



Stochastics and Statistics

Switching regression metamodels in stochastic simulation

M. Isabel Reis dos Santos^{a,*}, Pedro M. Reis dos Santos^b^a Department of Mathematics and Centre for Management Studies, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal^b Department of Computer Science and Engineering, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

ARTICLE INFO

Article history:

Received 5 February 2015

Accepted 27 November 2015

Available online 4 December 2015

Keywords:

Simulation

Metamodel

Regression mixture

Maximum likelihood

ABSTRACT

Simulation models are frequently analyzed through a linear regression model that relates the input/output data behavior. However, in several situations, it happens that different data subsets may resemble different models. The purpose of this paper is to present a procedure for constructing switching regression metamodels in stochastic simulation, and to exemplify the practical use of statistical techniques of switching regression in the analysis of simulation results. The metamodel estimation is made using a mixture weighted least squares and the maximum likelihood method. The consistency and the asymptotic normality of the maximum likelihood estimator are established. The proposed methods are applied in the construction of a switching regression metamodel. This paper gives special emphasis on the usefulness of constructing switching metamodels in simulation analysis.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Linear regression analysis plays an important role in many fields. A regression metamodel may be used for interpreting the input/output of a simulation model and, consequently, for analysing the real world data. A simulation metamodel, usually a simple mathematical function, is an approximation of the input/output function that is defined by the underlying simulation model (Kleijnen, 2008). Kleijnen (1975) proposed some statistical tools for making the regression metamodels commonly usable, and the most popular methods for constructing simulation metamodels are the polynomial regression ones; see also (Biles, 1974). The construction and use of metamodels continues today and comprises several types of metamodels like, for example, linear regression metamodels (Kleijnen, 1992), nonlinear regression metamodels (Santos & Nova, 2006; Santos & Santos, 2008), Kriging metamodels (Kleijnen, 2009) among others. A metamodel may be used with different purposes; for example, it may be used as a surrogate of a simulation model or as a building block inside a simulation model (Santos & Santos, 2009).

However, in simulation practice sometimes we may obtain a poor fit when a single regression metamodel is used. It happens when simulation model behavior isn't likely to follow one unique regime, and that different subsets of the input/output data may favor different submodels. A different approach may be using switching regression techniques for constructing metamodels in stochastic

simulation. A switching regression model assumes that we have a random variable y such that $E[y]$ is a linear function of explanatory variables with $y \sim N(\mathbf{x}^T \boldsymbol{\theta}_s, \sigma_s^2)$ with probability λ_s , $s = 1, \dots, S$. So, there are a set of S regression models characterized by the parameters $(\boldsymbol{\theta}_1, \sigma_1^2), \dots, (\boldsymbol{\theta}_S, \sigma_S^2)$, and for each observation pair (y_i, \mathbf{x}_i) the indicator λ_{si} chooses one among several models to obtain y_i ; the unknown parameters $\boldsymbol{\theta}_1, \sigma_1^2, \dots, \boldsymbol{\theta}_S, \sigma_S^2$ are estimated from the data.

Switching regression models are dated to at least (Quandt, 1958), and find applications in a wide variety of areas such as economics (Chen, 2007; McKenzie & Takaoka, 2008), finance (Fukuda, 2009), and marketing (DeSarbo & Cron, 1988). Goldfeld and Quandt (1973) introduced the Markov-switching models, in which a latent state variable (instead of a fixed probability) following a Markov-chain controls regime shifts, meanwhile (Quandt, 1972) studies a mixture of normal linear regression models where the choice between regimes is based on fixed probabilities. In the clusterwise linear regression context, Späth (1979) considers the regression problem where the error sums of squares is computed over all regimes (referred by clusters) is minimized using an exchange algorithm. Lau, Leung, and Tse (1999) propose a programming procedure to estimate clusterwise linear regression models based on combinatorial optimization problems; see also Carbonneau, Caporossi, and Hansen (2011). Quandt (1972) proposed the maximum likelihood method for estimating switching regressions, and Kiefer (1978) studied the problem of data covering two regression regimes and maximum likelihood methods for unknown parameters estimation. In the maximum likelihood context, DeSarbo and Cron (1988) generalize the Quandt (1972) and Hosmer (1974) stochastic switching regression models to more than two regimes. This article extend these developments to the construction of simulation switching regressions metamodels, where the unknown

* Corresponding author. Tel.: +351 936166769.

E-mail addresses: isabel.santos@tecnico.ulisboa.pt, isabel.santos@ist.utl.pt (M.I. Reis dos Santos), reis.santos@tecnico.ulisboa.pt (P.M. Reis dos Santos).

Table 1
Illustrative data.

x_i	y_{ij}		
1	8.705	8.105	5.200
2	8.760	9.160	3.700
3	9.715	9.115	3.100
4	9.670	10.170	1.500
5	10.725	10.425	1.000

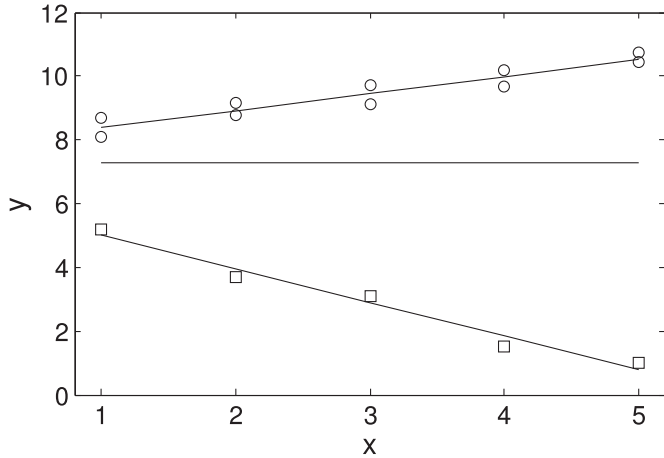


Fig. 1. Graphical representation of data from Table 1.

variances are estimated using the replications available from a simulation experiment and the switching probabilities may vary with the experimental points, and are also estimated from de data.

