

REINSCH'S SMOOTHING SPLINE SIMULATION METAMODELS

Pedro R. Santos

Isabel R. Santos

Department of Computer Science and Engineering
Technical University of Lisbon (IST)
Av. Rovisco Pais, 1049-001 Lisboa, PORTUGAL

Department of Mathematics, Centre for Management Studies
Technical University of Lisbon (IST)
Av. Rovisco Pais, 1049-001 Lisboa, PORTUGAL

ABSTRACT

Metamodels have been used frequently by the simulation community. However, not much research has been done with nonparametric metamodels compared with parametric metamodels. In this paper, smoothing splines for performing nonparametric metamodeling are presented. The use of smoothing splines on metamodeling fitting may provide functions that better approximate the behavior of the target simulation model, compared with linear and nonlinear regression metamodels. The smoothing splines tolerance parameter can be used to tune the smoothness of the resulting metamodel. A good experimental design is crucial for obtaining a better smoothing spline metamodel fitting, as illustrated in the examples.

1 INTRODUCTION

Simulation models are commonly used for providing insight on the behavior of particular real systems. They allow to estimate the changes on system performance produced by changes to the system design, and consequently are used for decision making. However, experimentation with these models is, frequently, time consuming, and metamodels are often constructed. A metamodel is a mathematical approximation of the observed input-output relationship defined by the more detailed simulation model. By representing the behavior of the original simulation model by a mathematical function, the metamodel captures the essential characteristics of the system (Barton 1992). Metamodels may also be used as building blocks of larger simulation models (Santos and Santos 2009). Metamodels may be parametric (for example, $Y = \theta X^2 + \epsilon$), and nonparametric in which the metamodel structure is not specified *a priori*, that is, the nature and the number of the unknown parameters are not fixed in advance. Little research has been done with nonparametric metamodels compared with parametric metamodels (Keys and Rees 2004).

The general linear metamodel has been often used because of its simple construction and utilization (Kleijnen 1975, Cheng and Kleijnen 1999, Kleijnen and Sargent 2000). The weakness of polynomial metamodels is their inability to provide a global fit to smooth response functions of arbitrarily complex shapes (Barton 1992). Some alternative metamodels have been proposed like, for example, rational metamodels (Hendrickx and Dhaene 2005) and several *nonlinear* metamodels that allow a better and more realistic global approximation of the input-output simulation behavior; for example, Bayesian metamodels (Cheng and Kleijnen 1999), Kriging metamodels (Kleijnen and van Beers 2004), nonlinear regression metamodels (Santos and Nova 2006), and neural nets (Badiru and Sieger 1998).

High order polynomial fits are non robust. Consequently, if a quadratic approximation to the input-output is inadequate, then a more accurate class of functions must be used. The drawbacks of the polynomial approach are avoid if the experimental region is divided in several small sub-regions, and on each sub-region a low order polynomial approximation is used (piecewise polynomial approximants). Consequently, an alternative for metamodeling is to use spline functions. In simulation optimization context (Daughety and Turnquist 1978) used splines for estimating, not only the optimal response from the simulation, but also for obtaining the estimated response surface over the entire experimental region of the problem.

A description of spline metamodel construction techniques is presented by (Barton 1992). The same author points out that the selection of an experimental design for fitting alternative models, such as spline

metamodels, was still a drawback in their widespread use. However, interpolating spline approximation can produce good approximations, if sufficiently accurate input-output values are available (de Boor 2001). This is the case when the simulation model output response is deterministic, or when in stochastic simulation the variance along the experimental region is not very large. Our approach is to use smoothing splines for metamodeling. This approximation method using smoothing splines is based on the tradeoff between the accuracy of the approximation at observed values, and the smoothness associated with the metamodel.

This paper addresses, in following Section, the use of smoothing spline functions as an approximation method. In Section 3 the application of the Reinsch's smoothing splines to the construction of simulation metamodels is presented. Some issues in metamodel experimental design are explained in Section 4. Two examples are used, in Section 5, to illustrate and discuss the proposed metamodeling procedure comparing with other metamodeling approaches, and using different experimental designs. The Section 6 is reserved for conclusions.

