Sensitivity analysis and optimization of system dynamics models

Kleijnen, J.P.C.

Publication date:
1995

Citation for published version (APA):
Kleijnen, J. P. C. (1995). *Sensitivity analysis and optimization of system dynamics models: Regression analysis and statistical design of experiments*. (CentER Discussion Paper; Vol. 1995-4). Unknown Publisher.

\sysdyn.3

SENSITIVITY ANALYSIS AND OPTIMIZATION OF SYSTEM DYNAMICS MODELS: REGRESSION ANALYSIS AND STATISTICAL DESIGN OF EXPERIMENTS.

Jack P.C. Kleijnen
Center for Economic Research (CentER) and
Department of Information Systems
Tilburg University
Postbox 90153
5000 LE Tilburg, Netherlands

Fax:+3113-663377
E-mail: kleijnen@kub.nl

Version 3: January 16, 1995
Version 2: July 1992
Version 1: May 1992

Sensitivity Analysis and Optimization of System Dynamics Models: Regression Analysis and Statistical Design of Experiments

Jack P.C. Kleijnen

Abstract

This tutorial discusses what-if analysis and optimization of System Dynamics models. These problems are solved, using the statistical techniques of regression analysis and design of experiments (DOE). These issues are illustrated by applying the statistical techniques to a System Dynamics model for coal transportation, taken from Wolstenholme's book "System Enquiry: a System Dynamics Approach" (1990). The regression analysis uses the least squares algorithm. DOE uses classic designs, namely, fractional factorials and central composite designs. Compared with intuitive approaches, DOE is more efficient: DOE gives more accurate estimators of input effects. Moreover DOE is more effective: interactions are estimable too. The System Dynamics model is also optimized. A heuristic is derived, inspired by Response Surface Methodology (RSM) but accounting for constraints. Some remaining pertinent problems are briefly discussed, namely DOE for cases with many factors, and DOE for random System Dynamics models. Conclusions are presented for the case study, and general principles are derived. Finally 23 references are given for further study.

Introduction

Typically, analysts spent most of their time on developing the System Dynamics model, and little time on the sensitivity analysis of their model. Nevertheless, it is important to answer questions such as: what are the effects of changing input values; are there interactions among inputs? This article uses a case study that concerns a well-structured problem, so one more question is raised: which input values give optimal output? Many System Dynamics studies, however, concern ill-structured problems, so optimization is of academic interest, at most. But even models of ill-structured systems do require sensitivity analysis. Hopefully, this article will inspire System Dynamics analysts to apply the techniques of regression analysis and statistical design of experiments, to answer questions about what-if analysis, optimization, goal seeking, and model validation,.

System Dynamics uses simulation to 'solve' its models. Simulation is a mathematical technique that is very popular, because it is flexible, simple, and realistic. By definition, simulation involves experimentation, namely with the model of the real system. Experimentation, however, requires an appropriate *design and analysis*, if reliable results are desired.

Note that experiments with *real* systems have been frequently subjected to the design and analysis techniques developed in the field of mathematical statistics. In the 1930s Fisher focussed on agricultural experiments. Since the 1950s Box concentrated on chemical experimentation. Nowadays Taguchi's designs are very popular in industrial quality control. (Paragraphs starting with 'Note that' may be considered to be footnotes.)

Experiments with *simulated* worlds can also benefit of an appropriate design and analysis, as the case study in this article will illustrate.

Note the following *didactic* dilemma: should the theory of DOE be explained, starting from one or more case studies (inductive approach) or starting from general principles (deductive approach)? Because DOE is not well known in the Systems Dynamics community, it seems wise to introduce DOE to that community though a case study. For the *discrete-event* simulation community, Kleijnen (1994) gives an overview starting from general principles, summarizing several case studies.

Though DOE is not well known in System Dynamics, there are counterexamples. Optimization of System Dynamics models is investigated in Gustafsson and Wiechowski (1986) and Wolstenholme and Al-Alusi (1987); their techniques will be briefly discussed in the section on optimization. A recent book on System Dynamics is surprisingly brief on sensitivity analysis and optimization; see Wolstenholme, Henderson, and Gavine (1993, p. 54, 232). Also see Barlas and Carpenter (1990) and Richardson and Pugh (1981, pp. 277-292).

Crucial questions in simulation experimentation are: which combinations of inputs should be simulated, and how can the resulting output be analyzed? Obviously, these questions are asked in both random and in deterministic simulations. The case study in this paper concerns a deterministic model; random models will be briefly discussed near the end of this article.

Note that the term *inputs* refers not only to parameters and variables, but also to 'behavioral relationships', described as follows. Parameters are quantities that are not directly observable so they must be estimated; examples are the average delays in System Dynamics models. Examples of variables are resources such as bunker capacities in the case study of this article. Changing a behavioral relationship may mean that a certain policy is replaced by a different rule, as the case study will demonstrate. All three types of inputs are called 'factors' in DOE.

