DESIGN EXPERIMENTS FOR THE CONSTRUCTION OF SIMULATION METAMODELS

Pedro Reis dos Santos
Department of Computer Science and Engineering
Instituto Superior Técnico (TULisbon)
Av. Rovisco Pais, 1049-001 Lisboa, PORTUGAL
Email: pedro.reis@tagus.ist.utl.pt

M. Isabel Reis dos Santos
Department of Mathematics, Center for Management Studies
Instituto Superior Técnico (TULisbon)
Av. Rovisco Pais, 1049-001 Lisboa, PORTUGAL
Email: maria.santos@tagus.ist.utl.pt

KEYWORDS

Metamodels; Experimental Design; Replications; Regression.

ABSTRACT

Metamodels are used as analysis tools for solving optimization problems or as surrogates used as building blocks in larger scale simulations. The metamodel replaces the simulation model by a simplified input-output relationship, frequently a mathematical function with customized parameters. The construction of a metamodel is based on the simulation results for a set of design points. In order to collect statistical information each design point may be simulated repeatedly. This paper explores the precision of the resulting metamodel based on the tradeoff between higher number of design points versus a higher number of replications at a smaller number of design points.

INTRODUCTION

Simulation models are built as replacements of real systems, and are used to assist in the analysis and optimization of those systems. The resulting model can be simulated for new inputs, giving clues about the expected behaviour of the real system under untested conditions. Furthermore, a model can be repeatedly simulated for the same inputs using different random number seeds. These replications provide valuable information about the expected uncertainties of the real system.

A simulation metamodel is a model based on a simulation model. The metamodel represents the simulation model, and thus the real system, through a simplified representation of the relationship between the inputs and outputs (Davis and Bigelow, 2003). Thus a metamodel purports only to represent the behaviour of the simulation model in an input-output sense. A straightforward method to construct a metamodel is to regress a polynomial function based on the input-output values collected from a set of design points (Kleijnen and Sargent, 2000; Santos and Santos, 2009). This polynomial function is regarded as a highly abstracted version of the reality. A simplified low-resolution representation of the system enables the simulation analyst to quickly comprehend the overall behaviour. More elaborate metamodels use radial basis functions (Jin et al., 2000), Bayesian approaches (Chick, 2004) or rational metamodels (Hendrickx and Dhaene, 2005). For a more realistic global response fitting, nonlinear metamodels including nonlinear regression (Santos and Nova, 2006), Kriging (Kleijnen and van Beers, 2004), and neural networks (Hurrion, 1992) are used for complex behaviours since they provide more flexible fitting.

The precision of the constructed metamodel relies heavily on the set of design points, or experimental design, used to exercise the simulation model. If those points are carefully selected, the collected results from the simulation experiments will expose the true behaviour of the system (Santos and Santos, 2008). Since simulation runs are expensive, it is desirable to infer the system's behaviour from a small number of design points. The statistical nature of the underlying model implies that a single run may not capture the mean expected behaviour of the system in regard to some entity of interest. In fact, only a number of experiments under the same input conditions expose the expected variance of a given entity. An accurate value for the statistical aspects of some entity must rely on a large number of experiments. As stated above, simulation runs are expensive and only a small number of replications can be performed at each design point.

The construction of a metamodel should require a small number of simulation runs in order to be an accessible tool. On the other hand, the information gathered must be sufficiently precise in order to produce a credible metamodel. This paper explores the precision of the resulting metamodel based on the tradeoff between higher number of design points versus a higher number of replications at a smaller number of design points. The next Section explores the construction of metamodels, specifically linear regression metamodels. The 'experimental design' Section discusses the selection of design points for the metamodel's construction. The Section entitled 'metamodel validation' is reserved for techniques to evaluate the fitting of each metamodel and to enable the comparison of previously estimated metamodels. In the 'construction tradeoff' Section the computational effort necessary to develop a metamodel is discussed. Two regions, a low variance and a high variance, of the classical M/M/1 numerical example are used to collect simulation data and built metamodels based on experimental designs using different numbers of replications

Proceedings 23rd European Conference on Modelling and Simulation ©ECMS Javier Otamendi, Andrzej Bargiela, José Luis Montes, Luis Miguel Doncel Pedrera (Editors)
ISBN: 978-0-9553018-8-9 / ISBN: 978-0-9553018-9-6 (CD)

and design points. In the 'conclusions' Section the obtained results are analysed and summarized.