This paper is organized as follows. A framework of the metamodel with relation to simulation metamodels and switching regression models is described in Section 2. In Section 3 the estimation procedure for constructing switching regression metamodels is presented. Section 4 describes an application example related to a simple manufacturing process. The conclusions are presented in Section 5.

2. Simulation metamodels

A simulation model viewed as a black-box may be represented through a mathematical function $g(\cdot, \cdot)$ as

$$y = g(\mathbf{d}, \mathbf{r}_0)$$

where, \mathbf{y} is a vector of simulation outputs, \mathbf{d} is the vector of input factors of the simulation model and \mathbf{r}_0 is a vector of pseudo-random seeds. Typically, one metamodel is constructed for each component of \mathbf{y} , so we consider metamodels where \mathbf{y} has one component

$$y = f(\mathbf{x}; \boldsymbol{\theta}) + \epsilon$$

where $\boldsymbol{\theta}$ is the unknown metamodel parameters, and \mathbf{x} is a vector of metamodel inputs; for example, in the simulation of the M/M/1 system we may choose $x_1 = d_1/d_2 = \lambda/\mu$, where λ is the arrival rate and μ is the service rate. If f is a linear regression function, then a common set of regression parameters is enough for describing the input/output characteristics of the simulation program and, consequently, the simulation data may be described with only one regime. However, sometimes the input/output data exhibit some heterogeneity which produces the variation of the set of regression parameters over the data and, consequently, one regime may not be adequate for approximating the simulation input/output data. Switching regression metamodels (mixture of linear regressions) may help to overcome the lack-of-fit problem in these situations.

An illustrative example, where the usual regression leads to misleading results, is given in Table 1 and is depicted in Fig. 1. The ad-

justed metamodel based on all data points is

$$\hat{y} = 0 x + 7.270$$

and we may observe that this line poorly approximates the data. If the observations are split into two subsets then the following metamodel, which allows a good fit, is obtained

$$\hat{y} = \begin{cases} -1.06 x + 6.080, & \text{with probability } \hat{\lambda}_1 = 2/3 \text{ (subset 1)} \\ 0.53 x + 7.865, & \text{with probability } \hat{\lambda}_2 = 1/3 \text{ (subset 2)} \end{cases}$$

3. Switching regression metamodels in stochastic simulation

Consider an experimental design consisting of n different design points, $\{x_{il} : i = 1, \dots, n; l = 1, \dots, p\}$, with p explanatory variables. For each design point i , r independent replications of the simulation model are carried out and the experiment yields $\Omega_{ij} = \{\tilde{z}_{ijk} : i = 1, \dots, n, j = 1, \dots, r, k = 1, \dots, o\}$, where \tilde{z} is the relevant system response, with o observations per replication. For each experimental point i and replicate j the observations $(\tilde{z}_{ij1}, \tilde{z}_{ij2}, \dots, \tilde{z}_{ij o})$ are split into S regimes sequentially ordered:

- regime 1: $\Omega_{1ij} = \{z_{i,j,1}, z_{i,j,2}, \dots, z_{i,j,t_{1j}}\}$
- regime 2: $\Omega_{2ij} = \{z_{i,j,t_{1j}+1}, z_{i,j,t_{1j}+2}, \dots, z_{i,j,t_{1j}+t_{2j}}\}$
- regime 3: $\Omega_{3ij} = \{z_{i,j,t_{1j}+t_{2j}+1}, z_{i,j,t_{1j}+t_{2j}+2}, \dots, z_{i,j,t_{1j}+t_{2j}+t_{3j}}\}$
- ...
- regime S : $\Omega_{Sij} = \{z_{i,j,t_{1j}+t_{2j}+\dots+t_{jS-1}+1}, z_{i,j,t_{1j}+t_{2j}+\dots+t_{jS-1}+2}, \dots, z_{i,j,t_{1j}+t_{2j}+\dots+t_{jS}}\}$

where

$$\Omega_{ij} = \bigcup_{s=1}^S \Omega_{sij}$$

The probability associated with each set Ω_{sij} is estimated by

$$\hat{\lambda}_{si} = \frac{1}{r} \sum_{j=1}^r \hat{\lambda}_{sij} \quad \text{where} \quad \hat{\lambda}_{sij} = \frac{\#\Omega_{sij}}{\#\Omega_{ij}} = \frac{t_{sij}}{o} \tag{1}$$

where $\#\Omega$ represents the number of elements belonging to the set Ω .