2 SPLINE ADJUSTMENT

To adequately approximate a function of various variables given only the value of the function, frequently perturbed by noise, at several points in the dependent variable interval is a problem that appear in many disciplines. The objective is to model the dependence of the response y as a function of one or more explanatory variables x_1, \dots, x_s . Suppose that the system that produced the data is described by

$$y = f(x) + \varepsilon$$

over some interval $[a; b] \in \mathbb{R}$, where ε is an additive stochastic component with zero expected value, and x is one explanatory variable. A common used approach is to fit a parametric function to the data by least squares. This approach does not permit large flexibility and is likely to generate accurate approximation functions only when the hypothetical curve is close to the true functional form.

An alternative that allow more flexibility of the approximation is to combine piecewise and local parametric fitting and roughness penalty methods (Friedman 1991). In piecewise parametric fitting the function f is approximated by simple parametric functions on each sub-interval $[x_i; x_{i+1}]$ of $[a; b]$ ($i = 1, \dots, n-1$); frequently these simple functions are low order polynomials. The approximating function f is assumed to be continuous and sometimes have continuous low order derivatives on $[a; b]$. The number of sub-intervals, and lowest order derivative that is discontinuous, allow to control the tradeoff between smoothness and flexibility. The most used piecewise polynomial fitting methods are based on splines. In this case, on each sub-interval, the approximating function is a polynomial of degree m and the derivatives of the approximating function are assumed to be continuous until order $m-1$; a common choice is $m=3$.

Splines functions are inherently better compared to polynomials for approximating input-output relationships whose behavior in a sub-region may be unrelated to the behavior in another sub-region or when the response vary rapidly. Polynomials have global dependence on local behavior, that is, the behavior in a small sub-region, of the experimental region, has a strong influence on their behavior along the entire experimental region. Splines do not have this drawback because they are piecewise low polynomial functions. The idea behind spline smoothing is to find an approximating function which fits accurately the input-output data, but also is a smooth function. The lack of accuracy is frequently defined by the residual sum of squares

$$E(f) = \sum_{i=1}^n [y_i - f(x_i)]^2$$

so the fit term is formed by a sum of squared differences between the response values and the value of the function at the predictor variable values. The function smoothness is measured by the integral of the square of a derivative of the function over the interval $[a; b]$

$$I(f) = \int_a^b [f^{(d)}(x)]^2 dx$$

The integral $I(f)$ is a functional that increases with increasing roughness of the function f .

A univariate piecewise polynomial, frequently constructed by interpolation or smoothing, is characterized by its break sequence, and its polynomial coefficients on each sub-interval. The breaks are the design points used for constructing the spline function, and it is assumed to be in a strictly increasing sequence, $X_1 < \dots < X_n$; so the number of polynomial pieces is $n-1$. The degree of the polynomial may vary along the sub-intervals,

but within each spline all polynomials are of the same order m , i.e.,

$$\theta_{i,m}x^m + \dots + \theta_{i,2}x^2 + \theta_{i,1}x + \theta_{i,0}, \text{ if } x_i \leq x < x_{i+1}, \quad i = 1, \dots, n-1$$

Consequently, a univariate piecewise polynomial may be represented by the matrix $(n-1) \times (m+1)$

$$\Theta = \begin{bmatrix} \theta_{1,m} & \dots & \theta_{1,2} & \theta_{1,1} & \theta_{1,0} \\ \theta_{2,m} & \dots & \theta_{2,2} & \theta_{2,1} & \theta_{2,0} \\ \vdots & & \vdots & \vdots & \vdots \\ \theta_{n-1,m} & \dots & \theta_{n-1,2} & \theta_{n-1,1} & \theta_{n-1,0} \end{bmatrix} \quad (1)$$

and the vector of breaks $\mathbf{x} = (x_1, \dots, x_n)^T$. When more and more design points are used, the row rank of the matrix Θ becomes higher and the class of metamodels becomes more rich, allowing the representation of the true input-output function with more and more detail. In spline modeling the number of unknown coefficients depends on the number of design points. In this case, f is a function that is smooth in some sense. A common assumption is to consider f to be twice continuously differentiable in the region of interest.

The smoothing spline functions introduced by (Schoenberg 1964) and (Reinsch 1967) has become the most frequently used spline. The Reinsch's smoothing spline f minimizes the total squared approximating function curvature

$$\int_a^b [f^{(2)}(x)]^2 dx \quad (2)$$

among all functions such that

$$\sum_{i=1}^n \left[\frac{y_i - f(x_i)}{\delta_{y_i}} \right]^2 \leq S, \quad f \in C^2[a; b], \quad (3)$$

where δ_{y_i} 's are positive weights, and $S \geq 0$ is a scaling parameter.