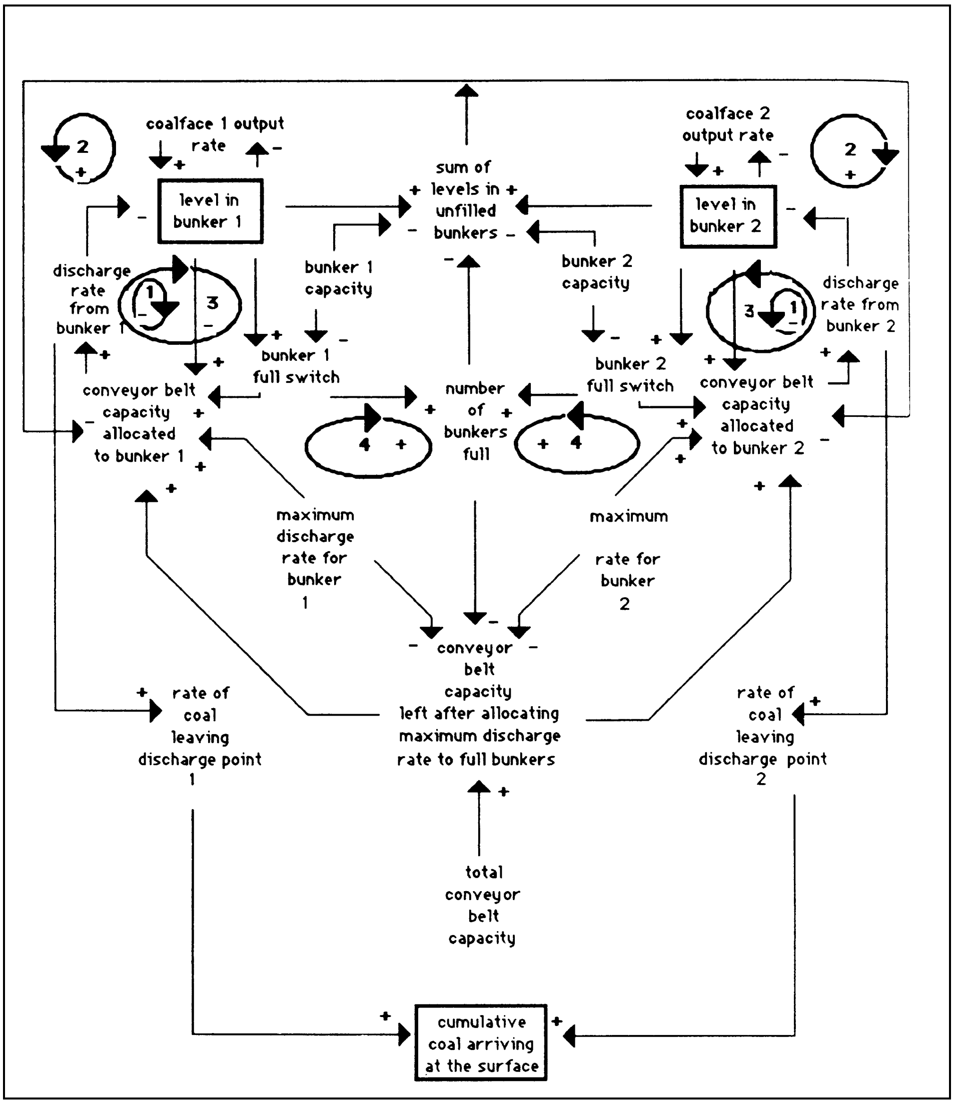
This article is further organized as follows. First the case study is presented, namely Wolstenholme's coal transportation model. Then his intuitive design is reproduced, for comparison with DOE. Next DOE is presented in the following subsections. First a simple, additive approximation to the input/output (I/O) behavior of the System Dynamics model is discussed; next a more general approximation is proposed, which accounts for interactions between factors and requires a design with 2^3 input combinations. A next section presents the results of the formal regression analysis of the 2^3 design. First the validity of the regression model is checked. Next individual input effects are examined. Three policies are modeled as a single qualitative factor; it is shown that the third policy is best. Finally the System Dynamics model is optimized: total cost is minimized such that efficiency remains 100%. A heuristic is derived, inspired by Response Surface Methodology (RSM) but accounting for constraints. Some remaining pertinent issues are briefly discussed, namely DOE for cases with many factors, and DOE for random System Dynamics models. Conclusions are presented for the case study, and general principles are derived. Finally 23 references are given for further study.

Case study: Wolstenholme's (1990) coal transportation System Dynamics model

Wolstenholme (1990, pp. 107-128) presents the following case (he spent eight years with British Coal); also see Figure 1 (which gives the influence diagram for only two bunkers and a simplified version of policy III).

A certain coal mine has three coalfaces, each linked to its own bunker. These bunkers have specific capacities (which are a topic of investigation in both Wolstenholme's book and this article). Each bunker receives its input from a single coalface, and discharges its output onto a conveyor belt that serves all three bunkers. This belt transports the coal to the surface of the mine. Whenever a bunker is full, the corresponding coalface must stop its output; obviously this congestion decreases the efficiency. The objective of management is to optimize the amount of coal moved from the coalface to the surface, each day.

Figure 1. Influence diagram of coal clearance model incorporating bunker discharge policy III (reproduced from Wolstenholme 1990, p. 128, figure 7.11).


coalface 1 output rate
level in bunker 1
sum of levels in unfilled bunkers
coalface 2 output rate
level in bunker 2
discharge rate from bunker 1
bunker 1 capacity
bunker 2 capacity
discharge rate from bunker 2
bunker 1 full switch
bunker 2 full switch
conveyor belt capacity allocated to bunker 1
number of bunkers full
conveyor belt capacity allocated to bunker 2
maximum discharge rate for bunker 1
maximum rate for bunker 2
conveyor belt capacity left after allocating maximum discharge rate to full bunkers
rate of coal leaving discharge point 1
rate of coal leaving discharge point 2
total conveyor belt capacity
cumulative coal arriving at the surface

When optimizing the coal flow, it is important to study capacity restrictions and control rules. *Vital questions* are: what are the efficiency effects of changing input values; which input values give optimal output; which decision rule leads to the best result; are there interactions among inputs, and so on. So this case study is representative of many problems that arise in real life, especially in physical distribution and production planning.

Wolstenholme's model assumes that bunker #3 is the first bunker that discharges coal onto the conveyor belt, followed by bunker #2, and finally bunker #1. The simulation response is *efficiency* that can *not* exceed 100%. Wolstenholme presents the following three control rules for managing the discharge rate of the bunkers.

Policy I: The discharge rate of each bunker can only be either zero or maximal (no intermediate values). The maximum is used as long as there is coal in the bunker and room on the conveyor belt.

Policy II: The bunker discharge rate can be any value between zero and its maximum: this rate equals the ratio of the bunker level and the bunker capacity. As in policy I, the discharge rate is subject to coal being available in the bunker and room being available on the conveyor belt.

Note that both policies lead to coal losses: production is stopped at the coalfaces whenever a bunker is full and no room is available on the conveyor belt. The order in which bunkers discharge onto the belt implies that the last coalface in line (coalface #1) suffers the heaviest losses. Therefore policy III is formulated.

