cur} = 11.03% and CV_{new} = 12.81%). For EOUP, we use the square root transformation for symmetry and consistency with Table III. The significant increase in the CV in the reference scenario for the CVaR waiting times reflects a more right-skewed distribution with observations ranging from 63.40 to 121.17 min. Without the four largest observations in the reference scenario the CV reduces to 4.13%, which is still considerably high compared with the new settings. The overall mean CVaR is estimated to be 63.77 and 70.23 min with the new and reference settings, respectively. The initial analysis suggests that the new settings give lower CVaR on average and the performance is less sensitive to changes in the uncontrollable factors.

Fitting the linear model (LM in Table III) does indicate significant differences in mean CVaR among the 20 new settings with setting 2 having the lowest CVaR waiting time. Furthermore, the variances around the means for each setting of the controllable factors show evidence of being heterogeneous with $p = 0.005$ for Bartlett's test of variance homogeneity. From Table III it can also be seen that the residual variation in the reference scenario is 7 times higher compared with the new settings, which indicates that the new settings are more robust against changes in the uncontrollable factors.

5.2. LME

The REML estimates of the variance components in the LME analysis of the CVaR waiting times are also included in Table III. The two components for CVaR are seen to be comparable in size for D_N and a bootstrapped 95% confidence band²⁰ for the intraclass

correlation²¹ gives $0.46 \leq \sigma_E^2 / (\sigma_E^2 + \sigma_S^2) \leq 0.55$. For the reference setup, the variance components of the CVaR are seen to be significantly larger. This shows that not only is the current setup inferior to the proposed setups on average, but it also tends to be more sensitive to changes in the uncontrollable factors and the seed. The total reduction in variance with the new settings compared with the reference settings is 86% with the largest relative reduction for variation corresponding to the seed being 90%.

From Figure 3 it can be seen that the CVaR waiting times in the reference scenario have a highly right-skewed distribution with 4 runs with CVaR-values greater than 100 min, whereas the remainder of the runs are contained in the interval [63.40, 79.87]. Furthermore, the four observations violate the model assumptions: $E_j \sim N(0, \sigma_E^2)$ and $S_k \sim N(0, \sigma_S^2)$. Omitting the observations from the analysis gives $\sigma_E^2 = 2.36^2$ and $\sigma_S^2 = 1.67^2$, which means an increase in σ_E^2 and a decrease in σ_S^2 with the average also decreasing from 70.23 to 70.00. The reduction in total variation without the four observations from the reference settings to the new settings is 68%. The diagnostics after omitting the observations do not indicate problems with the model assumptions. It is seen that the current setting (with or without the four observations) is more sensitive to changes in the uncontrollable factors. The sizes of the variance components for TT and EOUT are equivalent for the two experiments, whereas the sample means are higher with the new settings.

5.3. GAM

To identify the significant uncontrollable factors, a GAM model is fitted to the CVaR waiting times. From the estimated model, it can be seen that four uncontrollable factors are significantly affecting the CVaR waiting times while the remaining four uncontrollable factors do not have an effect on the CVaR waiting times. The significant factors are the incoming rate of acute patients and the amount of time the anesthesiologists, porters and the recovery beds are occupied by other processes.

The estimated effects of the significant uncontrollable factors are shown in Figure 4. The curves fitted individually for each of the five repetitions for the new settings show that the functional form is consistent from one repetition to the next. In Table III it can also be seen that the residual variation is estimated to be $\sigma_S^2 = 1.15^2$, which matches the component from the LME model. This compared with the LME indicates that no information is lost by restricting the smooth functions to be additive. Moreover, the adjusted R^2 's show that the benefit of including the uncontrollable factors is significant with 50% or more improvements in R^2 compared with the linear model.