Smoothing spline estimators arise with (Schoenberg 1964) whose work was based on the idea of smoothing presented by (Whittaker 1923), and a detailed survey on smoothing splines can be found in (Wahba 1990). Smoothing splines was first developed in the numerical analysis context, and Wahba showed, followed by other authors, that they had useful statistical properties. As a result, they began to be widely used in data analysis and in various applications areas. In simulation context, (de Boer, Nicola, and Rubinstein 2000) used cubic smoothing splines for estimating transition probability functions in the estimation of rare-event probabilities, and (Keys and Rees 2004) used thin-plate smoothing splines for presenting a sequential simulation optimization strategy.

3 SPLINE METAMODELS

In this paper, the simulation model is treated as the following black-box

$$\mathbf{z} = \eta(\mathbf{x}; \boldsymbol{\tau}) \quad (4)$$

where only the inputs \mathbf{x} , and the outputs \mathbf{z} are observed and analyzed. The vector $\boldsymbol{\tau}$ is composed by the pseudo-random number streams that drive the simulation at \mathbf{x} ; in the black-box point of view, the values of the internal variables and other knowledge about the internal functioning, of the simulation model, are unobservable.

The combinations of the individual inputs (parameters or/and input variables) are named design points. The set of design points defines an experimental design which is critical for providing a good fit. The output may also be influenced by other factors like random error and initialization bias. The magnitude and distribution of the random error can be estimated through the observation of the output multiple times at the same design point (replication). In stochastic simulation, replication means that a given combination of inputs i is simulated r_i times using different pseudo random numbers, and the corresponding output are observed y_{i1}, \dots, y_{ir_i} . Setting $r_i = r$, the resulting sample means of the replicated responses

$$\bar{y}_i = \frac{1}{r} \sum_{j=1}^r y_{ij}$$

have smaller variances than each individual response y_{ij} . In this paper, for all design points, r simulation runs must be carried out, and the simulation outputs data collected:

$$\{y_{ij} : i = 1, \dots, n, j = 1 \dots, r\}$$

Selecting an input x and an output variable of interest y (y can be equal to some z_j or a combinations of z_j 's), let the following metamodel be an approximation of the input-output relationship that is defined by the simulation model (4):

$$\bar{y}_i = f(x; \theta) + \bar{\epsilon}_i.$$

where f is a smoothing spline function, \bar{y}_i is the output of the simulation program correspondent to i -th simulation configuration (or design point), with Y_{ij} being the replication j of the i -th design point, and where $\bar{\epsilon}_i$ is the noise due to lack of fit of the metamodel, and inaccuracy caused by the pseudo-random streams, with $\bar{\epsilon}_i \sim N(0, \sigma^2)$. The vector θ is given by the vectorization of the transpose of Θ defined in (1):

$$\theta = \text{vec}(\Theta^T) = (\theta_{1,m}, \theta_{2,m}, \dots, \theta_{n-1,m}, \dots, \theta_{1,0}, \theta_{2,0}, \dots, \theta_{n-1,0})^T$$

The estimated smoothing spline function $f(x, \hat{\theta})$ is obtained applying the Reinsch's approach (Reinsch 1967) to the simulation context. This function minimizes

$$\int_a^b [f^{(2)}(x; \theta)]^2 dx \tag{5}$$

among all functions such that

$$\sum_{i=1}^n \left[\frac{\bar{y}_i - f(x_i; \theta)}{\delta_{\bar{y}_i}} \right]^2 \leq S, \quad f \in C^2[a; b] \tag{6}$$

The weights $\delta_{\bar{y}_i}$, $i = 1, \dots, n$ control the magnitude of smoothing and are rescaled through the variation of S .

4 EXPERIMENTAL DESIGN

The experimental design should effectively explore an experimental region of an input space through a small number of design points. An experimental design is a set of input values, or design points, selected over a range of interest. The points are chosen to efficiently investigate the relationship between the input and the response.

A careful choice of an experimental design may better expose the relation between the inputs and the response. For instance, a bad experimental design may suggest a linear relation between the inputs and the response while the response is in fact nonlinear. The detection of such nonlinearities is important and can be achieved with a larger number of input values or, better still, a careful selection of input values.

