Sensitivity Analysis of System Dynamics Models: Regression Analysis and Statistical Design.

Jack P.C. Kleijnen
Katholieke Universiteit Brabant
(Tilburg University)
Tilburg, Netherlands)

Abstract

This tutorial gives a survey of strategic issues in the statistical design and analysis of experiments with System Dynamics models. These models may be either deterministic or random. The strategic issues include what-if analysis and optimization. The analysis uses regression (meta)models and Least Squares. The design uses classical experimental designs such as 2^{k-p} factorials, which are efficient and effective. If there are very many inputs, then special techniques such as group screening and sequential bifurcation are useful. Some applications are discussed briefly.

Introduction

System Dynamics uses the technique of simulation to 'solve' its models. Simulation is a mathematical technique that is very popular because of its flexibility, simplicity, and realism. By definition, simulation involves experimentation, namely with the model of a real system. Consequently it requires on appropriate *design and analysis*. For real systems, mathematical statistics has been applied since the 1930s: Sir Ronald Fisher focussed on agricultural experiments in the 1930s; George Box concentrated on chemical experimentation, since the 1950s; see Box and Draper (1987). My first book (Kleijnen, 1974/1975) covered both the 'tactical' and 'strategic' issues of experiments with random and deterministic simulation models. The term *tactical* was introduced into simulation by Conway (1963); it refers to the issues of runlength and variance reduction, which arise only in random simulations such as System Dynamics studies of the effects of information quality; see the survey in Kleijnen (1980, pp. 137-143). *Strategic* questions are: which combinations of input variables should be simulated, and how can the resulting output be analyzed? Obviously strategic issues arise in both random and deterministic simulations. Mathematical statistics can be applied to solve these questions, also in deterministic simulation; see Kleijnen (1987, 1990), and Sacks et al. (1989). This contribution focusses on these strategic issues in simulation experiments with System Dynamics models.

Strategic issues are also addressed under names like *model validation*, *what-if analysis*, *goal seeking*, and *optimization*; see Table 1, reproduced from Kleijnen (1987, p. 136). We shall return to this table.

Table 1: Terminology

Computer program	System Dynamics model	Regression model	User view
Output	*Response*	*Dependent variable y*	*Result*
Input	*Parameter*	*Independent variable x*	*Environment*
	Variable		
	Enumeration	Continuous	Validation Risk Analysis
	Function	Discrete	*Controllable* Optimization
	Scenario	Binary	
			Goal output (control)
			Satisfy (what-if)
	Behavioral relationship		

Regression Metamodels

Before systems analysts start experimenting with a System Dynamics model, they have accumulated *prior* knowledge about the real system: they may have observed the real system, tried different System Dynamics models, debugged the final simulation program, and so on. This tentative knowledge is formalized in a regression or Analysis of Variance (ANOVA) model. ANOVA models are presented in the basic statistical theory on the design of experiments: Sums of Squares (SSs) are compared through the F test to detect significant main effects and interactions. The simplest ANOVA models can be easily translated into regression models; see Kleijnen (1987, pp. 263-293). Because regression analysis is more familiar than ANOVA is, we shall use regression terminology henceforth.

So prior knowledge is formalized in a *tentative* regression model. In other words, this model must be tested later on to check its validity as we shall see. The regression model specifies which *inputs* seem important, which *interactions* among these inputs seem important, and which *scaling* seems appropriate. We shall discuss these items next.

Table 1 showed that 'inputs' are not only parameters and variables but may also be 'behavioral relationships'. Parameters are quantities that are not directly observable so they must be estimated. Changing a behavioral relationship may mean that a module of the System Dynamics model is replaced by a different module. In the regression model such a qualitative change is represented by one or more binary (0,1) variables. 'Inputs' are called 'factors' in experimental design terminology. 'Interaction' means that the effect of a factor depends on the values (or 'levels)' of another factor:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{j=1}^{k} \sum_{g=j}^{k} \beta_{jg} x_j x_g + \sum_{j=1}^{k} \sum_{g=j}^{k} \sum_{h=g}^{k} \beta_{jgh} x_j x_g x_h + \dots + e, \qquad (1)$$

where y is the response of a simulation run; β_0 is the overall or grand mean; β_j is the main or first-order effect of factor j; β_{jg} is the two-factor interaction between the factors j and g ($g \neq j$); β_{jj} is the quadratic effect of factor j; β_{jgh} is the three-factor interaction among the factors j, g, and h ($h \neq g \neq j$) and so on; e denotes 'fitting errors' or noise. Under certain mathematical conditions the 'response curve' in Eq. (1) is a Taylor series expansion of the System Dynamics model $y(x_1,\dots,x_k)$. Unfortunately these conditions do not hold in System Dynamics. Therefore we propose to start with an initial model that excludes interactions among three or more factors: such high-order interactions are popular in ANOVA but they are hard to interpret. The purpose of the regression model is to guide the design of the simulation experiment and to interpret the resulting simulation data; a regression model without high-order interactions suffices, as we observed repeatedly in practice.