METAMODEL FITTING

The relationship between the output w and an infinite number of inputs $\mathbf{v} = v_1, v_2, \ldots$ of the real system, where some inputs may be stochastic, can be represented by a transfer function f

$$w = f(\mathbf{v}) \tag{1}$$

A simulation model of a real system can be represented by

$$y = g(\mathbf{z}, \mathbf{r}) \tag{2}$$

where y is the response, $\mathbf{z} = (z_1, \ldots, z_k)^T$ is a vector of input variables and $\mathbf{r}$ is a vector of random numbers or randomly selected seeds of the pseudo-random number generators. This simulation model attempts to describe the relationship f through a few inputs. The output y of the simulation model represents the output w of the real system. The simulation program can be represented by a empirical function g.

Assume that the simulation model can be represented by a metamodel. Specifically, the general linear regression simulation metamodel

$$y_{ij} = \theta_0 + \sum_{l=1}^{q-1} \theta_l x_{il} + \epsilon_{ij} \tag{3}$$

Denoting ϵ_{ij}, where $\epsilon_{ij} \sim \mathrm{N}(0, \sigma_i^2)$ with $\sigma_i > 0$, as the additive error term correspondent to the j-th replication of the experimental point i of the response. The unknown q metamodel parameters θ_l ($l = 0, \ldots, q-1$) must be estimated. The explanatory variable x_l, or input l, may be the same as the simulation variable z_l, or a transformation of one or more variables z_l; for example, in the $M/M/1$ system, the utilization factor $x_1 = z_1/z_2 = \rho$, with $z_1 = \lambda$ (arrival rate) and $z_2 = \mu$ (service rate) may be a better explanatory variable than z_1 and z_2 separately (Kleijnen, 2008). When using polynomial metamodels, the linear regression model is also applied. For example, if $\tilde{x}_{i1} = x_{i1}$ and $\tilde{x}_{i2} = x_{i1}^2$ in the regression model

$$y_{ij} = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i1}^2 + \epsilon_{ij} \tag{4}$$

then we obtain $y_{ij} = \theta_0 + \theta_1 \tilde{x}_{i1} + \theta_2 \tilde{x}_{i2} + \epsilon_{ij}$ which is linear in the parameters θ_0, θ_1, and θ_2.

Fitting the regression metamodel to the individual responses y_{ij} is equivalent to fitting the metamodel to the averages $\bar{y}_{i\cdot} = \sum_{j=1}^{r_i} y_{ij}/r_i$ with weights r_i; see (Kleijnen, 1987, p. 195). Consequently, instead of problem (3), the equivalent least squares problem is considered in which the individual observations, at each design point, are replaced by their averages across simulation runs

$$\bar{y}_{i\cdot} = \theta_0 + \sum_{l=1}^{q-1} \theta_l x_{il} + \bar{\epsilon}_{i\cdot} \tag{5}$$

with $\bar{\epsilon}_{i\cdot} \sim \mathrm{N}(0, \sigma_i^2/r_i)$.