If $\hat{\lambda}_{si} \approx 1$ for some i and $s = 1, \dots, S$, then we may assume one regime only at experimental point i . Since λ_s may depend on the experimental point, when predicting the response at \mathbf{x} between \mathbf{x}_i and \mathbf{x}_{i+1} the corresponding probabilities may be computed using, for example, interpolation of first degree. For a single input:

$$\hat{\lambda}_s(x) = \hat{\lambda}_{si} + \frac{x - x_i}{x_{i+1} - x_i} (\hat{\lambda}_{s,i+1} - \hat{\lambda}_{s,i})$$

For each regime $s = 1, \dots, S$, and replication j of each design point i a measure of interest y_{sij} is determined from $z_{i,j,k}$. For instance, the mean value for each regime is

$$y_{sij} = \frac{1}{t_{sij}} \sum_{k=k_{1ij}}^{k_{2ij}} z_{i,j,k}$$

where

$$k_{1ij} = 1 + \sum_{m=1}^{s-1} t_{mij} \quad \text{and} \quad k_{2ij} = \sum_{m=1}^s t_{mij}$$

For each experimental point i the mean and variance values of the measure of interest each regime can now be computed

$$\bar{y}_{si} = \frac{1}{r} \sum_{j=1}^r y_{sij}$$

$$\hat{\sigma}_{si}^2 = \frac{1}{r-1} \sum_{j=1}^r (y_{sij} - \bar{y}_{si})^2 \tag{2}$$

To represent the simulation model, the following mixture of linear regressions is assumed

$$\bar{y}_i = \begin{cases} \theta_{10} + \sum_{l=1}^p \theta_{1l}x_{li} + \bar{\epsilon}_{1i}, & \text{with probability } \lambda_{1i} \\ \vdots \\ \theta_{S0} + \sum_{l=1}^p \theta_{Sl}x_{li} + \bar{\epsilon}_{Si}, & \text{with probability } \lambda_{Si} \end{cases} \quad (3)$$

where $\bar{\epsilon}_{si}$ are iid as $N(0, \sigma_{si}^2/r)$, $i = 1, \dots, n$, represents the inaccuracy of the metamodel ($\sigma_{si}^2 > 0$), $s = 1, \dots, S$. Assume that the mixing weights λ_{si} , $0 < \lambda_{si} < 1$, and the regression parameters $\theta = (\theta_{10}, \dots, \theta_{1p}, \dots, \theta_{S0}, \dots, \theta_{Sp})$ are unknown, and

$$\sum_{s=1}^S \lambda_{si} = 1 \quad i = 1 \dots n$$

We assume $\lambda_{si} \neq 0$ and $\lambda_{si} \neq 1$ because if $\lambda_{si} = 0$ or $\lambda_{si} = 1$, the parameters of some regime is not identified. If the regime's probabilities are unknown, then they may be estimated by (1).

Consequently, \bar{y} has a finite mixture probability density function

$$f(\bar{y}) = \sum_{s=1}^S \frac{\lambda_s}{\sqrt{2\pi\sigma_s^2/r}} \exp \left[-\frac{1}{2\sigma_s^2/r} (\bar{y} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_l)^2 \right]$$

3.1. Mixture weighted least squares

Späth (1979) first introduced the term *clusterwise regression* as a separation technique when clustering data. Instead of a separation criterion given a set of previously known mathematical functions, the approach proposed in this paper estimates the function parameters from observations already clustered into different groups. Whereas in the traditional regression model the approximating function is the same for all subsets of the sample, that is, it assumes only one regime ($\theta_{sl} = \theta_l$, for all $s = 1, \dots, S$ and $l = 0, \dots, p$), the clusterwise model assumes a mixture of linear regressions like in (3). Späth presents an exchange algorithm to minimize the global problem

$$\sum_{i=1}^n \sum_{s=1}^S \left(\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li} \right)^2$$

In this sense, the clusterwise linear regression is a regression problem where the error sums of squares computed over all clusters is minimized. In order to assure the existence of a solution, it is required that the number of observations is significantly larger than the number of unknown parameters ($n \gg dS$).

In this paper, the following global estimated weighted error sum of squares used is

$$\sum_{i=1}^n \sum_{s=1}^S \frac{\hat{\lambda}_{si}}{\hat{\sigma}_{si}^2/r} \left(\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li} \right)^2$$

In mixture weighted least squares (MWLSQ), all regimes must be adjusted simultaneously, taking into account the relative estimated variances $\hat{\sigma}_{si}^2$ and the estimated regime probabilities $\hat{\lambda}_{si}$, obtained from the previously clustered observations.

Initial values are computed using estimated separated weighted least squares (WSLQ) method for each regime.

$$\sum_{i=1}^n \frac{1}{\hat{\sigma}_{si}^2/r} \left(\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li} \right)^2$$

Then an unconstrained nonlinear minimization is performed, using the previously determined initial values. The use of an initial value near the solution improves the robustness of the nonlinear search.

3.2. Maximum likelihood estimation

The log-likelihood function can be expressed as

$$\ln L(\lambda, \theta, \sigma^2) = \sum_{i=1}^n \ln \left\{ \sum_{s=1}^S \frac{\lambda_{si}}{\sqrt{2\pi\sigma_{si}^2/r}} \exp \left[-\frac{1}{2\sigma_{si}^2/r} (\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li})^2 \right] \right\}$$

where $\lambda = (\lambda_{11}, \dots, \lambda_{1n}, \dots, \lambda_{S1}, \dots, \lambda_{Sn})$, $\theta = (\theta_{10}, \dots, \theta_{1p}, \dots, \theta_{S0}, \dots, \theta_{Sp})$, and $\sigma^2 = (\sigma_{11}^2, \dots, \sigma_{1n}^2, \dots, \sigma_{S1}^2, \dots, \sigma_{Sn}^2)$.