The regression variables x in Eq. (1) may be *transformations* of the original System Dynamics parameters and variables; for example, x_1 may equal $\log(z_1)$ where z_1 denotes the original System Dynamics input. *Scaling* is also important: if the lowest value of z_1 corresponds with $x_1 = -1$ and its highest value corresponds with $x_1 = +1$, then β_1 measures the relative importance of factor 1 when that factor ranges over the experimental area.

In optimization through *Response Surface Methodology* (RSM) we explore response curves locally. The local regression model is the first-order model:

$$y = \gamma_0 + \sum \gamma_j z_j + e, \qquad (2)$$

where the importance of factor j at $\bar{z}_j$, the midpoint of the local experiment, is measured by $\gamma_j \bar{z}_j$ with $\bar{z}_j = \sum_{i=1}^{n} z_{ij}/n = (L_j+H_j)/2$ where $L_j \leq z_{ij} \leq H_j$ with local experimental area $[L_1,H_1] \times \dots \times [L_k,H_k]$; z_{ij} denotes the value of factor j in simulation run or observation i. See Bettonvil and Kleijnen (1990), and Box and Draper (1987).

In all experiments, analysts use models such as Eq. (1), explicitly or implicitly. For example, if they change one factor at a time (as, for example, Wolstenholme, 1990 does), then (implicitly) they assume that all interactions (β_{jg}, β_{jgh},...) are zero. Of course it is better to make the regression model explicit and to find a design that fits that model, as we shall see next. But first note that we call the regression model a *metamodel* because it models the input/output behavior of the underlying System Dynamics model; that model is treated as a black box in our approach.

Experimental Design

Based on a tentative regression metamodel we select an experimental design. The design matrix D = (d_{ij}) specifies the n combinations of the k factors that are to be simulated. (In multi-stage experimentation such as RSM this set of n combinations is followed by a next set.) Classical statistical theory gives designs that are 'efficient' and 'effective'. *Efficiency* means that the number of factor combinations or simulation runs is 'small'. Suppose there are Q effects in the regression metamodel. The number of runs should then satisfy the condition $n \geq Q$; for example, we need k + 1 runs if there are no interactions at all. Table 2 shows that we may observe one base run (say) (-1, -1,..., -1), and then change one factor at a time: (+1, -1,..., -1), (-1, +1, -1,..., -1),..., (-1,..., -1, +1). To estimate the effects $\beta' = (\beta_0, \beta_1, \dots, \beta_k)$ we fit a curve to the simulation data

(X,y) where the first-order model implies X=(1,D); 1 denotes a vector of n ones. The classic fitting criterion is *Least Squares*. This criterion yields the effects estimator.

Table 2: Two designs for three factors.
(- denotes -1; + denotes +1)

	One at a time			2^{3-1} Design		
Run	d_1	d_2	d_3	d_1	d_2	d_3
1	-	-	-	-	-	+
2	+	-	-	+	-	-
3	-	+	-	-	+	-
4	-	-	+	+	+	+

$$\hat{\beta} = (X'X)^{-1} X'y. \quad (3)$$

Now consider the classic fractional factorial 2^{3-1} design of Table 2. It is easy to check that the corresponding X is orthogonal. Hence Eq. (3) reduces to the scalar expression

$$\hat{\beta}_{j'} = \sum_{i=1}^{n} x_{ij'} \cdot y_i/n \qquad (j'=0,1,\ldots,k). \quad (4)$$

How can we choose between the two designs of Table 2? Classical theory assumes that the fitting errors e are *white noise*: e is normally and independently distributed with zero mean and constant variance (say) σ^2. Then Eq. (3) yields the variance-covariance matrix

$$\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}. \quad (5)$$

It can be proved that an orthogonal matrix X is *'optimal'*. Actually, there are several optimality criteria; see Federov (1972) and Kleijnen (1987, p. 335). An orthogonal X minimizes $\text{var}(\hat{\beta}_{j'})$, the elements on the main diagonal of Eq.(5). There are straightforward procedures for deriving 'good' design matrices, in case n equals 2^{k-p} with (p=0,1...); for other n values there are tables and software; see Box and Draper (1987), and Kleijnen (1987).