In the estimation of the parameter vector $\boldsymbol{\theta}$, we considered the weighted least squares method that yield the estimator

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{V}_{\bar{Y}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_{\bar{Y}}^{-1} \bar{\mathbf{Y}} \tag{6}$$

where $\mathbf{V}_{\bar{Y}} = \mathrm{diag}(\sigma_1^2/r_1, \sigma_2^2/r_2, \ldots, \sigma_n^2/r_n)$ is a $n \times n$ diagonal matrix, and $n \geq q = rank(\mathbf{X})$. $\mathbf{X} = (x_{il})$ denotes the $n \times q$ matrix of explanatory regression variables with x_{il} the value of explanatory variable l (i.e., input l) in design point i and $x_{i0} = 1$. If $\bar{\epsilon}_{i\cdot} \sim \mathrm{N}(0, \sigma^2/r_i)$ with $\sigma > 0$, then

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \bar{\mathbf{Y}}, \tag{7}$$

with $\mathbf{W} = \mathrm{diag}(r_1, r_2, \ldots, r_n)$. If the variance of the response is approximately constant for all the design points, the estimator yields

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \bar{\mathbf{Y}} \tag{8}$$

In most applications the variances are unknown and $\mathbf{V}_{\bar{Y}}$ must be estimated using the classical estimators

$$\hat{\sigma}_i^2 = \frac{1}{r_i - 1} \sum_{j=1}^{r_i} (y_{ij} - \bar{y}_{i\cdot})^2 \tag{9}$$

for estimating the n different variances σ_i^2 (Kleijnen and van Groenendaal, 1992). The predicted response at experimental point i for the resulting metamodel is given by

$$\hat{y}_i = \hat{\theta}_0 + \sum_{l=1}^{q-1} \hat{\theta}_l x_{il} \tag{10}$$

EXPERIMENTAL DESIGN

The selection of input values, or experimental design, is critical to provide a good fit. An experimental design D corresponds to the values of the inputs over an experimental region $R \subset \mathbb{R}^q$ where the metamodel is useful. Each design point i in this region is designated by a specific input vector $\mathbf{x}_{i\cdot}$ which must be different from every other points. These points are chosen to efficiently investigate the relationship between the design factors and the response. Assuming that the interesting features of the simulation model may belong to any part of the experimental region, the experimental design should be based on a selection of evenly distributed points throughout the region. A careful choice of an experimental design may better expose the relation between the input and the response and require less design points. In order to improve the detail of the response new points should be added (Santos and Santos, 2008) in subregions where the input-output behaviour has more interest.

Besides the selected input vector values, the response may be affected by other factors such as bias and random

error (or noise). The estimation of the magnitude and distribution of the random error may be performed using replication, i.e. observing the response multiple times at the same inputs. The resulting sample means of the replicated responses have smaller variances than each individual response. To improve the detail of the response more replications (Kleijnen and Groenendaal, 1995) should be performed in high variance regions.

For each design point, a simulation run must be executed and the simulation output data collected. However, before the simulations are executed, a number of issues must be defined. These issues include the initial and final conditions, whether to perform terminating or steady-state simulations and the run length. When executing terminating simulations, that are run a pre-determined amount of time or until a certain event occurs, it may be necessary to censor the results if rare events are simulated. Otherwise, all the simulation output data should be gathered. The results obtained from steady-state simulations, that have no natural point of termination, may be biased by the warm-up period. In such cases, an initial data deletion must be performed to control the bias and reduce the estimation error in the parameters of the metamodel (Alexopoulos, 2006). Since a single mean value is gathered from each design point, simulations should not be longer than a batch or an independent replication. Then a simple mean value Y_i, at each design point i, is determined from gathered simulation data, for $i = 1, \ldots, n$.

METAMODEL VALIDATION

In the context of validation of simulation models, Kleijnen et al. (1998) consider a simulation model valid if the real system and the simulation model have identical real and simulated responses; see also Kleijnen et al. (2000). The validation process addresses the quality assurance of the metamodel fit. It is responsible for establishing that the metamodel closely resembles the real system and the simulation output data on which the metamodel is based. The comparison could be done using numerical statistics up to a significance level. Alternatively, the assessment could be made by using graphical plots.