From Figure 4 it can be seen that the same functional relationships exist for both the current setting and the new settings for the occupancy of the anesthesiologists and the porters. The occupancy of the recovery beds has a steeper increase in CVaR in the reference settings compared with the new settings, which is likely to be caused by the fact that fewer beds are available in the reference settings. The smoothed curves for the occupancy of the recovery beds show that the new settings are more robust against variations in this factor. It can further be seen that the new settings are less sensitive to the arrival rate of the acute patients (Figure 4(a)). Moreover, it

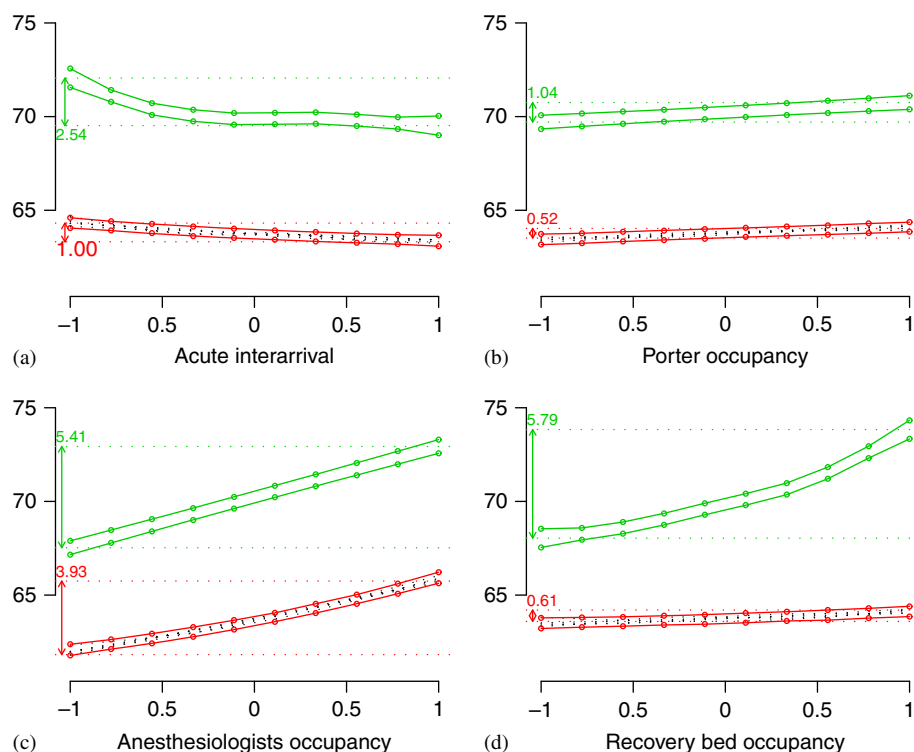


Figure 4. Estimated effects of the significant uncontrollable factors: (a) acute interarrival time; (b) amount of time porters are occupied by other procedures; (c) amount of time anesthesiologists are occupied by other procedures; and (d) amount of time the recovery beds are used for other patients. The two top curves in each of the four sub figures correspond to the 95% confidence limits for the estimated effects in the reference design. The bottom curves consist of two solid curves corresponding to a 95% confidence limits for the estimated effect in a model with all five repetitions in D_N included and five dashed curves corresponding to a model for each of the five repetitions. This figure is available in colour online at www.interscience.wiley.com/journal/qre

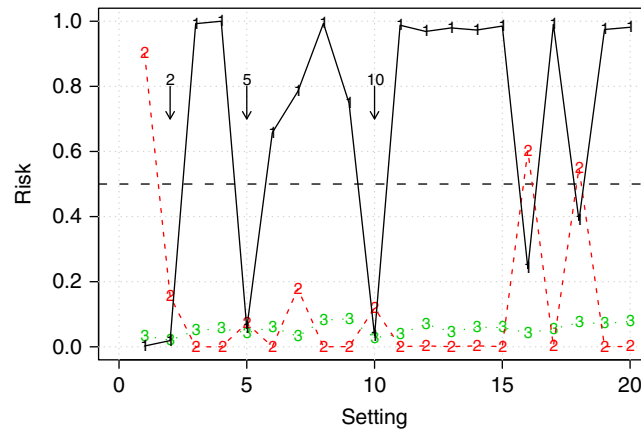


Figure 5. Risk profiles for CVaR (dotted line marked 3), TT (dashed line marked 2) and EOUT (solid line marked 1). Arrows indicate settings with risks lower than 0.5 (marked by dashed line) for TT and EOUT. This figure is available in colour online at www.interscience.wiley.com/journal/qre