A good explanation of policy III would require understanding the details of the coal transportation model, and hence would take more space than seems warranted for the purpose of this article. Details on the model and the policies can be found in Wolstenholme. For this article it suffices to understand that policy III is more sophisticated than policies I and II; hence the following crude explanation should do.

Policy III: The maximum discharge rate for a specific bunker is used, whenever that bunker is full and enough capacity is available on the conveyor belt. Next this policy calculates the discharge rate for the other partly empty bunkers. This rule determines whether the tentative allocation of the remaining conveyor belt capacities to these bunkers would result in violating their maximum discharge rates. In case of such a violation, these discharge rates are reduced to their maxima. If there is no such violation, remaining conveyor belt capacities are allocated among the remaining bunkers in proportion to their bunker levels.

For each policy, a System Dynamics model is programmed (this article uses the *POWERSIM 1.1* software, whereas Wolstenholme uses STELLA). Diagrams and equations of the model variants I, II, and III can be found in Wolstenholme.

This article will show that DOE treats the Systems Dynamics model as a black box; that is, the details of that model are not important. Nevertheless it is interesting to see that the case study concerns a truly complicated model. Therefore Figure 1 was added.

Intuitive Experimental Design

In the preceding section a number of 'vital' questions were raised, related to sensitivity analysis. Wolstenholme (1990, p. 115) answers these questions by varying *three inputs*, namely total belt capacity, maximum discharge rate per bunker, and capacity of each bunker. He then selects the *input combinations* (or runs) reproduced in Table 1. He seems to use intuition and common sense to select these particular combinations: for example, the number of values per factor is three, two, and five respectively.

Next he runs the nine combinations of Table 1; this yields the efficiency values in that table. The following problem is how to find a pattern in the I/O behavior of the coal transport model.

Table 1. Input/output of Wolstenholme's design per policy (source: Wolstenholme 1990, p.115, table 7.1).

	Total belt cap. (tons /hour)	Maximum discharge rate of each bunker (tons/hour)	Capacity of each bunker (tons)	Policy I	Policy II	Policy III
Run 1	2000	1000	500	72.84	73.75	75.68
Run 2	2000	700	500	55.73	68.20	75.04
Run 3	2000	1000	1000	76.30	77.86	83.00
Run 4	2000	1000	1200	78.47	80.73	85.60
Run 5	2500	1000	150	64.59	64.79	69.54
Run 6	2500	1000	500	73.17	84.48	90.78
Run 7	3500	1000	150	72.63	65.48	72.63
Run 8	3500	1000	500	98.46	86.16	98.46

To solve this problem, Wolstenholme (1990, pp. 116-121) again uses intuition and common sense, studying run after run. Moreover, based on this analysis he formulates policy III.

Statistical Design of Experiments (DOE)

DOE does not tell *which factors* should be studied; therefore the same three inputs as Wolstenholme used, are considered in this section. Neither does DOE tell *which range of input values* to consider; hence the same minimum and maximum values per factor are used, as in the preceding section. DOE does tell *which combinations of input values* to use, as follows.

It should be emphasized that the selection of input combinations depends on the kind of I/O behavior that is assumed to hold for the (simulated) system that is experimented with. The simpler that I/O behavior is, the fewer combinations are necessary (which makes sense intuitively). So in the following subsections several types of I/O behavior are discussed.

Additive metamodels for approximating I/O behavior of System Dynamics models

The simplest I/O behavior arises when the output (efficiency of coal transport) equals the sum of the effects (say) β_k of the individual inputs k (k = 1, 2, 3), plus an overall value (say) β_0 (this β_0 is roughly 80 in Table 1). So the I/O behavior of the System Dynamics model is modeled by an *additive metamodel* (it is called a metamodel because it is a model of an underlying System Dynamics model). The System Dynamics model is treated as a black box.

However, what exactly is meant by *the* effect of the individual input? Is it the effect of changing the input by one unit? But different inputs have different dimensions, as Table 1 demonstrates (tons/hour and tons respectively). In this article *sensitivity analysis is* defined as determining the effects of changing the inputs *drastically*, that is, changing the input from its minimum to its maximum value; the simulation model is not valid outside that range or in practice that factor can vary only over that domain. This type of sensitivity analysis should be distinguished from marginal analysis: what is the effect of infinitely small changes (perturbations). Bettonvil and Kleijnen (1990) show that the effects in sensitivity analysis are measured by β_k provided all inputs are standardized (scaled), as follows.

Let the original (non-standardized) input be denoted by z_k. In the simulation experiment z_k ranges between a lowest value L_k and a highest value H_k. For example, in Table 1 z_1 denotes total belt capacity (measured in tons/hour), so L_1 = 2000 and H_1 = 3500; z_2 has a much smaller range (namely, 1000 - 700 = 300). The variation (or spread) of factor k is measured by $a_k = (H_k - L_k)/2$ and its location (or mean) by $b_k = (H_k + L_k)/2$. Then input k may be *standardized* as follows:

$$x_{ik} = (z_{ik} - b_k)/a_k \tag{1}$$

where x_{ik} denotes the value of the standardized factor k in combination i ($i = 1, ..., n$).

So the additive metamodel is

$$y_i = \beta_0 + \sum_{k=1}^{3} \beta_k x_{ik} + e_i \tag{2}$$

where y_i denotes the simulation response of factor combination i, and e_i represents approximation error (the other symbols have already been defined).

This metamodel is a first degree polynomial in x; in other words, the I/O behavior within the experimental domain (determined by the ranges of the inputs) is approximated by a *Taylor* series cut off after the first-order effects.

Further this metamodel is a regression model linear in its regression parameters β (which quantify the factor effects). Hence these parameters can be estimated through the classic *least squares* criterion. Software for this fitting algorithm is abundant. The least squares estimator $\hat{\beta}$ of the parameter vector $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is

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

where $\boldsymbol{y}$ denotes the vector of simulation responses; obviously $\boldsymbol{X} = (x_{ik})$ must have full rank.