The comparison can be performed with respect to the information used to build the metamodel. The available information can be analysed at three validation levels. A model adequacy relates the metamodel predicted output $\hat{y}_i$ in Equation (10) to the simulation data used to fit the metamodel $\bar{y}_{i\cdot}$ in Equation (5). A validation with respect to the simulation model uses additional model data not directly used in the metamodel estimation, y in Equation (2). Finally, the validation with respect to the real system compares the output of the metamodel with the real data w in Equation (1). The real data may have been used to built the simulation model on which the metamodel is based.

The model adequacy compares the replicated averages $\bar{y}_{i\cdot}$ at design point i used to built the metamodel with the respective output values which the metamodel is actually able to predict. The difference between the two values, at each design point, represents the error introduced by the adjustment. To estimate the unknown parameter values of the metamodel the weighted least squares method may be used. The method consists on minimizing the error sum of squares

$$\mathrm{SSE}(\boldsymbol{\theta}, \bar{y}) = \sum_{i=1}^{n} \frac{r_i}{\sigma_i^2} \left(\bar{y}_{i\cdot} - \theta_0 - \sum_{l=1}^{q-1} \theta_l x_{il} \right)^2 \qquad (11)$$

If the variances σ_i^2 / r_i are approximately equal, assessed using a test for homogeneity of variances (Conover, 1971, page 241), then the ordinary least squares method may be used. After the parameter estimation, a goodness of fit test must be evaluated in order to assess the accuracy of the resulting metamodel. The rooted mean square error statistic

$$\mathrm{RMSE}(\xi) = \sqrt{\mathrm{SSE}(\hat{\boldsymbol{\theta}}, \xi)/(n-q)} \qquad (12)$$

where $\xi = \hat{y}$, also known as the fit standard error, is used as a fit statistical measure for parametric fit. A RMSE value closer to 0 indicates a better fit.

The validation with respect to the simulation model uses data collected from the simulation runs. This data can be collected at input values not used in the metamodel estimation, called predictive data. The predictive validation may also use the RMSE statistical measure taken at the new input values with the metamodel estimated with the old data, as given in Equation (12) with $\xi = y$. When using replications, each replication response value y_{ij} can be independently used as predictive data. In this case, rather than using the averages $\bar{y}_{i\cdot}$ in Equation (11) the errors are computed for each individual y_{ij} data not directly used in the metamodel estimation, y in Equation (2),

$$\mathrm{SSE}(\hat{\boldsymbol{\theta}}, y) = \sum_{i=1}^{n} \sum_{j=1}^{r} \frac{r_i}{\sigma_i^2} \left(y_{ij} - \hat{y}_i \right)^2 \qquad (13)$$

with $\hat{y}_i$ given by Equation (10).

When data from the real system is available, the metamodel can be validated with respect to the actual data on which the simulation model was based. The RMSE, with $\xi = w$, compares the metamodel predictions at the input values from which the real data was collected, with

$$\mathrm{SSE}(\hat{\boldsymbol{\theta}}, w) = \sum_{i=1}^{n_w} \frac{r_i}{\sigma_i^2} \left(w_i - \hat{y}_i \right)^2 \qquad (14)$$

where w_i is the real value and n_w is the number of observation available from the real system.

CONSTRUCTION TRADEOFFS

The construction of a metamodel requires a significant computational effort since a large number of simulations

may be required. In fact, to obtain precise information about the behaviour of the simulation model a large number of observations may be required. Additionally, a large number of replications at each design point may be necessary to quantify the magnitude and distribution of the random error. Replications are also necessary to estimate the variance at each experimental point. If the variance is not approximately constant along the experimental region we may use estimated weighted least squares or use additional replications until $\hat{\sigma}_1^2/r_1 \simeq \ldots \simeq \hat{\sigma}_n^2/r_n$ (Kleijnen and Groenendaal, 1995; Kleijnen, 2006; Santos and Santos, 2008).