Using (1) and (2), the maximum likelihood approach consists on maximizing the log-likelihood function with respect to θ , that is, consists on solving the following system of equations:

$$\frac{\partial \ln L(\hat{\lambda}, \theta, \hat{\sigma}^2)}{\partial \theta} = 0 \quad (4)$$

or equivalently,

$$\frac{\partial \ln L(\hat{\lambda}, \theta, \hat{\sigma}^2)}{\partial \theta_{st}} = 0, \quad s = 1, \dots, S, \quad t = 0, \dots, p$$

where,

$$\frac{\partial \ln L(\hat{\lambda}, \theta, \hat{\sigma}^2)}{\partial \theta_{st}} = \sum_{i=1}^n \frac{1}{V_i} \frac{\hat{\lambda}_{si}}{\sqrt{2\pi\hat{\sigma}_{si}^2/r}} \frac{x_{ti}(\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li})}{\hat{\sigma}_{si}^2/r} \exp \left[-\frac{1}{2\hat{\sigma}_{si}^2/r} (\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li})^2 \right]$$

with

$$V_i = \sum_{s=1}^S \frac{\hat{\lambda}_{si}}{\sqrt{2\pi\hat{\sigma}_{si}^2/r}} \exp \left[-\frac{1}{2\hat{\sigma}_{si}^2/r} (\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li})^2 \right]$$

In Proposition 1, the existence of a consistent root of the likelihood Eq. (4) is established; also, the asymptotic distribution of this consistent root is obtained (see the verification of these results in Appendix).

Proposition 1. Let the random variable \bar{y}_{si} has probability density function

$$f(\bar{y}_i; \theta) = \sum_{s=1}^S \frac{\hat{\lambda}_{si}}{\sqrt{2\pi\hat{\sigma}_{si}^2/r}} \exp \left[-\frac{1}{2\hat{\sigma}_{si}^2/r} (\bar{y}_{si} - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_{li})^2 \right] \quad (5)$$

where $\theta = (\theta_{10}, \dots, \theta_{1p}, \dots, \theta_{S0}, \dots, \theta_{Sp})$, with $\theta_{sl} \in \mathbb{R}$. Then for n sufficiently large there exists a unique consistent root $\hat{\theta}$ of the likelihood Eq. (4), and the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta^*)$ is multivariate normal with zero mean vector and variance-covariance matrix given by $I(\theta^*)^{-1}$, where θ^* represents the true value of the parameter θ and $I(\theta^*)$ is the Fisher information matrix.

The system of nonlinear Eq. (4) yields the maximum likelihood estimates (MLE). However, this system cannot be solved explicitly and, frequently, has non-unique solution. Consequently, numerical methods like, for example, Newton method or Nelder–Mead may be used for obtaining maximum likelihood estimates. These methods may converge to a local solution instead of the global one, especially if the initial guess is not sufficiently near the exact solution. However, Proposition 1 assures the existence of a unique global solution but, if there is more than one solution, it does not provide any information about which solution is consistent. When conventional algorithms fail in detecting the global optimum, stochastic methods like e.g. Simulated Annealing might prove to be a good alternative. Although sometimes slower, they converge to the global optimum.

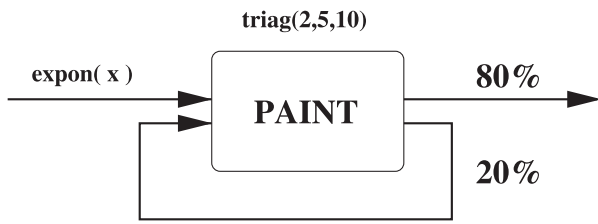


Fig. 2. Simulation model of the simplified manufacturing process.

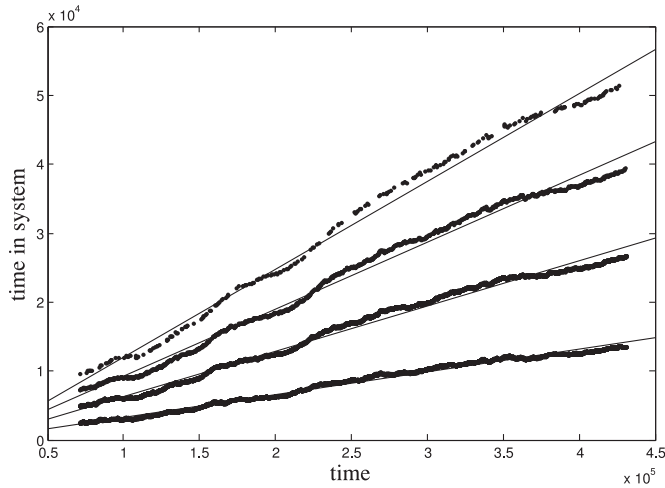


Fig. 3. Time in system growth rate (measure of interest).

4. Application example

Several systems exhibit switching output regimes. Some of these systems include a feedback component. The depicted application example is a system with feedback that exhibits output regimes in certain range of inputs. The example is a simplification of a manufacturing process where items must be reprocessed if the result does not meet specifications.

The system used for depicting the construction of the proposed metamodels is a simple parts painting processing unit. The parts to be painted arrive at the unit according to a Poisson process. The mean time between arrivals is the decision variable x . The painting time is triangularly distributed between 2 and 10 minutes, with a mode of 5 minutes. However, 20 percent of the painting operations have to be repeated due to imperfections (see Fig. 2).