Let Q denote the number of regression parameters. In the additive metamodel, $\boldsymbol{X}$ is n by $Q = K + 1$. To give $\boldsymbol{X}$ full rank, it suffices to simulate $n = Q$ input combinations; in this case study, four combinations is enough. Moreover, it can be proved that the accuracy of the estimator $\hat{\beta}$ is maximized when $\boldsymbol{X}$ is *orthogonal*. An example is provided by the four combinations 1, 4, 6, and 7 in Table 2, or by the remaining four combinations.

For an orthogonal matrix $\boldsymbol{X}$ the least squares estimator $\hat{\beta}$ obviously reduces to the scalar expression

$$\hat{\beta}_k = \sum_{i=1}^{n} x_{ik} y_i / n \quad (k = 0, 1, ..., K). \tag{4}$$

Many analysts use a *one factor at a time* design: they first simulate the base scenario; next they change one input at a time. This design would also require four combinations, but since this design yields a non-orthogonal $\boldsymbol{X}$, it gives a less accurate estimator.

However, the additive metamodel may very well be too simple. Therefore it seems wise to be prudent, and to assume the following type of metamodel.

Metamodels with interactions for approximating I/O behavior

Wolstenholme (1990, p. 123) states '... the bunker discharge policy itself interacts with the bunker level'. Intuitively, interaction means that the effect of one factor also depends on the value of another factor. Formally, the following metamodel can express interactions:

$$y_i = \beta_0 + \sum_{k=1}^{3} \beta_k x_{ik} + \sum_{k=1}^{2} \sum_{k'=k+1}^{3} \beta_{k,k'} x_k x_{k'} + e_i \qquad (5)$$

where $\beta_{k,k'}$ denotes the interaction between factors k and k' with $k < k'$ (obviously $\delta y / \delta \beta_k$ depends on both k and k').

In this metamodel, Q (the number of regression parameters) is 1 + 3 + 3(3 -1)/2 = 7, so n (the number of combinations or rows of $\boldsymbol{X}$) must be at least seven. However, the standard DOE solution takes eight combinations, namely the 2^3 combinations in Table 2. Observe that the combinations in this table are formed systematically, whereas the combinations in Table 1 are selected intuitively. Both tables use approximately the same number of combinations (nine and eight respectively), but Table 2 gives an orthogonal $\boldsymbol{X}$ (so Table 2 gives a more accurate estimator $\hat{\beta}$, given n = 8).

In general, an orthogonal $\boldsymbol{X}$ (such as the one in Table 2) can be found through straightforward procedures in case n equals 2^{k-p} where p denotes a non-negative integer such that $n \geq Q$. For general n values there are tables and software. See Box and Draper (1987) and Kleijnen (1987).

The three original inputs are 'total belt capacity' z_1, 'maximum discharge rate of the bunker' z_2, and 'capacity per bunker' z_3. Hence the 2^3 design of table 2 gives Table 3, which also shows the simulated outputs.

Note that some combinations in Wolstenholme's design and in the 2^3 design have identical input combinations (see Tables 1 and 3). The fact that the corresponding output values are identical (apart from numerical rounding errors) strongly suggests that both models have been programmed correctly (verification).

Note that interactions may be important, but they may be ignored in the analysis (that is, the additive model of Eq. 2 is used, whereas the model with interactions of Eq. 5 holds). It can be proven that even then the 2^3 design gives unbiased estimators of the main effects. (The 2^3 design is said to have at least 'resolution 4'; see Kleijnen 1987, p. 301).

Table 2. Full factorial 2^3 design in standardized factors x_{ik}.

Combination	x_1	x_2	x_3
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

Table 3. Input/output of 2^3 design per policy.

	input			efficiency		
Run	z1	z2	z3	policy I	policy II	policy III
1	2000	700	150	55.78	56.87	65.87
2	3500	700	150	66.34	57.65	66.34
3	2000	1000	150	56.48	62.09	66.42
4	3500	1000	150	72.63	65.48	72.63
5	2000	700	1200	62.62	74.61	85.16
6	3500	700	1200	87.94	74.76	87.94
7	2000	1000	1200	78.47	80.73	85.60
8	3500	1000	1200	100	93.18	100

Regression Analysis of Experiment

Practitioners often make a *scatter plot* with on the x-axis the values of one factor (for example, capacity per bunker z_3) and on the y-axis the simulation response (efficiency). This graph indicates the input/output behavior of the simulation model, treated as a black box. It shows whether this factor has a positive or negative effect on the response, whether that effect remains constant over the relevant domain of the factor, and so on.

This scatter plot can be further analyzed, *fitting a curve* to the (x, y) data; for example, a straight line ($y = \beta_0 + \beta_1 x$). Moreover, this graphical analysis should account for the variations in the other factors (total belt capacity z_1 and maximum discharge rate per bunker z_2): *interaction*.

To study these interactions, scatter plots per factor can be superimposed. For example, the scatter plot for different values of z_3 may be drawn, given a certain combination of values for z_1 and z_2. Plots of y versus z_3 for different combinations of z_1 and z_2 can now be superimposed. Intuitively, the curve for a combination of *high* values for z_1 and z_2 lies above the curve for a combination of *low* values (if not, the System Dynamics model is probably wrong; see Kleijnen 1995). If the response curves are not parallel, there are interactions, by definition. To save space, these pictures are not shown in this article (they are available from the author upon request).

However, superimposing many plots is cumbersome. Moreover, their interpretation is subjective: are the response curves really parallel, etc.? These shortcomings are removed by regression analysis, as this section will demonstrate.

In his *intuitive* analysis of the I/O data in Table 1, Wolstenholme (like many other analysts) does not explicitly mention interactions. DOE, however, implies that if interactions are conjectured to be important, then they are estimated. This estimation uses the least squares estimator in Eq. (3) where in case of interactions $\boldsymbol{X}$ becomes an eight by seven matrix (there are $n = 2^3$ runs in Table 3, and $Q = 7$ effects in Eq. 5).