Furthermore, the length of each simulation must be determined when dealing with steady-state simulations. The length of the collected data, after the initial deletion, determines the precision of the collected data. In order to collect high precision data, longer simulation runs must be executed. However, when modeling the average behaviour of the system we may split the simulation run into batches or replications and use the observed mean response as the response value come from a large simulation. This allows us to obtain, with the same computational effort, additional statistical information that could not be obtained from one simulation run only. In short, the problem of computational effort required in metamodel building may resolve to a compromise between the number of design points and the number of replications or batches.

The aim of this paper is to check whether if the metamodel fitting is improved by an increase of the number of design points or if, on the contrary, a larger number of replications with a small number of design points produces better results. Henceforth, it is necessary to create identical conditions with respect to the computational effort used in metamodel's construction and compare the precision of the resulting matamodel. To gather meaningful conclusions it is necessary to use different experimental designs where the computational effort is successively transformed from additional replications into additional design points.

NUMERICAL EXAMPLE: M/M/1

The simple $M/M/1$ queuing system was simulated using AweSim version 3.0 and the metamodels built in MATLAB 6.5 using built-in and custom made routines.

We assume that customers arrive according to a Poisson process with a constant expected arrival rate, λ, and that service times follow an exponential distribution with an expected service time, $1/\mu$. The performance measure of interest is the expected waiting time in queue. The objective is to express this response as a function of the mean arrival rate (decision variable or input). We fix $1/\mu = 1$ resulting in a single input, the queue utilization factor $\rho = \lambda/\mu = \lambda$. The experiments were carried out in two different regions: a low utilization factor region, $.1 \leq \rho \leq .2$, and a high utilization factor region, $.8 \leq \rho \leq .9$. The low utilization factor experiment assumes a low and constant variance to construct the metamodel (low metamodel). The high utilization factor experiment is used to built two distinct metamodels, where the parameters are estimated using the ordinary least squares ($high$ metamodel) and the weighted least squares methods ($weight$ metamodel).

For each of the three metamodels (low, $high$ and $weight$) a number of equally spaced experimental designs were selected; see Table tab:design.

Table 1: Experimental designs.

design	n	r	runs
D_a	6	20	120
D_b	11	10	110
D_c	21	5	105
D_d	51	2	102
D_e	101	1	101

For each run j and for each design point i, in the low utilization factor region an initial set of 1500 observations were removed to mitigate the initial bias. The next 3000 observations were collected and averaged to produce y_{ij}. For the high utilization factor region an initial set of 6000 observations were deleted due to a longer warm-up period, and the following 3000 were collected. In fact, there were executed 20 replications at each of the 101 possible design points. Then a subset of design points and replications were selected to fulfill each experimental design requirements. For instance, the design D_a uses all 20 replications but only at design points $0.1, 0.12, 0.14, 0.16, 0.18, 0.2$, while design D_e uses only one replication at each of the 101 design points. The weight estimator $r_i/\hat{\sigma}_i^2$ at each design point i was based on all 20 replications, even when the selected design did not use all available replications, in order to obtain the best possible estimator with the available data; see Equation (9).

Figure 1 represents the averaged mean time in queue for each experimental design in the low experimental region. As expected, as the number of replications increases, the resulting mean value gets closer to the exact value but the design points are further apart. From the figure we can infer that the variance can be assumed constant, since observations can be represented within a band of almost fixed width.

Figure 2 represents the averaged mean time in queue for each experimental design in the $high$ experimental region. Unlike the low experimental region, the variance increases as the value of ρ gets closer to 0.9. Such observation leads us to introduce an estimated weighted least squares metamodel ($weight$) as well as an ordinary least squares metamodel ($high$).