The paint process is repeated up to 3 times for each part, if the previous painting operation did not meet specifications, resulting in 4 regimes. The time in system grows during the day, but the growth rate varies with the interarrival time. The purpose of the simulation experiment is to express the time in the system, z , as a function of the mean time between arrivals of parts. The metamodels were built in MATLAB 7.10 with some custom made routines. The observations z_{ijk} were gathered from 6 independent replications of 9 equally spaced design points in the [6, 6.8] experimental region. From each replication of 51000 observations, the first 1000 observations were removed to mitigate initialization transient and provide a good separation between regimes.

Every observation from each simulation run is grouped into one of four switching regimes. From each regime, a growth rate is determined by the slope of the line that is obtained by applying a least squares to the regime observations (see Fig. 3). The probability of each regime is computed from the relative number of observations of each regime. At each design point, a set of six replications are performed. The average values of six computed slopes \bar{y}_{si} . (see Table 2) are determined (measure of interest), as well as their variances $\sigma_{si} = \hat{\sigma}_{si}$

Table 2 Simulation output data for metamodel construction (response).

x_i	\bar{y}_{1i}	\bar{y}_{2i}	\bar{y}_{3i}	\bar{y}_{4i}
6.0	1.2748×10^{-1}	2.3841×10^{-1}	3.3526×10^{-1}	4.1945×10^{-1}
6.1	1.1591×10^{-1}	2.1833×10^{-1}	3.0894×10^{-1}	3.8882×10^{-1}
6.2	1.0169×10^{-1}	1.9222×10^{-1}	2.7238×10^{-1}	3.4445×10^{-1}
6.3	8.9629×10^{-2}	1.7132×10^{-1}	2.4585×10^{-1}	3.1369×10^{-1}
6.4	7.9326×10^{-2}	1.5237×10^{-1}	2.1977×10^{-1}	2.8234×10^{-1}
6.5	6.9566×10^{-2}	1.3383×10^{-1}	1.9355×10^{-1}	2.4918×10^{-1}
6.6	5.4843×10^{-2}	1.0664×10^{-1}	1.5555×10^{-1}	2.0215×10^{-1}
6.7	4.1868×10^{-2}	8.1969×10^{-2}	1.2057×10^{-1}	1.5667×10^{-1}
6.8	3.2370×10^{-2}	6.3870×10^{-2}	9.4196×10^{-2}	1.2392×10^{-1}

Table 3 Estimated variances for the simulation output data.

x_i	$\hat{\sigma}_{1i}$	$\hat{\sigma}_{2i}$	$\hat{\sigma}_{3i}$	$\hat{\sigma}_{4i}$
6.0	1.2748×10^{-1}	2.3841×10^{-1}	3.3526×10^{-1}	4.1945×10^{-1}
6.1	1.1591×10^{-1}	2.1833×10^{-1}	3.0894×10^{-1}	3.8882×10^{-1}
6.2	1.0169×10^{-1}	1.9222×10^{-1}	2.7238×10^{-1}	3.4445×10^{-1}
6.3	8.9629×10^{-2}	1.7132×10^{-1}	2.4585×10^{-1}	3.1369×10^{-1}
6.4	7.9326×10^{-2}	1.5237×10^{-1}	2.1977×10^{-1}	2.8234×10^{-1}
6.5	6.9566×10^{-2}	1.3383×10^{-1}	1.9355×10^{-1}	2.4918×10^{-1}
6.6	5.4843×10^{-2}	1.0664×10^{-1}	1.5555×10^{-1}	2.0215×10^{-1}
6.7	4.1868×10^{-2}	8.1969×10^{-2}	1.2057×10^{-1}	1.5667×10^{-1}
6.8	3.2370×10^{-2}	6.3870×10^{-2}	9.4196×10^{-2}	1.2392×10^{-1}

Table 4 Estimated mixing weights for metamodel construction.

x_i	$\hat{\lambda}_{1i}$	$\hat{\lambda}_{2i}$	$\hat{\lambda}_{3i}$	$\hat{\lambda}_{4i}$
6.0	8.2487×10^{-1}	1.4550×10^{-1}	2.5290×10^{-2}	4.3433×10^{-3}
6.1	8.2374×10^{-1}	1.4590×10^{-1}	2.5713×10^{-2}	4.6467×10^{-3}
6.2	8.2125×10^{-1}	1.4746×10^{-1}	2.6560×10^{-2}	4.7233×10^{-3}
6.3	8.1946×10^{-1}	1.4873×10^{-1}	2.6950×10^{-2}	4.8600×10^{-3}
6.4	8.1505×10^{-1}	1.5184×10^{-1}	2.7890×10^{-2}	5.2233×10^{-3}
6.5	8.1440×10^{-1}	1.5166×10^{-1}	2.8663×10^{-2}	5.2767×10^{-3}
6.6	8.1248×10^{-1}	1.5287×10^{-1}	2.8957×10^{-2}	5.7000×10^{-3}
6.7	8.1059×10^{-1}	1.5416×10^{-1}	2.9530×10^{-2}	5.7133×10^{-3}
6.8	8.0803×10^{-1}	1.5699×10^{-1}	2.9047×10^{-2}	5.9267×10^{-3}