Of course intuition should never be discarded lightly. So it is good practice to eyeball the results of the experiment, before performing a formal analysis. For example, in Table 3 inputs 1 and 2 seem to have some effects with policy I, but input 3 has the greatest effect. With policy II input 1 has little influence. The factors 2 and 3 have relatively strong effects. With policy III it seems difficult to see how much influence inputs 1 and 2 have; probably input 1 has more effect than input

2 has. For sure, input 3 has a very strong effect on efficiency. For all three policies, factor 3 has a great effect. When input 3 reaches its maximum, efficiency improves.

The formal regression analysis considers the three inputs z_1 through z_3, which are quantitative inputs, plus the three policies (I through III), which are qualitative. Qualitative factors are slightly more difficult to represent in a regression model. Therefore the effects of the three quantitative factors are first examined *per policy*. Next policy is incorporated as a factor.

In practice, analysts often try to interpret individual effects before they check that the regression model as a whole makes sense. In this article, however, first it is checked that the estimated regression model is a *valid* approximation of the System Dynamics model's I/O behavior. If the metamodel seems valid, its individual effects are examined.

Checking the validity of the metamodel

To check whether the estimated regression model is a valid approximation, two approaches can be followed. The second approach is only supported by modern statistical software (such as SAS).

(i) How well the regression model fits the simulated data, can be measured through the *R-square coefficient*:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{6}$$

where $\bar{y}$ denotes the average output; $\hat{y}$ denotes the fitted values ($\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$); $\hat{y} - y$ is called the residual. So a perfect fit gives $R^2 = 1$. However, obviously R^2 increases as Q increases (any regression model with $Q = n$ yields $R^2 = 1$). Therefore the *adjusted* R^2 is defined:

$$R^2_{adj} = 1 - (1 - R^2)\frac{(n - 1)}{(n - Q)}. \tag{7}$$

The factor $n - Q$ is called the degrees of freedom. When $Q = n$, then R^2_{adj} is undefined.

(ii) The regression model can also be used to predict the simulation output for a *new* combination. To save computer time, *cross-validation* is used: the regression model is estimated using only seven of the eight combinations in Table 3. In that table, first combination 1 is deleted; then combination 2, and so on. So when combination 1 is deleted, the regression parameters are estimated from the remaining seven combinations (2 through 8). This estimator (say) $\hat{\beta}_{-1}$ is used to predict the simulation response $\hat{y}_1$. The actual simulation response is already known ($y_1 = 55.78$ for policy I; see Table 3). Hence prediction errors can be computed. Table 4 gives the *relative prediction errors* $\hat{y}/y$.

Table 4. Relative prediction errors in cross-validation, for policy I.

Combination deleted:	1	2	3	4	5	6	7	8
$\hat{y}_{-i}/y_i$	0.766	1.039	1.053	1.102	1.194	0.981	0.973	0.916

The I/O data of the 2^3 design in Table 3 are also analyzed through other regression (meta)models. Searching for a 'good' regression model requires intuition, common sense, and knowledge of the underlying system that generated the I/O data (the System Dynamics model and the real system). This search receives more attention in econometrics than in DOE. To save space, these details are skipped. The final conclusion is that in this case study a regression model with the three main effects seems best (interactions turn out to be insignificant; deleting the main effect of input #2 increases the relative prediction error); see Table 5.

Note that if the fit of the estimated regression model is not acceptable, then it may be worthwhile to replace the variables z by log(z). In general, transformations of variables may improve the validity of the regression (meta)model (also see the subsection on random System Dynamics models, later on).

Individual input effects

It is well-known that the estimated effects can be tested statistically, assuming that the approximation errors e are *white noise*, that is, e is normally and independently distributed with zero mean and constant variance (say) σ^2. Then the least squares estimator, defined in Eq. (3), yields the variance-covariance matrix

$$\hat{\mathbf{cov}}(\hat{\beta}) = (\mathbf{X}'X)^{-1}\hat{\sigma}^2 \tag{8}$$

where the elements on the main diagonal of $\hat{\mathbf{cov}}(\hat{\beta})$ are the estimated variances of the estimated input effects. Taking the squares gives the *standard errors* (say) $s(\hat{\beta}_k)$. To test if β_k is zero (unimportant effect), Student's t statistic is computed:

$$t_{n-Q} = \hat{\beta}_k / s(\hat{\beta}_k). \tag{9}$$

The critical value of this statistic can be looked up in a table, provided a *significance level* (say) α is fixed. A usual value is 0.10, but to reduce the probability of falsely eliminating important inputs, α = 0.20 is also used in this article.

It is interesting to see how the estimated individual effects β_k change, as combinations are deleted. Obviously, if the specified regression model (see Eqs. 2 and 5) is a good approximation, then the estimates remain stable. Table 5 illustrates this approach for the additive metamodel in Eq. (2) when using the I/O data of Table 3 (2^3 design). Observe that the three estimated main effects have the correct signs: increasing input capacities increase the efficiency. Moreover, the intuitive analysis suggested that input 3 has more effect than input 1, which in turn exceeds the effect of input 2.

For policy II, however, the best metamodel turns out to have main effects only for inputs #2 and #3: deleting the nonsignificant main effect of input #1, decreases the maximum relative prediction error; interactions are not significant. For policy III a model with the three main effects gives again a good approximation.

Note that in another case study, concerning a Flexible Manufacturing System (FMS), only a regression model *with* interactions (besides main effects) gives valid predictions and sound explanations; see Kleijnen and Standridge (1988).

Regression metamodels with policy as a qualitative factor

Next policy is modeled as a single qualitative factor. Hence there are now three quantitative inputs, each simulated for only two values, and there is one qualitative factor with three 'levels' (policy I, II, III). (Technically, regression analysis handles this qualitative factor through two binary (0, 1) variables; see Kleijnen 1987.)

Table 5. Estimates of main effects, upon deleting a combination, in policy I; blanks mean nonsignificant; * denotes significance at level 0.20, other values are significant at level 0.10.