Assuming that the interesting features of the metamodel can belong to *any* part of the experimental region, the design should be based on a selection of evenly distributed points throughout the region. Space-filling designs provide coverage of the entire experimental region, providing a broad exploration of the model and a valuable overview of what the response surface might look like. Whenever some parts of the experimental region exhibit more interesting characteristics, the design should include more points to explore these parts. These hybrid variations of evenly distributed and space-filling designs, may provide a better metamodel fit since more detailed information on the relation between the inputs and the response is available.

The collection of simulation data requires some tactical decisions: perform a terminating or a steady-state simulation; determine the initial conditions; choose the final conditions such as run time or number of events completed; and decide on an appropriate balance between run length and the number of replications.

Terminating simulations are those that run until a specific event have occurred (including the event of simulating a fixed amount of time). For terminating simulations, it may be necessary to censor results if we are simulating rare events. Steady-state simulations have no natural termination point, and can keep generating data for their analysis. For steady-state simulations, the warm-up period must be chosen carefully, and the length of the warm-up period affects the total experimentation time.

In steady-state simulations, the initial conditions may generate rare sequences of events, that introduce a warm-up period, until the system achieves statistical equilibrium. This warm-up creates bias in the simulation output values. For example, when all queues are empty, the simulation program begins with a sequence of

events that may have a low probability of occurring. So, the simulation output can be contaminated with an initial bias (Law 2007, Section 9.5). To control the initial bias, that can be the greatest source of error in the estimation of the metamodel parameters, an initial data deletion is performed. However, the initial data deletion selection involves a compromise, since in a small deletion the bias can still affect the result, while a large one may produce a large variance and it is not efficient. A good bias detection is important in the context of the independent replications method, because usually many runs are needed. A survey of the initialization bias problem can be found in (Pawlikowski 1990).

In this paper, we use the Welch's procedure to detect the initial bias (Law 2007). For each replication of the i th experimental point, we ignore the observations until the corresponding truncation point and collect the remaining simulation results.

The selection of the number of replication in each design point is important since the response may be affected by other factors, besides the selected input values. These factors introduce effects that can be referred as unsystematic (random error or noise) and as systematic (bias). Replication allows the estimation of the magnitude and distribution of random error, and the sample means of the replicated responses have smaller variances than the individual responses.

5 ILLUSTRATIVE EXAMPLE

To illustrate the proposed fitting procedure, two examples are presented. The first example is the classical $M/M/1$ queuing system and the second example is an automobile parts factory. All simulation experiments were simulated using AweSim version 3.0 and the metamodels built in MATLAB 7.0 using the MATLAB's routine `spaps` which is based on the Reinsch's method, and custom made routines.

5.1 Classical $M/M/1$ queuing system

In this example, the customers arrive according to a Poisson process with a constant expected arrival rate, $\lambda = 0.01$, and that service times follow an exponential distribution with a mean service time, $1/\mu$. The performance measure wherein we will concentrate is the mean time in system (response). The objective is to express the mean time in system as a function of the mean service time (decision variable). The mean service time values are between 5 and 75 minutes, corresponding to $0.05 \leq \rho \leq 0.75$, where $\rho = \lambda/\mu = 1/(100\mu)$ is the traffic intensity.

For metamodel estimation, 8 different equally spaced values for mean service time were considered:

$$\{1/\mu_i : i = 1, 8\} = \{5, 15, 25, 35, 45, 55, 65, 75\}. \quad (7)$$

For the metamodel validation, the following additional design points are used:

$$\{1/\mu_i : i = 1, 7\} = \{10, 20, 30, 40, 50, 60, 70\}. \quad (8)$$

The measure of metamodel predictive accuracy is the error sum of squares

$$\text{PSSE} = \sum_{i=1}^7 (\bar{z}_i - \hat{z}_i)^2 \quad (9)$$

where \hat{z}_i is the predicted value using the fitted metamodel based on (7), and $\bar{z}_i = \sum_{j=1}^r z_{ij}/r$ is the output of the simulation program correspondent to i -th design point in (8).

For each of the 15 design points, $r = 10$ independent replications were used. All replications started with an empty and idle system and the Welch's procedure is used for initial data deletion (Law 2007).