The estimated metamodel values, for a second degree polynomial fitting, are depicted in Table 2. For each of these 15 fitted metamodels a set of statistical validation

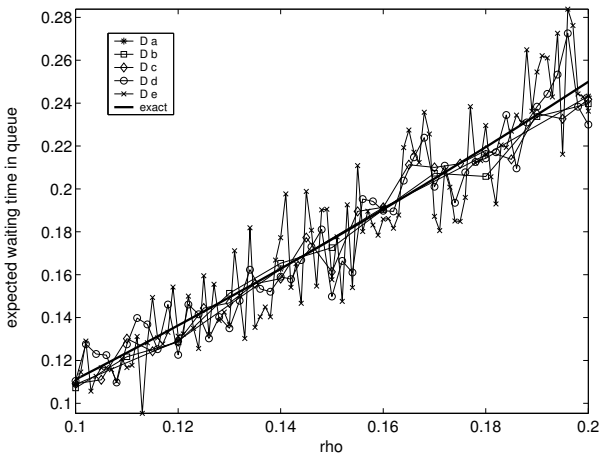

Figure 1: Waiting time in queue for the low designs.

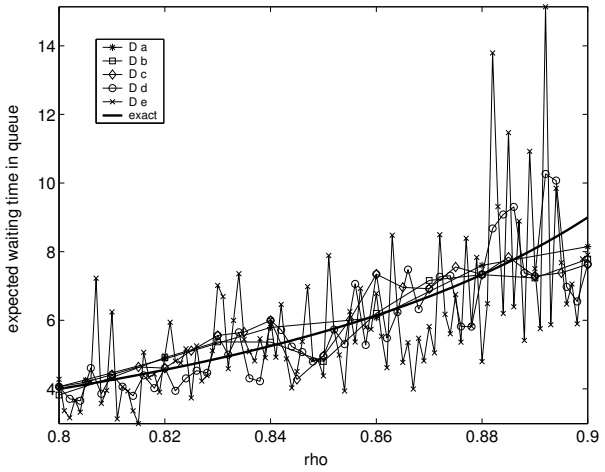

Figure 2: Waiting time in queue for the $high$ and $weight$ designs.

values were collected. Table 3 displays the adequacy statistical values of the metamodel with respect to the averaged simulation data used in the fitting process. It can be observed that, as the number of replications increases, the $\mathrm{RMSE}(\bar{y})$ shows a steady increase as the number of design points increases and the number of replications decreases at each design point. This results from fact that the interpolation polynomial yields a null $\mathrm{RMSE}(\bar{y})$, and as new design points are added the $\mathrm{RMSE}(\bar{y})$ value increases as it is the sum of non-negative error values. On the other hand, as the number of replications at each design point decreases, thus the averaged value represents a rougher approximation to the true value.

A more meaningful $\mathrm{RMSE}(y)$ value may be based on all 20 replications of all 101 design points for any of the analysed metamodels. This $\mathrm{RMSE}(y)$ predictive validation values are presented in Table 4. The values obtained are very similar for each case, leading us to conclude that the number of simulation runs are fundamental to the precision of a metamodel. The amount of gathered data obtained is the important factor, whether collected as replications or as new design points.

Table 2: Estimated values for the metamodels.

	design	metamodel
low	D_a	$-0.02324\ x^2 + 1.379\ x - 0.03117$
	D_b	$-0.9165\ x^2 + 1.619\ x - 0.04649$
	D_c	$-1.285\ x^2 + 1.751\ x - 0.05627$
	D_d	$4.402\ x^2 + 0.05655\ x + 0.06582$
	D_e	$3.71\ x^2 + 0.3559\ x + 0.03824$
$high$	D_a	$49.9\ x^2 - 43.56\ x + 7.007$
	D_b	$45.93\ x^2 - 38.91\ x + 5.688$
	D_c	$-35.08\ x^2 + 99.14\ x - 52.91$
	D_d	$209.1\ x^2 - 306.3\ x + 115$
	D_e	$248.2\ x^2 - 379\ x + 148.4$
$weight$	D_a	$43.33\ x^2 - 32.94\ x + 2.713$
	D_b	$182.8\ x^2 - 273.4\ x + 105.8$
	D_c	$93.22\ x^2 - 119.7\ x + 40.19$
	D_d	$280.2\ x^2 - 429.7\ x + 168.2$
	D_e	$191.8\ x^2 - 284.2\ x + 108.5$