can also be seen from the curves for the occupancy of the porters and the anesthesiologists that the curves for the new settings are flatter compared with the current settings. This indicates an interaction between the controllable and the uncontrollable factors, and shows that with the new controllable settings, the system is more robust against changes in the arrival rate and the occupancy of the recovery beds. Compared with Figure 2, the effect of increasing the arrival rate shown in Figure 4(a) corresponding to shortening the time between arrivals is similar to the M/M/1 and M/M/2 queues for which it also increases the waiting time.

5.4. Risk profiles

The risk profiles of CVaR, TT and EOUT for each combination of the controllable factor settings are shown in Figure 5. The risks are defined as the risk of not fulfilling the quality requirements defined in Section 2.2. In addition to the already defined requirements, we require that the new settings have a lower CVaR-value than the 5% quantile in the reference setting (65.43 min). From Table III, it can be seen that the performances in mean value and variance components are similar for TT and EOUT with the new settings compared with the baseline scenario. On average the TT is 6% better (higher) in the new settings, whereas EOUT is 2% worse (higher). This implies that it can be expected that meeting the requirement for EOUT will be more challenging.

The risks are estimated with the model in Equation (5), which estimates the effect of the uncontrollable factors on the linear predictor with smooth functions. For the risks corresponding to TT and EOUT, it can be seen that settings 2, 5 and 10 perform well for both measures. It can also be seen that the TT and EOUT risks are negatively correlated with Spearman's ρ equal to -0.89 , i.e. that lowering the risk of treating too few patients increases the risk of treating more elective patients outside regular hours. Settings 2, 5 and 10 are quite similar, that is they operate with 4 operating days, 4 ORs and an increase in elective patients by 2 per day. The three settings use more resources compared with the current setup with the lowest additional costs for setting 5. It can be seen that settings 2, 5 and 10 on average fulfill all the requirements in more than 80% of the runs. Compared with the reference setting, the most interesting difference in the controllable factors is the use of 4 operating days compared to 5 as in the current setting.

6. Conclusion

In this paper, we present the analysis of a simulation model with two types of variations due to changing seed and changes in the settings of the uncontrollable factors. The usefulness of using a GAM and a linear mixed model was illustrated by a theoretical queuing system, which showed that the suggested modeling framework performed equally well for the well-known queuing systems. The analysis for our case study shows that the variation in the CVaR waiting time with an LME model can be split into two equally large variance components for a set of new settings, whereas the seed variance in the reference scenario is lower compared with the variance caused by changes in the uncontrollable factors. A GAM shows that the main source of variation for the new settings is the use of the anesthesiologist for other tasks. Moreover, the new settings eliminate the impact of one of the most important uncontrollable factors.

The use of the LME model provides additional insight on the variation related to the settings of the uncontrollable factors and the GAM identifies the most important uncontrollable factors. This may assist decision makers in constructing focused strategies for controlling the uncontrollable factors better and if possible to improve the robustness of the system. In this application, for example, to ensure a more reliable access to the anesthesiologist seemed to be beneficial. The analysis also shows that the uncontrollable factors interacted with the controllable factors. Given the new settings, the system was deemed more robust to changes in the uncontrollable factors.