For estimation purpose, we used several linear regression polynomials and cubic splines functions. For each metamodel, we evaluate the error sum of squares based on the observations used on the fitting, and the predictive error sum of squares (9); see Table 1. Among polynomials, the fourth degree polynomial has the better SSE + PSSE and PSSE, although the fifth degree polynomial has a better SSE. However, using cubic splines we obtain the lowest SSE, as well as SSE + PSSE, compared with the other metamodels.

When using a polynomial a specific error sum of squares is obtained, whereas using spline metamodels, the error sum of squares may be controlled with the tolerance S . When S approaches zero, SSE tends to zero as expected from the observation of the condition (6). However, the sum SSE + PSSE does not tend to zero and exhibits a minimum value of 187.0970 when the tolerance of approximately $S = 230$ is used; see Figure 1.

Table 1: Error sum of squares for several $M/M/1$ metamodels.

Metamodel	SSE	PSSE	SSE + PSSE
second degree polynomial	1326.4	670.8715	1997.3
third degree polynomial	149.3073	148.6823	297.9896
fourth degree polynomial	103.5446	94.7659	198.3105
fifth degree polynomial*	92.4988	115.1810	207.6798
cubic spline with tolerance $S = 230$	24.8766	162.2204	187.0970

* numerical ill-conditioned fitting

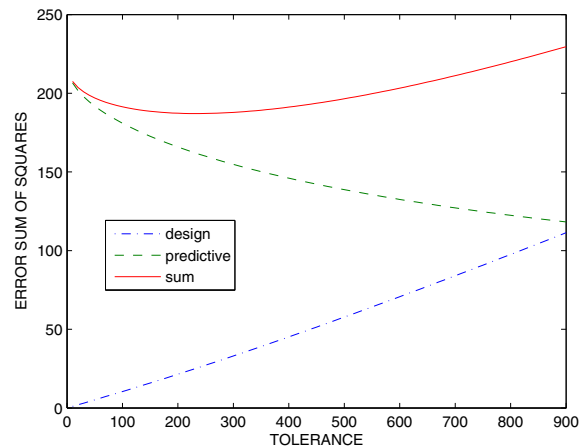


Figure 1: Error sum of squares ($M/M/1$ queuing system).

Letting $m = 3$ and a tolerance of $S = 230$, and using the matrix notation (1), the obtained spline metamodel is represented by:

$$\Theta = \begin{bmatrix} 0.000345710 & -4.44089e-17 & 1.26095 & 5.11365 \\ 6.77226e-5 & 0.0103713 & 1.36466 & 18.0688 \\ 0.00179552 & 0.0124030 & 1.59240 & 32.8203 \\ -0.00323691 & 0.0662687 & 2.37912 & 51.7802 \\ 0.0108490 & -0.0308387 & 2.73342 & 78.9614 \\ -0.00521276 & 0.294631 & 5.37135 & 114.061 \\ -0.00460828 & 0.138248 & 9.70014 & 192.025 \end{bmatrix} \quad (10)$$

5.2 Automobile Parts Factory

An automobile parts factory model drills two different types of parts and then finishes the parts. Both operations process only one part at a time. On the average, finishing operations have to be repeated 20 percent of the time. If a part has been routed through finishing twice but still needs to be refinished, it must be drilled again. The parts arrive according to $\text{gamma}(2, \lambda_1)$ and $\text{gamma}(2, \lambda_2)$, and take two and five minutes to be routed to the drill area, respectively. Drilling time is triangularly distributed between 10 and 30 minutes, with a mode of 15 minutes, while the finishing operation has a uniform distribution on the interval $[2, 5]$ minutes.

The response Y , average time in system, is expressed as a function of the mean time between arrivals of one of the parts, corresponding to $X = 2\lambda_1$ (decision variable). The other part is assumed to arrive at a constant rate $\lambda_2 = 20$. Two experimental designs were alternatively used, one for construction and the other for validation. The first is the *fixed* design which includes 19 equally spaced points:

$$\{2\lambda_i : i = 1, 19\} = \{0.5, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90\}.$$

The second is a *fill* experimental design where the points are determined by a sequential procedure described by (Santos and Santos 2008) since the response exhibits a nonlinear behavior:

$$\{2\lambda_i : i = 1, 19\} = \{0.5, 11.6875, 22.875, 28.4688, 31.2656, 31.9648, 32.6641, 33.3633, 33.5381, 34.0625, 34.3138, 35.4609, 36.8594, 39.6563, 45.25, 56.4375, 67.625, 78.8125, 90\}.$$

At each point, for both experimental designs, 10 independent replications of 1500 observations each were performed and 200 initial observations were removed to mitigate bias.