Combination deleted	ß$_0$	ß$_1$	ß$_2$	ß$_3$	R^2	adj. R^2
1	70.900	10.823	5.995	11.358	0.9772	0.9543
2	72.858	9.520	4.038 *	9.400	0.9316	0.8632
3	72.906	8.821	4.736 *	9.351	0.9202	0.8404
4	73.466	10.129	5.300	8.791	0.9478	0.8955
5	74.053	7.675	2.843 *	11.245	0.9730	0.9461
6	72.320	8.983	4.575 *	9.613	0.9194	0.8387
7	72.271	9.456	4.101	9.463	0.9310	0.8620
8	71.486	8.149		8.679	0.9026	0.8052
None	72.535	9.195	4.363	9.725	0.9314	0.8799

The best regression model includes the main effects of all four factors. Policy III is the best policy; policy II is worse than policy I, even though policy I is the simplest policy. These regression results agree with an intuitive analysis of the raw I/O data in Table 3: calculate the efficiency per policy, averaged over all 2^3 combinations of the three other factors (these averages are 72.50, 70.75, 78.75).

Note that there are different types of metamodels, besides the linear regression models of this article. For example, piecewise linear models can perfectly fit the observed deterministic simulation responses y_i with i = 1, ..., n. A perfect fit can also be obtained without discontinuities at the observations, by means of splines. Of course, new observations will deviate from the fitted metamodel. However, these deviations may be assumed to be smaller, the closer the new observation lies to an old observation. This assumption implies that the approximation errors e_i no longer are white noise. The errors may then be assumed to form a covariance stationary process with positive correlation such that the correlation coefficients decrease as the distance between the new observation and the neighboring old observation increases. Sacks, Welch, Mitchell, and Wynn (1989) use such a process to model the systematic effects of the inputs (not the noise part). Different types of metamodels are surveyed in Barton (1992). Obviously the metamodels of this article are of the simplest type.

Optimization of Simulated Coal Transportation

In the introduction, optimization has already been mentioned as one of the crucial questions when modeling well-structured systems. Wolstenholme (1990, pp. 125-127) assumes that the maximum discharge rate is fixed at an average of 1000 tons/hour. Therefore he optimizes only the two remaining inputs, 'total belt capacity' z_1, and 'capacity per bunker', denoted by z_2 from now on. The optimization is further restricted to the best control rule, namely policy III. Wolstenholme additionally assumes that *total costs* -denoted by C- is to be minimized, under the condition that the efficiency (y) remains at its maximum (100%). He assumes that the cost parameters are £1000/ton/hour for z_1 and £2000/ton for z_2.

This yields the following mathematical optimization problem:

$$\underset{z_1,z_2}{Min}\ C = 1000z_1 + 2000z_2 \tag{10}$$

such that

$$y(z_1,z_2) = 1. \tag{11}$$

There are many optimization procedures in the fields of mathematical programming and computer science (the best known being linear programming, but Eq. 11 is not linear; other techniques are genetic algorithms, simulated annealing, tabu search, etc.). Optimization of System Dynamics models is investigated in Gustafsson and Wiechowski (1986); they apply 'Fletcher and Reeves's conjugate gradients' to a toy problem. Optimization of System Dynamics models is also studied in Wolstenholme and Al-Alusi (1987). Their heuristic 'might take 100 or more iterations' (see their page 102); no constraints seem to be accounted for, but no details of the heuristic are presented. They apply the heuristic to a defence system. A recent overview of optimization in discrete event simulation is Fu (1994) with 133 references; also see Merkuryev and Visipkov (1994). This article uses the following heuristic.

A new heuristic for optimizing system Dynamics models

The heuristic in the sequel is inspired by *Response Surface Methodology (RSM)*, which approximates the I/O behavior -now called the response surface- *locally* by low-order polynomials in the input variables (here z_1 and z_2). In the first stages, RSM uses first-order polynomials; in the final stage it uses a second-order polynomial. As the preceding sections showed, the estimation of these local approximations can be performed through regression analysis and DOE.

Note that classic RSM determines the direction of better performance (higher efficiency), using the *steepest ascent* algorithm. Given a value for y, the first-order polynomial defines a K-dimensional hyperplane (here K = 2); the steepest ascent path is perpendicular to that plane. Obviously, as fitted local planes change, so does the direction of the path. Classic RSM, however, maximizes a single criterion, ignoring restrictions (side-conditions).

Step 1. Find an *initial* combination $\mathbf{z} = (z_1, z_2)'$ that yields a simulated efficiency of 100% (y = 1; see Eq. 11). Such a combination is already available: see the element in the last row and column of Table 3.

Step 2. Reduce each input by (say) 10%. Simulate the System Dynamics model with this input. Obtain the corresponding output.

Step 3. If the output of step 2 is still 100% (y = 1), then return to step 2; else ($y < 1$) proceed to the next step.

Step 4. Find the most recent input combination with y = 1 (see steps 1 and 2), and reduce the step size to (say) 5%. Simulate the model with this new input combination, and obtain the corresponding output.

Step 5. Further 'explore' the most recent local area that includes a combination with y = 1; that is, simulate the model for four input combinations, specified by the 2^2 design (the standardized values are given by Table 2's first four rows, eliminating the last column; the original values are determined by the heuristic of this section).

To save computer time, the combination (x_1 = 1, x_2 = 1) is taken equal to the last input combination that yielded y = 1. As the combination (-1, -1) the last combination that gave $y < 1$ is taken. So two new combinations must be simulated: (1, -1) and (-1, 1).

The heuristic yields L_1 = 2693.25, H_1 = 2835.00, L_2 = 923.40, and H_2 = 972.00; also see Figure 2 (these four points form a rectangle).

Since no further progress can be made, the optimum seems to be close. Now a *second-order polynomial* is used to approximate the production function $y(z_1, z_2)$: such an approximation can model a 'hill top', whereas a first-order polynomial can model only a 'hill side', even when interactions are included.