Moreover, specific settings of the controllable factors improved the long waiting times significantly while keeping a low risk of treating fewer patients or more patients outside regular hours. The drawback of the improvements in the CVaR waiting time was the cost of the additional resources needed. By combining the cost and performance, it may be possible to find cost-effective solutions

balancing cost and waiting time. The cost-effectiveness issue is important for further analysis as resources are a constraint. This could be done by translating waiting time into cost or by letting waiting time serve as a risk measure in a Pareto frontier analysis.

References

1. Santner TJ, Williams BJ, Notz WI. *The Design and Analysis of Computer Experiments*. Springer: Berlin, 2003.
2. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Statistical Science* 1989; **4**(4):409–423.
3. Kleijnen JP. Kriging metamodeling in simulation: A review. *European Journal of Operational Research* 2009; **192**(3):707–716.
4. Martin JD, Simpson TW. Use of kriging models to approximate deterministic computer models. *AIAA Journal* 2005; **43**(4):853–863.
5. Kleijnen JP. *Design and Analysis of Simulation Experiments*. Springer: Berlin, 2008.
6. Sanchez SM. Robust design: Seeking the best of all possible worlds. *Proceedings of the 2000 Winter Simulation Conference*, Orlando, FL, U.S.A., 2000; 69–76.
7. Kral D. The extend simulation environment. *Proceedings of the 2002 Winter Simulation Conference*, San Diego, CA, U.S.A., 2002; 205–213.
8. Bielen F, Demoulin N. Waiting time influence on the satisfaction–loyalty relationship in services. *Managing Service Quality* 2007; **17**(2):174–193.
9. Dehlendorff C, Kulahci M, Merse S, Andersen KK. Conditional value at risk as a measure for waiting time in simulations of hospital units. *Quality Technology and Quantitative Management* 2009; accepted. URL: <http://www.cc.nctu.edu.tw/~qtqm/>.
10. Rockafellar RT, Uryasev S. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 2002; **26**:1443–1471.
11. Kibzun A, Kuznetsov E. Comparison of var and cvar criteria. *Automation and Remote Control* 2003; **64**(7):153–164.
12. Kibzun AI, Kuznetsov EA. Analysis of criteria var and cvar. *Journal of Banking and Finance* 2006; **30**(2):779–796.
13. Pineiro J, Bates D. *Mixed Effects Models in S and S-PLUS*. Springer: Berlin, 2000.
14. Venables WN, Ripley BD. *Modern Applied Statistics with S*. Springer: Berlin, 2002.
15. Hastie T, Tibshirani R. *Generalized Additive Models*. Chapman & Hall, CRC: London, Boca Raton, 1990.
16. R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria, 2007. ISBN 3-900051-07-0.
17. Wood S. *Generalized Additive Models—An Introduction with R*. Chapman & Hall, CRC: London, Boca Raton, 2006.
18. Fang KT, Li R, Sudjianto A. *Design and Modeling for Computer Experiments*. Chapman & Hall, CRC: London, Boca Raton, 2006.
19. Dehlendorff C, Kulahci M, Andersen KK. Designing simulation experiments with controllable and uncontrollable factors. *Proceedings of the 2008 Winter Simulation Conference*, Miami, FL, U.S.A., 2008.
20. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. Chapman & Hall: London, 1993.
21. Montgomery DC. *Design and Analysis of Experiments* (7th edn). Wiley: New York, 2009.

Authors' biographies

Christian Dehlendorff is a PhD student in Informatics and Mathematical Modeling at the Technical University of Denmark. He has an MSc in Engineering with data analysis and statistics. His research interests include design of experiments and computer experiments.

Murat Kulahci is an Associate Professor in Informatics and Mathematical Modeling at the Technical University of Denmark. His research interests include design of experiments, statistical process control, and financial engineering. He is a member of the American Statistical Association, European Network of Business and Industrial Statistics (ENBIS), and the Institute of Operations Research and the Management Sciences.

Klaus K. Andersen is an Associate Professor in Informatics and Mathematical Modeling at the Technical University of Denmark. He has a PhD in Time Series Analysis and his research interests include design of experiments and statistical consulting.