The mean response is modeled by the parametric arc-tangent function (Santos and Santos 2008) and two spline metamodels. Using both designs, two metamodels were fitted through the nonlinear least-squares using Trust-Region Reflective Newton method. The adjustments were performed by the Matlab's `fit` routine, with normalized points; see Table 2. Since there are no other controllable parameters, a better fitting can only be obtained using a different function.

Table 2: Estimated metamodels for the parametric arc-tangent function.

Design	Metamodel
<i>fill</i>	$\text{Arctan}_1 : y = 116.0 - 57.08 \text{ Arctan}(-14.91 x - 3.776)$
<i>fixed</i>	$\text{Arctan}_2 : y = 116.5 - 56.97 \text{ Arctan}(-19.83 x - 8.246)$

The splines defined in 5 and 6 can be tailored using a smooth tolerance parameter. Varying the tolerance parameter values, metamodels can be constructed to match the SSE and PSSE error of sum squares of the Arctan_1 metamodel, respectively, using the same *fill* design; see Table 3. Both metamodels present a worst fitting than the Arctan_1 as shown by the total error sum of squares $\text{SSE} + \text{PSSE}$. Nevertheless, a minimum value of $\text{SSE} + \text{PSSE}$ can be obtained with null tolerance ($S = 0$), as depicted in Figure 2. The parameters for the resulting null tolerance cubic spline₃ metamodel are the following

$$\Theta = \begin{bmatrix} -0.0038063 & 3.5404e-15 & -0.25336 & 207.45 \\ 0.0046669 & -0.051385 & -0.48459 & 205.96 \\ -0.0004535 & 0.018619 & -0.64842 & 202.84 \\ -0.0047819 & 0.011817 & -0.49623 & 200.01 \\ 0.02845 & -0.059911 & -0.73671 & 197.22 \\ -0.24952 & 0.36684 & 0.79794 & 195.6 \\ 0.40863 & -3.3759 & -14.248 & 177.57 \\ -0.15873 & 2.7535 & -17.36 & 73.011 \\ -0.028238 & 0.37259 & -1.7291 & 35.21 \\ 0.004951 & -0.050977 & -0.12103 & 32.35 \\ -0.00064488 & 0.023287 & -0.25948 & 31.089 \\ -0.001115 & 0.013614 & -0.074971 & 30.293 \\ 0.00020449 & -0.0031106 & -0.022454 & 30.12 \\ 0.00019862 & -4.3258e-5 & -0.038223 & 29.955 \\ -0.00015337 & 0.0029361 & -0.023759 & 29.788 \\ -0.00015377 & 0.00063547 & -0.0059015 & 29.723 \\ 0.00020294 & -0.0016711 & -0.01108 & 29.69 \\ -9.1535e-5 & 0.001373 & -0.01257 & 29.619 \end{bmatrix} \tag{11}$$

Table 3: Error sum of squares for metamodels fitted using the *fill* design.

Metamodel	SSE	PSSE	SSE + PSSE	Tolerance (S)
Arctan_1	161.0416	138.1521	229.1937	
spline ₁	161.0416	183.1152	344.1568	88.27195
spline ₂	100.4395	138.1521	238.5916	50.5452
spline ₃	0.0	45.0701	45.0701	0.0

When using the equally spaced *fixed* design, the minimum total error sum of squares is obtained for approximately $S = 379$; see Figure 3. However, the large gap for $2\lambda_2 = [30, 35]$ in the *fixed* design introduces oscillations in the spline₃ metamodel specially for $2\lambda_2 = [20, 30]$; see Figure 4. This results in a poorer fitting