To estimate the six parameters in this approximation (see Eq. 12 below), the 2^2 design is expanded to a so-called *central composite design*; also see Figure 2. So the central point is added: (0, 0) in standardized values; (2764, 948) in original values. Moreover two one-at-a-time points are added: (a, 0) and (0, a) with $a \neq 0$, $a \neq 1$. Two more points are added, namely the mirror images (-a, 0) and (0, -a). The value a is set to 0.75: if $a > 1$ then (0, a) and (a, 0) give high cost, and (0, -a) and (-a, 0) give low efficiency. Altogether the number of combinations is 9 (= $2^2 + 1 + 4$). (Obviously this design simulates five values per input; see the numbers along the axes, except for the estimated optimal values.) This design yields Table 6.

From these I/O data the second-order polynomial is estimated:

$$\begin{aligned}\hat{y} = {} & 601.138514 - 0.1239693 z_1 - 0.7161188 z_2 + \\ & - 0.0000334 z_1 z_2 + 0.0000291 z_1^2 + 0.0004282 z_2^2\end{aligned} \qquad (12)$$

Note that the second-order and interaction coefficients are relatively small, but z_1^2, z_2^2, and $z_1 z_2$ are relatively large, so these coefficients should not be ignored.

Step 6. Combine the restriction in Eq. (11) and the estimated production function in Eq. (12); that is, replace $\hat{y}$ in (12) by the value 1:

$$\begin{aligned}& 601.138514 - 0.1239693 z_1 - 0.7161188 z_2 + \\ & - 0.0000334 z_1 z_2 + 0.0000291 z_1^2 + 0.0004282 z_2^2 = 1\end{aligned} \qquad (13)$$

Under this restriction, minimize the total cost given by Eq. (10). This mathematical problem can be solved through a *Lagrangean multiplier*. In this case study, a unique solution for the estimated optimal input combination follows: $(\hat{z}_1^*, \hat{z}_2^*)$ = (2841.35; 968.17).

Note that Eq. (13) gives the estimated *efficiency frontier* in Figure 2: outside the ellipsoid, inputs are wasted; inside that ellipsoid the efficiency is too low. The optimum combination is the point where the efficiency frontier is touched by an iso-cost line. There are infinitely many parallel iso-cost lines (all with angle -1/2; see Eq. 10), but the figure shows the unique iso-cost line that forms the tangent of the ellipsoid.

Because the estimated optimal combination is based on an approximation, this solution is checked by simulating the System Dynamics model with this input combination. It indeed gives 100% efficiency. However, its cost is £ 4.77769 mln, which exceeds the lowest cost in Table 6 that corresponds with a combination that gives 100% efficiency: combination (2835.00; 923.40) gives £ - 4.68180 mln (2.1% lower). The explanation is that the second-order polynomial has an R-square of only 0.80357.

Compared to Wolstenholme (1990, pp. 125-127), this solution saves 15.12% (£ 0.72231 mln), a substantial cost reduction indeed.

Note that minor cost reductions can be obtained by further exploring the immediate area around the solution found above. In practice, however, the input may be restricted to multiples of (say) 100; then the optimal combination is (2900, 1000) with cost £ 4.9.

Figure 2. Central composite design, estimated efficiency frontier, and optimal iso-cost line.

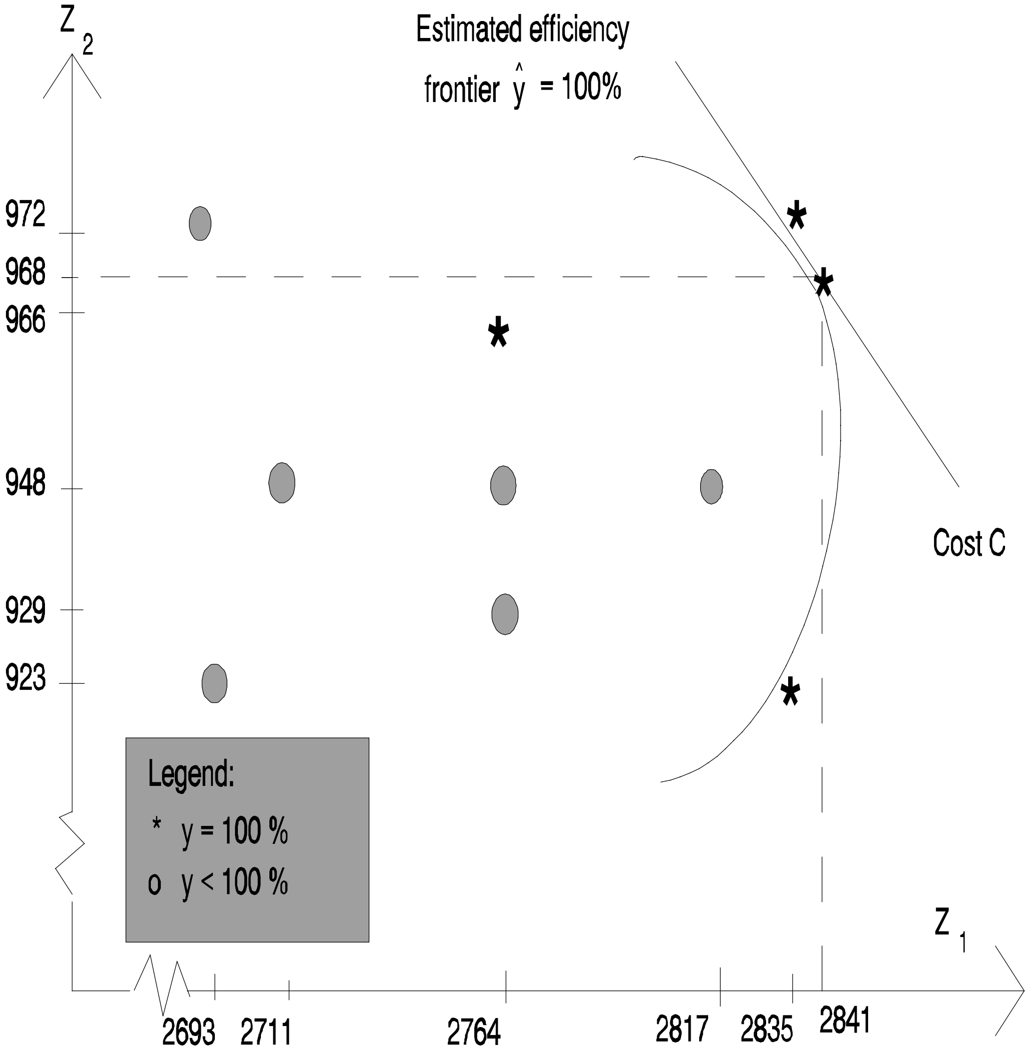


Table 6. Input combinations in the central composite design with corresponding costs and efficiencies.