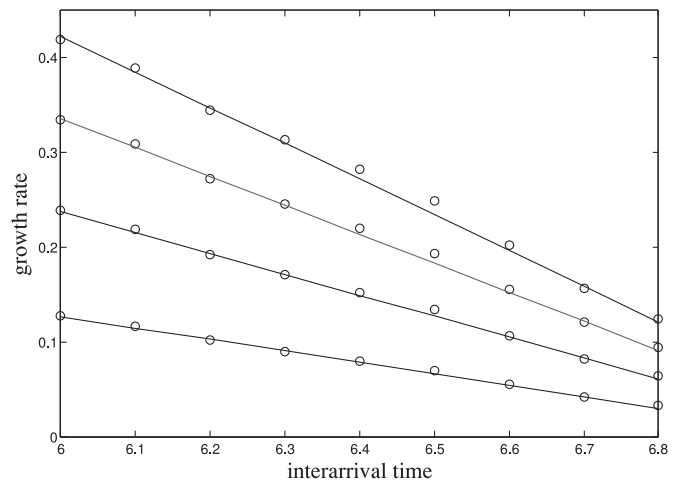


Fig. 4. Metamodel relates interarrival time with time in system growth rate.

(see Table 3) and the average probability of each regime $\bar{\lambda}_{si} = \hat{\lambda}_{si}$ (see Table 4).

The resulting metamodel relates the interarrival time with the growth rate (see Fig. 4). All four regimes must be adjusted simultaneously, taking into account the relative variances $\hat{\sigma}_{si}$ and the regime probability $\hat{\lambda}_{si}$. In the first approach, a mixture weighted least squares method is used to compute the θ values of the metamodel. In a second approach, a maximum likelihood method is applied.

Table 5
Estimated unknown parameters values.

$ax + b$	regime 1	regime 2	regime 3	regime 4
WLSQ a	0.8525389	1.5684717	2.1714385	2.6876232
WLSQ b	-0.1209539	-0.2217537	-0.3059143	-0.3774470
MWLSQ a	0.8369327	1.5570236	2.1684496	2.6730821
MWLSQ b	-0.1183631	-0.2194875	-0.3046008	-0.3739679
MLE a	0.8368609	1.5555085	2.1518431	2.6808206
MLE b	-0.1183521	-0.2192511	-0.3020517	-0.3753420

Table 6
Error sum of squares (SSE) for estimated metamodels.

WLSQ	MWLSQ	MLE
10.6968	4.4052	4.3948

The results of performing independent weighted least squares on each regime provide the initial θ values for the mixture weighted least squares and maximum likelihood estimation (see Table 5). These θ values are used as initial values for the minimization Nelder–Mead simplex method (see Lagarias, Reeds, Wright, & Wright (1998)), implemented by *fminsearch* in Matlab, also depicted in Table 5.

The adjusted values through mixture weighted least squares and maximum likelihood methods produce indistinguishable lines in Fig. 4. However, the SSE values, depicted in Table 6, provide an insight into the precision of the resulting metamodels

$$SSE = \sum_{i=1}^n \sum_{s=1}^S \frac{\hat{\lambda}_{si}}{\hat{\sigma}_{si}^2/r} (\bar{y}_{si} - f_s(x_i, \hat{\theta}_s))^2$$

where $\hat{\theta}_s = (\hat{\theta}_{s0}, \dots, \hat{\theta}_{sp})$.

If all regimes are adjusted simultaneously, a better adjustment is achieved. The maximum likelihood method provides a slightly better fit.

5. Conclusions

Metamodels provide a simple representation of the input/output relationship of a simulation model. When the output data exhibits independent groups of data, the metamodel should reproduce that grouping if a good representation of the original simulation model is to be achieved. In order to reproduce those groups, the metamodel must switch between different regimes. Each regime is modeled by a different mathematical function and switch mechanism is activated, usually a probability. The probability at each design point can be estimated using the relative frequencies of the observations. As the regimes result from the same output sequence, and are therefore dependent, they must be adjusted simultaneously. A number of simulation replications should be performed, if the variances of output are to be estimated. The metamodel construction processes the proposed fitting with a set of functions, one for each regime. Although each of the functions is usually linear, the resulting mixture is non-linear. An unconstrained minimization algorithm is employed to estimate the metamodel parameters for the set of functions. The initial values for the minimization are obtained from the independent estimation of each regime to improve the performance of the algorithm, and consequently the precision of the estimated parameters. The example presented displays four regimes that are fitted with first degree polynomial functions. The metamodels are compared using the sum of squared errors and show that a global estimation improves the quality of the fitted metamodel. The maximum likelihood method provides slightly better fitting results and the resulting estimators are consistent.

Appendix

The verification of Proposition 1 consists on applying the following Theorem stated by Chanda (1954) to switching regression.