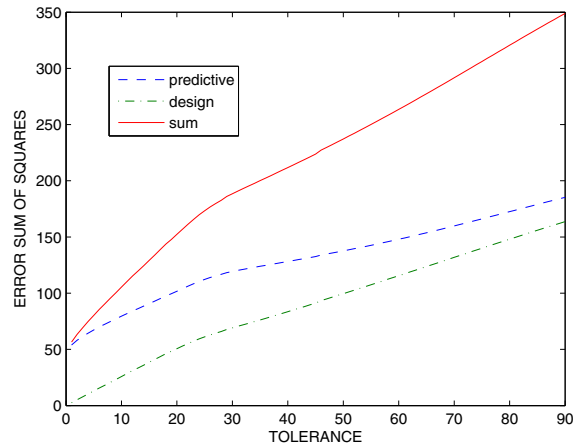


Figure 2: Error sum of squares (automobile parts factory).

than the Arctan_2 metamodel; see Table 4. Nonlinear functions provide a global adjustment where a small number of parameters need to be estimated. Any additional point, or small movement of an existing point, produces only slight changes on the estimated parameter values. Whereas, splines provide a piecewise fitting where local changes may result in significant local changes in the resulting behavior.

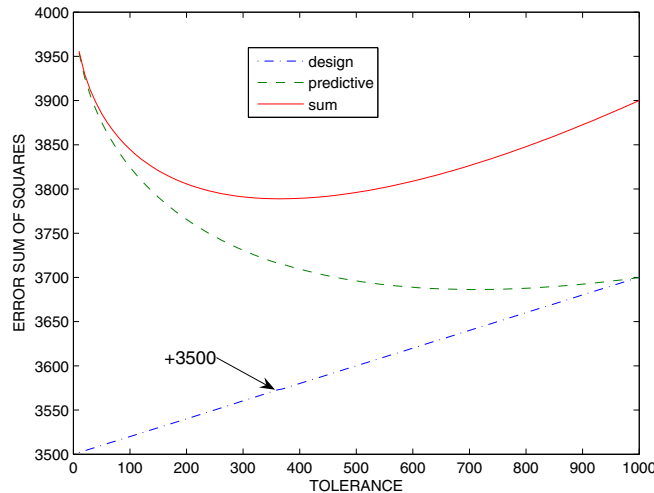


Figure 3: Error sum of squares (automobile parts factory).

As the above example depicts, the splines defined in (5) and (6) are quite sensitive to the selected experimental design. The design points should be equally spaced along the response curve in order to avoid large gaps.

6 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 and the correspondent outputs are collected. Parametric metamodels have global dependence since local behavior has a strong influence on their behavior along the entire experimental region. Spline metamodels are inherently better in approximating input-output relationships when responses vary rapidly and when the behavior in different sub-regions are unrelated. Reinsch's smoothing spline minimizes the total squared curvature of the

Table 4: Error sum of squares for metamodels fitted using the *fixed* design.

Metamodel	SSE	PSSE	SSE + PSSE	Tolerance (S)
Arctan ₂	57.01211	1427.354	1484.366	
spline ₄	74.0066	3714.9	3788.9	370

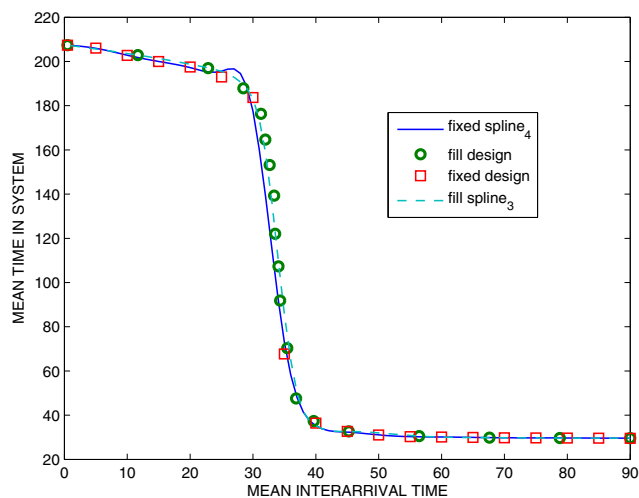


Figure 4: Spline metamodels with construction and validation design points.

spline restricted to a fixed tolerance that corresponds to the maximum error sum of squares. The tolerance parameter can be tailored to achieve a given degree of smoothness for the resulting metamodel. By controlling the tolerance parameter, the error sum of squares value (SSE) for the estimated metamodel may be minimized. However, reducing the SSE may reduce the accuracy of the metamodel's response for the values between the construction design points. To control the accuracy a set of predictive values may be used and a predictive error sum of squares (PSSE) computed. Thus, by minimizing SSE and PSSE, simultaneously, a good adjustment may be achieved. The selection of a good experimental design is essential to obtain an accurate metamodel. The procedure requires a set of evenly spaced design points along the response curve. The proposed Reinsch's smoothing spline metamodels provide a better approximation when the tolerance value that minimizes SSE + PSSE and a good experimental design are used.