Table 3: $\mathrm{RMSE}(\bar{y})$ adequacy values for the metamodels.

design	low	$high$	$weight$
D_a	0.0034	0.2928	0.0389
D_b	0.0052	0.4558	0.0942
D_c	0.0071	0.5666	0.1031
D_d	0.0114	0.8770	0.1199
D_e	0.0170	1.6320	0.2383

Table 4: $\mathrm{RMSE}(y)$ predictive validation values for the metamodels.

design	low	$high$	$weight$
D_a	0.0171	1.6562	0.2464
D_b	0.0172	1.6563	0.2384
D_c	0.0170	1.6641	0.2407
D_d	0.0171	1.6320	0.2378
D_e	0.0173	1.6341	0.2365

Since the M/M/1 queueing system has well known theoretical values, we used those values for the validation with respect to the real system. The results for $\mathrm{RMSE}(w)$ are depicted in Table 5. The results also lead us to conclude that there is no significant variation within each scenario. The concordance of the results between Table 4 and Table 5

strengthen the thesis that there is a simple tradeoff between the number of replications and the number of design points.

Table 5: $\mathrm{RMSE}(w)$ real system validation values for the metamodels.

design	low	$high$	$weight$
D_a	0.0030	0.3227	0.0597
D_b	0.0037	0.3517	0.0512
D_c	0.0028	0.3777	0.0446
D_d	0.0023	0.1377	0.0460
D_e	0.0029	0.1880	0.0318

CONCLUSIONS

Metamodels can be used as replacements for simulation models. The construction of a metamodel is based on the number of selected design points. At each design point a number of replications are executed in order to estimate the variance and collect other statistical values. The construction of a metamodel can become an expensive task if the number of simulation runs is large, as a result of the execution of many replications at several design points.

The precision of the metamodel is related to the number of design points and to the accuracy of the observed responses at each point. In this paper, we analysed the tradeoff between an increase in the number of design points at the expense of the number of replications at each point. Thus, by maintaining the simulation effort approximately constant the resulting precision of the fitted metamodel is evaluated as replications are replaced by new design points.

An M/M/1 example is used to exercise the proposed tradeoff since variances differ significantly from low to high utilization factors. The M/M/1 example is frequently used as work-bench for queueing systems due to the variability of expected behaviour, such as variance, as well as the availability of precise theoretical values. Furthermore, the M/M/1 theoretical values can be used as a reference for the validation of the metamodel with respect to the real system. Additional validation procedures are used to compare the metamodel with the simulation model and to check the accuracy of the metamodel.

The experiments carried out lead us to conclude that there is no significant increase in precision of the metamodel as the number of design point increases or decreases, as long as the total number of simulation runs is kept approximately constant. However, we conclude that it is preferable to have more replications and less design points, since it is possible to collect more meaningful statistical information without sacrificing the accuracy of the metamodel. The number of design points should be kept low, as far as it does not compromise the sampling of the simulation model response detail.

REFERENCES

Alexopoulos, C. (2006). A comprehensive review of methods for simulation output analysis. In Perrone, L., Wieland, F., Liu, J., Lawson, B., Nicol, D., and Fujimoto, R., editors, *Proceedings of the Winter Simulation Conference*, pages 168–178, Piscataway, NJ. IEEE.

Chick, S. E. (2004). Bayesian methods for discrete event simulation. In Ingalls, R., Rossetti, M., Smith, J., and Peters, B., editors, *Proceedings of the Winter Simulation Conference*, pages 89–100, Piscataway, NJ. IEEE.

Conover, W. J. (1971). *Practical Nonparametric Statistics*. John Wiley and Sons, New York, NY, USA.

Davis, P. K. and Bigelow, J. H. (2003). *Motivated Metamodels*. Rand.