Theorem 1. Let (u_1, u_2, \dots, u_n) a random sample such that u_i are independent and identically distributed with probability density function $f(u; \theta)$, where θ is a vector of unknown parameters. If the following conditions hold

1. For almost all u and for all $\theta \in \bar{\Omega} \subset \mathbb{R}^k$

$$\frac{\partial \ln f}{\partial \theta_r}, \frac{\partial^2 \ln f}{\partial \theta_r \partial \theta_s}, \text{ and } \frac{\partial^3 \ln f}{\partial \theta_r \partial \theta_s \partial \theta_t}, \quad r, s, t = 1, \dots, k. \quad (A.1)$$

2. For almost all u and for all $\theta \in \bar{\Omega} \subset \mathbb{R}^k$

$$\left| \frac{\partial f}{\partial \theta_r} \right| < F_r(u), \quad \left| \frac{\partial^2 f}{\partial \theta_r \partial \theta_s} \right| < F_{rs}(u)C,$$

$$\text{and } \left| \frac{\partial^3 f}{\partial \theta_r \partial \theta_s \partial \theta_t} \right| < H_{rst}(u), \quad r, s, t = 1, \dots, k. \quad (A.2)$$

where H is such that $\int_{-\infty}^{\infty} H_{rst}(u)f(u; \theta)du \leq M < \infty$ and $F_r(u)$ and $F_{rs}(u)$ are bounded for $r, s, t = 1, \dots, k$.

3. For all $\theta \in \bar{\Omega} \subset \mathbb{R}^k$, the following matrix is positive definite:

$$I(\theta) = \int_{-\infty}^{\infty} \left(\frac{\partial \ln f(u; \theta)}{\partial \theta} \right) \left(\frac{\partial \ln f(u; \theta)}{\partial \theta} \right)^T f(u; \theta) dx \quad (A.3)$$

Then, there exists a unique consistent estimator $\hat{\theta}$ which is the solution of the likelihood equations

$$\frac{\partial \ln L}{\partial \theta_j} = \frac{\partial}{\partial \theta_j} \left(\sum_{i=1}^n \ln f(u_i; \theta) \right) = 0, \quad j = 1, \dots, k$$

Moreover, $\sqrt{n}(\hat{\theta} - \theta^*)$ is asymptotic normally distributed with mean zero and variance-covariance matrix $I(\theta^*)^{-1}$, where θ^* is the true value of θ and $I(\theta^*)$ is the Fisher information matrix.

Let the density probability function (5) written as

$$f(y; \theta) = \sum_{i=1}^S f^s$$

where

$$f^s = \frac{\hat{\lambda}_s}{\sqrt{2\pi \hat{\sigma}_s^2/r}} \exp \left[-\frac{1}{2\hat{\sigma}_s^2/r} (y - \theta_{s0} - \sum_{l=1}^p \theta_{sl}x_l)^2 \right]$$

Then it may be observed that the following derivatives exist

$$\frac{\partial f}{\partial \theta} = \left(\frac{\partial f^1}{\partial \theta_{10}}, \dots, \frac{\partial f^1}{\partial \theta_{1p}}, \dots, \frac{\partial f^S}{\partial \theta_{S0}}, \dots, \frac{\partial f^S}{\partial \theta_{Sp}} \right)^T$$

$$\frac{\partial^2 f}{\partial \theta \partial \theta^T} = \begin{bmatrix} \frac{\partial^2 f^1}{\partial \theta_{10}^2} & \dots & \frac{\partial^2 f^1}{\partial \theta_{10} \partial \theta_{1p}} & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ \frac{\partial^2 f^1}{\partial \theta_{1p} \partial \theta_{10}} & \dots & \frac{\partial^2 f^1}{\partial \theta_{1p}^2} & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & \frac{\partial^2 f^S}{\partial \theta_{S0}^2} & \dots & \frac{\partial^2 f^S}{\partial \theta_{S0} \partial \theta_{Sp}} \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & \frac{\partial^2 f^S}{\partial \theta_{Sp} \partial \theta_{S0}} & \dots & \frac{\partial^2 f^S}{\partial \theta_{Sp}^2} \end{bmatrix}$$

Considering $w = (1, x^T)^T$, then, $w_1 = 1, w_2 = x_1, \dots, w_{p+1} = x_p$ the following may be written.

$$\frac{\partial f}{\partial \theta_{si}} = \frac{\partial f^s}{\partial \theta_{si}} = f^s \frac{w_{i+1}}{\hat{\sigma}_s^2/r} (y - \mathbf{w}^T \theta_s)$$

$$\frac{\partial^2 f}{\partial \theta_{si}^2} = \frac{\partial^2 f^s}{\partial \theta_{si}^2} = \frac{w_{i+1}^2}{\hat{\sigma}_s^2/r} \left[\frac{1}{\hat{\sigma}_s^2/r} (y - \mathbf{w}^T \theta_s)^2 - 1 \right] f^s$$

$$\frac{\partial^2 f}{\partial \theta_{si} \partial \theta_{kj}} = \frac{\partial^2 f^s}{\partial \theta_{si} \partial \theta_{kj}} = 0, \quad k \neq s$$

$$\frac{\partial^3 f^s}{\partial \theta_{si} \partial \theta_{kj} \partial \theta_{lt}} = 0, \quad k \neq s \text{ or } l \neq s$$

$$\frac{\partial^2 f}{\partial \theta_{si} \partial \theta_{sj}} = \frac{w_{i+1} x_{j+1}}{\hat{\sigma}_s^2/r} \left[\frac{1}{\hat{\sigma}_s^2/r} (y - \mathbf{w}^T \theta_s)^2 - 1 \right] f^s$$

$$\frac{\partial^3 f}{\partial \theta_{si} \partial \theta_{sj} \partial \theta_{sk}} = \frac{w_{i+1} w_{j+1} w_{k+1}}{(\hat{\sigma}_s^2/r)^2} (y - \mathbf{w}^T \theta_s) \left[1 + \frac{1}{\hat{\sigma}_s^2/r} (y - \mathbf{w}^T \theta_s)^2 \right] f^s$$