So the classical designs are efficient under the white noise assumption (recent research uses alternative assumptions; see Sacks et al., 1989). Moreover, these designs are *effective*: they permit the estimation of interactions. If we allow for two-factor interactions, then the number of effects Q increases to 1 + k + k(k-1)/2. If k is small, we may simulate $n \geq Q$ combinations; for example, if k = 5 then a 2^{5-1} design is suitable. (If k is large, we may hope that some factors will turn out to give nonsignificant main effects; we may assume that factors without main effects have no two-factor interactions either; there are designs with n = 2k that yield unbiased estimators for main effects; see Kleijnen, 1987, pp. 303-309, and Bettonvil and Kleijnen, 1990). If the factors are quantitative, then a second-order regression model includes k quadratic effects. In such a model, n must increase and more than two levels per factor must be simulated: Box and Draper (1987) and Kleijnen (1987) give such RSM designs.

Screening

For didactic reasons we discuss 'screening' designs *after* classical experimental designs. In practice, most System Dynamics models have a great many factors that may be important; of course the analysts assume that only a few factors are really important: principle of *parsimony*. So in the beginning of a System Dynamics study it is necessary to search for the few really important factors among the many conceivably important factors. Classical textbooks do not discuss such screening situations, because in real-life experiments it is impossible to control (say) a hundred factors. In simulation, however, we perfectly control all inputs and we indeed use models with many inputs.

One approach is *group screening*, introduced in the early 1960s by Watson, Jacoby and Harrison, Li, and Patel. This technique aggregates the many individual factors into a few group factors. Some simulation applications can be found in Kleijnen (1987, p. 327); these applications are not System Dynamics models but queuing simulations. Recently Bettonvil (1990) further developed group screening into *sequential bifurcation*, which is a very efficient technique that accounts for white noise and interactions. He applied this technique to an ecological model with nearly 300 factors. Also see Bettonvil and Kleijnen (1992).

Another approach uses *random* combinations of input values. These designs were discussed in Technometrics back in 1959; also see Kleijnen (1990, pp. 321-323). We do not further discuss these designs because they are less efficient than group screening is.

More efficient but complicated approaches do not treat the simulation model as a black box; they use analytical differential analysis; see Ho and Cao (1991) and McRae (1989).

Regression Analysis: Some Technicalities

Eq. (3) gave the Ordinary Least Squares (OLS) estimator $\hat{\beta}$. In *deterministic* System Dynamics models that estimator may suffice, although Sacks et al. (1989) give a better estimator if the white noise assumption is dropped and is replaced by a stationary covariance assumption. In *random* models the classic assumptions seldom hold. If the response variances differ with the inputs (as the response means do), then Weighted Least Squares (WLS) is better. If common random numbers drive the various factor combinations, then Generalized Least Squares (GLS) is best. See Kleijnen (1987, pp. 161-175).

Once the regression model is calibrated (that is, the parameters β are estimated), the *meta-model's validity* must be tested. For deterministic System Dynamics models we propose cross validation: delete factor combination i (that is, delete x_i', y_i); reestimate β from the remaining simulation data (X_{-i}, y_{-i}); predict the deleted System Dynamics response y_i through the reestimated regression model

$(\hat{y}_i = \hat{\beta}_{-i}' x_i)$; 'eyeball' the relative prediction errors $\hat{y}_i / y_i$: are these errors acceptable to the user?

In random simulation we prefer *Rao*'s adjusted lack-of-fit F-test based on GLS: the estimated response variances and covariances are compared with the residuals $(\hat{y} - y)$. If, however, the System Dynamics responses are not normally distributed, then cross validation based on OLS is better. See Kleijnen (1992a).

In practice, System Dynamics models generate time series for several responses of interest; for example, inventory paths for different products. These time series can be characterized by their averages and quantiles; for example, we may estimate which value is not exceeded 90% of the time. Classical experiment design theory concentrates on a single response variable (denoted by y in this paper). In practice we use these designs to specify simulation runs, and we observe *several* responses per factor combination. To analyze these responses we can apply regression models per response; more sophisticated multivariate regression analysis does not seem to be worthwhile in practice. See Kleijnen (1987) and (1992b).

Sensitivity analysis, which we emphasize in this paper, should be distinguished from *Risk Analysis*. The latter type of analysis is an interesting combination of deterministic System Dynamics models and Monte Carlo sampling. System Dynamics models are often deterministic. These models depend on a number of inputs that are unknown. Therefore the user may specify a prior distribution of possible values, such as a normal or a uniform distribution. The computer samples values from that distribution, and generates output values, which are summarized in an output distribution. See Iman and Helton (1988).