Hendrickx, W. and Dhaene, T. (2005). Sequential design and rational metamodeling. In Kuhl, M., Steiger, N., Armstrong, F., and Joines, J., editors, *Proceedings of the Winter Simulation Conference*, pages 290–298, Piscataway, NJ. IEEE.

Hurrion, R. D. (1992). Using a neural network to enhance the decision making quality of a visual interactive simulation model. *Journal of the Operational Research Society*, 43(4):333–341.

Jin, R., Chen, W., and Simpson, T. W. (2000). Comparative studies of metamodeling techniques under multiple modeling criteria. Technical Report AIAA-2000-4801, AIAA.

Kleijnen, J. P. C. (1987). *Statistical Tools for Simulation Practitioners*. Marcel Dekker, New York.

Kleijnen, J. P. C. (2006). White noise assumptions revisited: Regression metamodels and experimental designs in practice. In Perrone, L., Wieland, F., Liu, J., Lawson, B., Nicol, D., and Fujimoto, R., editors, *Proceedings of the Winter Simulation Conference*, pages 107–117, Piscataway, NJ. IEEE.

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

Kleijnen, J. P. C., Bettonvil, B., and Groenendaal, W. V. (1998). Validation of trace-driven simulation models: A novel regression test. *Management Science*, 44(6):812–819.

Kleijnen, J. P. C., Cheng, R. C. H., and Bettonvil, B. (2000). Validation of trace-driven simulation models: More on bootstrap tests. In Joines, J., Barton, R., Kang, K., and Fishwick, P., editors, *Proceedings of the Winter Simulation Conference*, pages 882–892, Piscataway, NJ. IEEE.

Kleijnen, J. P. C. and Groenendaal, W. V. (1995). Two-stage versus sequential sample-size determination in regression analysis of simulation experiments. *American Journal of Mathematical and Management Sciences*, 15(1):83–114.

Kleijnen, J. P. C. and Sargent, R. G. (2000). A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research*, 120:14–29.

Kleijnen, J. P. C. and van Beers, W. C. M. (2004). Application-driven sequential designs for simulation experiments: Kriging metamodeling. *Journal of the Operational Research Society*, 55(8):876–883.

Kleijnen, J. P. C. and van Groenendaal, W. (1992). *Simulation, A Statistical Perspective*. John Wiley and Sons, New York, NY, USA.

Santos, M. I. and Nova, A. M. P. (2006). Statistical fitting and validation of nonlinear simulation metamodels: a case study. *European Journal of Operational Research*, 171(1):53–63.

Santos, M. I. and Santos, P. M. (2008). Sequential experimental designs for nonlinear regression simulation metamodels. *Simulation Modeling Practice and Theory*, 16(9):1365–1378.

Santos, P. M. and Santos, M. I. (2009). Using subsystem linear regression metamodels in stochastic simulation. *European Journal of Operational Research*, 196(3):1031–1040.

AUTHOR BIOGRAPHIES

PEDRO SANTOS was born in Lisbon and is an assistant professor in the Department of Informatics at the Technical University of Lisbon. He received a B.S. and M.S. degrees in Electrical Engineering from the Technical University of Lisbon. He received a Ph.D. in Computer Science from the Technical University of Lisbon. His research interests include simulation, especially its programming, statistical aspects, applications, and metamodels. His email is `pedro.reis@tagus.ist.utl.pt` and his personal webpage at `<http://web.tagus.ist.utl.pt/~pedro.reis>`.

ISABEL SANTOS was born in Lisbon and is an assistant professor in the Department of Mathematics at the Technical University of Lisbon. She received a B.S. and M.S. degrees in Applied Mathematics and Computation from Technical University of Lisbon, and Ph.D. degree in Industrial Engineering from Technical University of Lisbon. She is a member of the Center for Management Studies of IST. Her research interests include probabilistic and statistical aspects of simulation, applications of simulation, stochastic models and simulation metamodels. Her email is `maria.santos@tagus.ist.utl.pt` and her personal webpage at `<http://web.tagus.ist.utl.pt/~maria.santos>`.