Consequently, the following derivatives exists and the first condition (A.1) is verified:

$$\frac{\partial \ln f}{\partial \theta_{si}} = \frac{\partial f^s}{\partial \theta_{si}} \frac{1}{f}$$

$$\frac{\partial^2 \ln f}{\partial \theta_{si} \partial \theta_{kj}} = \frac{1}{f} \frac{\partial^2 f^s}{\partial \theta_{si} \partial \theta_{kj}} - \frac{1}{f^2} \frac{\partial f^s}{\partial \theta_{si}} \frac{\partial f^k}{\partial \theta_{kj}} \tag{A.4}$$

If $k \neq s$ then

$$\frac{\partial^2 f^s}{\partial \theta_{si} \partial \theta_{kj}} = 0$$

and, as a result,

$$\frac{\partial^2 \ln f}{\partial \theta_{si} \partial \theta_{kj}} = - \frac{1}{f^2} \frac{\partial f^s}{\partial \theta_{si}} \frac{\partial f^k}{\partial \theta_{kj}}$$

Using Eq. (A.4) the following third degree derivative may be written

$$\begin{aligned} \frac{\partial^3 \ln f}{\partial \theta_{si} \partial \theta_{kj} \partial \theta_{tm}} &= 2 \frac{\partial f^s}{\partial \theta_{si}} \frac{\partial f^k}{\partial \theta_{kj}} \frac{\partial f^t}{\partial \theta_{tm}} \frac{1}{f^3} - \frac{\partial^2 f^s}{\partial \theta_{si} \partial \theta_{tm}} \frac{\partial f^k}{\partial \theta_{kj}} \frac{1}{f^2} \\ &\quad - \frac{\partial f^s}{\partial \theta_{si}} \frac{\partial^2 f^k}{\partial \theta_{kj} \partial \theta_{tm}} \frac{1}{f^2} - \frac{\partial f^t}{\partial \theta_{tm}} \frac{\partial^2 f^s}{\partial \theta_{si} \partial \theta_{kj}} \frac{1}{f^2} + \frac{\partial^3 f^s}{\partial \theta_{si} \partial \theta_{kj} \partial \theta_{tm}} \frac{1}{f} \end{aligned}$$

If k and t are both not equal to s then

$$\frac{\partial^3 \ln f}{\partial \theta_{si} \partial \theta_{kj} \partial \theta_{tm}} = 2 \frac{\partial f^s}{\partial \theta_{si}} \frac{\partial f^k}{\partial \theta_{kj}} \frac{\partial f^t}{\partial \theta_{tm}} \frac{1}{f^3}$$

The first and second derivatives of $f(\bar{y}; \theta)$ are continuous in both \bar{y} and θ . Consequently, they are bounded for $\theta \in \bar{\Omega}$ and $\bar{y} \in [a; b] \subset \mathbb{R}$.

If \bar{y} is arbitrarily large, the derivatives remain bounded because the largest of these derivatives is $O = (\bar{y}^4 \exp(-\bar{y}^2))$. Also, the largest value of the third derivatives of $\ln f$ is $O(\bar{y}^6)$. Moreover, all the moments of f exist and they are finite for all $\theta \in \bar{\Omega}$, so we may choose a constant M such that the second condition (A.2) is satisfied.

The third condition (A.3) is also satisfied because $I(\theta)$ is a variance-covariance matrix.

References

Biles, W. E. (1974). A gradient-regression search procedure for simulation experiments. In H. Highland (Ed.), *Proceedings of the Winter Simulation Conference* (pp. 490–497). Piscataway, NJ: IEEE.

Carboneau, R. A., Caporossi, G., & Hansen, P. (2011). Globally optimal clusterwise regression by mixed logical-quadratic programming. *European Journal of Operational Research*, 212, 213–222.

Chanda, K. C. (1954). A note on the consistency and maxima of the roots of likelihood equations. *Biometrika*, 41, 56–61.

Chen, S. (2007). Measuring business cycle turning points in japan with the markov switching panel model. *Mathematics and Computers in Simulation*, 76, 263–270.

DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *SIAM Journal on Computing*, 5(1), 249–282.

Fukuda, K. (2009). Distribution switching in financial time series. *Mathematics and Computers in Simulation*, 79(5), 1711–1720.

Hosmer, D. W. (1974). Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics B - Simulation and Computation*, 3(10), 995–1006.

Kiefer, N. M. (1978). Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica*, 46(2), 427–434.

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. (1992). Sensitivity analysis of simulation experiments: regression analysis and statistical design. *Mathematics and Computers in Simulation*, 34, 297–315.

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

Kleijnen, J. P. C. (2009). Kriging metamodeling in simulation: a review. *European Journal of Operational Research*, 192(3), 707–716.

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1), 112–147.

Lau, K., Leung, P., & Tse, K. (1999). A mathematical programming approach to clusterwise regression model and its extensions. *European Journal of Operational Research*, 116, 640–652.

McKenzie, C., & Takaoka, S. (2008). Underwriter reputation and switching. *Mathematics and Computers in Simulation*, 78, 215–222.

Quandt, R. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53, 873–880.

Quandt, R. E. (1972). A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338), 306–310.

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

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

Späth, H. (1979). Algorithm 39: Clusterwise linear regression. *Computing*, 22, 367–373.