Some applications

One case study concerns a set of deterministic ecological simulation models that resemble System Dynamics models (non-linear difference equations). These models require sensitivity analysis to support the Dutch government's decision making. Results for a model of the 'greenhouse' effect are given in Kleijnen, van Ham and Rotmans (1992).

Another case study concerns a Flexible Manufacturing System (FMS). Input to the deterministic simulation is the 'machine mix', that is, the number of machines of type i with i = 1,...,4. Intuitively selected combinations of these four inputs give inferior results when compared with a classical design. The throughput of the simulation is analyzed through two different regression metamodels. These models are validated. A regression model with only two inputs but including their interaction, gives valid predictions and sound explanations; see Kleijnen and Standridge (1988).

Applications of our approach are numerous in discrete-event simulation such as queuing simulation. An application is a decision support system (DSS) for production planning, developed for a Dutch company. To evaluate this DSS, a discrete-event simulation model is built. The DSS has 15 controllable variables that are to be optimized. The effects of these 15 variables are investigated, using a sequence of classical designs. Originally, 34 response variables were distinguished. These 34 variables, however, can be reduced to one criterion variable, namely productive machine hours, that is to be maximized, and one commercial variable measuring lead times, that must satisfy a certain side-condition. For this optimization problem the Steepest Ascent technique is applied to the experimental design outcomes. See Kleijnen (1992b).

Conclusions

Experimental design and regression analysis are statistical techniques that have been widely applied in the design and analysis of data obtained by *real life* experimentation and observation. In simulation, these techniques have gained popularity: a number of case studies have been published. The techniques need certain adaptations to account for the peculiarities of deterministic and random simulations. Their application to System Dynamics is straightforward, we claim.

References

Bettonvil, B., *Detection of Important Factors by Sequential Bifurcation*. Tilburg University Press, Tilburg, 1990.

Bettonvil, B. and J.P.C. Kleijnen, Measurement scales and resolution IV designs. *American Journal of Mathematical and Management Sciences, 10,* nos. 3 & 4, 1990, pp. 309-322.

Bettonvil, B. and J.P.C. Kleijnen, *Identifying the Important Factors in Simulation Models with Many Factors*. Tilburg University, Tilburg, March 1992.

Box, G.E.P. and N.R. Draper, *Empirical Model-Building And Response Surfaces*. John Wiley & Sons, New York, 1987.

Conway, R.W., Some tactical problems in digital simulation. *Management Science 10,* no. 1, Oct. 1963, pp. 47-61.

Federov, V.V., *Optimal Design of Experiments*. Academic Press, New York, 1972.

Ho, Y. and X. Cao, *Perturbation Analysis of Discrete Event Dynamic Systems*. Kluwer. Dordrecht (Netherlands), 1991.

Iman, R.L. and J.C. Helton, An investigation of uncertainty and sensitivity analysis techniques for computer codes. *Risk Analysis, 8,* 1988, pp. 71-90.

Karplus, W.J., The spectrum of mathematical models. *Perspectives in Computing, 3,* no. 2, May 1983, pp. 4-13.

Kleijnen, J.P.C., *Computers and Profits: Quantifying the Financial Benefits of Information*. Addison-Wesley, Reading (Massachusetts), 1980.

Kleijnen, J.P.C., *Statistical Techniques in Simulation*. Marcel Dekker, Inc., New York, 1974/1975. (Russian translation, Moscow, 1987).

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

Kleijnen, J.P.C., Statistics and deterministics simulation: Why not? *1990 Winter Simulation Conference Proceedings*, edited by O. Balci, ACM, New York, 1990.

Kleijnen, J.P.C., Regression metamodels for simulation with common random numbers: comparison of techniques. *Management Science* (accepted 1992a).

Kleijnen, J.P.C., Simulation and optimization in production planning: a case study. *Decision Support Systems* (accepted, 1992b).

Kleijnen, J.P.C., G. van Ham and J. Rotmans, Techniques for sensitivity analysis of simulation models: a case study of the CO_2 greenhouse effect. *Simulation* (forthcoming, 1992).

Kleijnen, J.P.C. and C.R. Standridge, Experimental design and regression analysis in simulation: an FMS case study. *European Journal of Operational Research, 33,* 1988, pp. 257-261.

McRae, G., Designing air quality control strategies. Chemical Engineering and Engineering and Public Policy, Carnegie-Mellon University, Pittsburgh (Pennsylvania), 1989.

Sacks, J., W.J. Welch, T.J. Mitchell, and H.P. Wynn, Design and analysis of computer experiments. *Statistical Science, 4,* no. 4, 1989, pp. 409-435.

Wolstenholme, E.F., *System Enquiry: a System Dynamics Approach*. John Wiley & Sons, Chichester (United Kingdom), 1990.