REFERENCES

- Badiru, A. B., and D. B. Sieger. 1998. Neural network as a simulation metamodel in economic analysis of risky projects. *European Journal of Operational Research* 105:130–142.
- Barton, R. R. 1992. Metamodels for simulation input-output relations. In *Proceedings of the Winter Simulation Conference*, ed. J. Swain, D. Goldsman, R. Crain, and J. Wilson, 289–299. Piscataway, NJ: IEEE.
- Cheng, R. C. H., and J. P. C. Kleijnen. 1999. Improved design of queueing simulation experiments with highly heteroscedastic responses. *Operations Research* 47 (5): 762–777.
- Daughety, A. F., and M. A. Turnquist. 1978. Simulation optimization using response surfaces based on spline approximations. In *Proceedings of the Winter Simulation Conference*, ed. H. Highland, 183–193. Piscataway, NJ: IEEE.
- de Boer, P. T., V. F. Nicola, and R. Y. Rubinstein. 2000. Adaptive importance sampling simulation of queueing networks. In *Proceedings of the Winter Simulation Conference*, ed. J. Joines, R. Barton, K. Kang, and P. Fishwick, 646–655. Piscataway, NJ: IEEE.
- de Boor, C. 2001. *A practical guide to splines*. New York, USA: Springer Verlag.
- Friedman, J. 1991. Multivariate adaptive regression splines. *Annals of Statistics* 19 (1): 1–67.

- Hendrickx, W., and T. Dhaene. 2005. Sequential design and rational metamodeling. In *Proceedings of the Winter Simulation Conference*, ed. M. Kuhl, N. Steiger, F. Armstrong, and J. Joines, 290–298. Piscataway, NJ: IEEE.
- Keys, A., and P. Rees. 2004. A sequential-design metamodeling strategy for simulation optimization. *Computers & Operational Research* 31:1911–1932.
- Kleijnen, J. P. C. 1975. A comment on blanning’s “metamodel for sensitivity analysis: the regression metamodel in simulation”. *Interfaces* 5 (3): 21–23.
- Kleijnen, J. P. C., and R. G. Sargent. 2000. A methodology for fitting and validating metamodels in simulation. *European Journal of Operational Research* 120:14–29.
- Kleijnen, J. P. C., and W. C. M. van Beers. 2004. Application-driven sequential designs for simulation experiments: Kriging metamodeling. *Journal of the Operational Research Society* 55 (8): 876–883.
- Law, A. M. 2007. *Simulation modeling and analysis*. fourth ed. New York, NY, USA: McGraw-Hill.
- Pawlikowski, K. 1990, June. Steady-state simulation of queueing processes: A survey of problems and solutions. *ACM Computing Surveys* 22 (2): 123–170.
- Reinsch, C. H. 1967. Smoothing by spline functions. *Numerische Mathematic* 10:177–183.
- Santos, M. I., and A. M. P. Nova. 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 P. M. Santos. 2008. Sequential experimental designs for nonlinear regression simulation metamodels. *Simulation Modeling Practice and Theory* 16 (9): 1365–1378.
- Santos, P. M., and M. I. Santos. 2009. Using subsystem linear regression metamodels in stochastic simulation. *European Journal of Operational Research* 196 (3): 1031–1040.
- Schoenberg, I. J. 1964. Spline functions and the problem of graduation. *Proceedings of the National Academy of Sciences of the United States of America* 52 (4): 947–950.
- Wahba, G. 1990. *Spline models for observational data*. Pennsylvania, USA: SIAM.
- Whittaker, E. T. 1923. On a new method of graduation. *Proceedings of the Edinburgh Mathematical Society* 41:63–75.

PEDRO M. REIS DOS SANTOS 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 web page can be found via <http://web.tagus.ist.utl.pt/~pedro.reis>.

M. ISABEL REIS DOS SANTOS 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 web page can be found via <http://web.tagus.ist.utl.pt/~maria.santos>.