Input combination standard variables (x_1, x_2)	Input combination original variables (z_1, z_2)	Cost (£ mln) C	Efficiency (%) y
(1.00, 1.00)	(2835.00; 972.00)	4.7790	100.00
(-1.00, 1.00)	(2693.25; 972.00)	4.63725	99.59
(1.00, -1.00)	(2835.00; 923.40)	4.68180	100.00
(-1.00, -1.00)	(2693.25; 923.40)	4.54005	99.36
(0.00, 0.00)	(2764.13; 947.70)	4.65953	99.35
(0.00, 0.75)	(2764.13; 965.97)	4.69607	99.59
(0.75, 0.00)	(2817.28; 947.70)	4.71268	100.00
(0.00, -0.75)	(2764.13; 929.48)	4.62309	99.36
(-0.75, 0.00)	(2710.97; 947.70)	4.60637	98.83

Remaining Pertinent Issues

There are several issues that did not come up in the case study discussed above, but that are relevant.

DOE for cases with many factors

The number of factors in the case study is very small: K = 3. Wolstenholme (1990, p. 114) does mention that the number of shifts may be a factor too. Now several case studies are briefly summarized, to illustrate that more factors may indeed be present.

One case study concerns a deterministic ecological simulation model of the greenhouse phenomenon (increase of temperatures worldwide). Mathematically, this model resembles System Dynamics models: the model consists of non-linear difference equations. This model requires sensitivity analysis to support the Dutch government's decision making. Results for a submodel with ten factors, investigated in a 2^{10-5} design, are given in Kleijnen, Van Ham, and Rotmans (1992).

In general, if the regression metamodel includes two-factor interactions, then the number of effects increases to $1 + K + K(K - 1)/2$. If K is small (as in Wolstenholme's case study), then it is practical to simulate $n = 2^K$. Otherwise, it is better to simulate only a fraction (such as a 2^{10-5} design).

Another submodel of the same greenhouse simulation has 62 inputs. A 2^{62-55} design is presented in Kleijnen, Van Ham, and Rotmans (1992). However, such a large number of factors may be better handled as follows.

If a System Dynamics model has a great many factors, then the analysts should assume that only a few factors are really important: principle of *parsimony*. So in the beginning of a study it is necessary to search for the few really important factors among the many conceivably important factors. Classic statistics books do not discuss such screening situations, because in real-life experiments it is impossible to control (say) a hundred factors. In simulation, however, the analysts perfectly control all inputs and indeed use models with many inputs. For example, Bettonvil and Kleijnen (1994) examine 281 factors for the ecological model mentioned above.

One approach is *group screening*, which aggregates the many individual factors into a few group factors. Recently Bettonvil and Kleijnen (1994) further developed group screening into *sequential bifurcation*, which is a very efficient technique that accounts for interactions. They detected the 15 most important factors among the 281 factors, simulating only 144 combinations.

Random System Dynamics models

Some System Dynamics models are random or stochastic, for example, models for the effects of inaccurate information; see Kleijnen (1980, 137-143) and Wolstenholme, Henderson, and Gavine (1993).

If not only the output means but also the output variances differ with the inputs, then *Weighted Least Squares (WLS)* gives the best estimator of the effects β. WLS gives smaller weight to an output, the higher its variance is.

If *common random numbers* drive the various input combinations, then *Generalized Least Squares (GLS)* is best. GLS estimates the correlations between outputs at different combinations. See Kleijnen (1987, 161-175).

The *metamodel's validity* can be tested in random simulation, using *Rao*'s adjusted lack-of-fit F-test based on GLS: the estimated response (co)variances are compared with the residuals. If, however, the output *y* is not normally distributed, then cross-validation based on OLS is better. See Kleijnen (1992).

If the fit of the estimated regression model is not acceptable, then it may again be useful to *transform* variables: for example, replace *z* by log(*z*). In general, transformations may improve the validity of the regression (meta)model, and give simulation responses that better satisfy the statistical assumptions of normality, constant variances, and lack of autocorrelation. See Kleijnen (1987).

Other questions are: how to start up the simulation run; how long to continue that run; how often to repeat that run with different random numbers; how to reduce the variance of the output? See the textbook by Kleijnen and Van Groenendaal (1992) for an introduction; for statistical details see Kleijnen (1987).

Applications of the approach outlined in this article, are numerous in discrete-event (random) simulations such as queuing simulations. An example is the 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 inputs that are to be optimized. The effects of these inputs are investigated, using a sequence of classic designs. One criterion variable, namely productive machine hours, is to be maximized, and one commercial variable measuring lead times, must satisfy a certain constraint (so mathematically this problem looks like the case discussed in the section on optimization). RSM is applied in this case study. See Kleijnen (1993).

Conclusions

This article was written to make DOE better known in the System Dynamics community. To illustrate the benefits of DOE, a case study was presented, namely Wolstenholme's coal transportation model. First, conclusion for this case study are presented; next general principles are summarized.

Case study results

Wolstenholme's intuitive design and analysis is not most efficient. DOE gives a good approximation of the I/O behavior of his System Dynamics model. DOE is objective, scientific, efficient, and effective.

The conjectures made in the informal analysis are confirmed by the regression analysis. Policy III is the best policy; it achieves the maximum efficiency, namely 100%. Policy I is better than policy II, even though the latter is more sophisticated.

DOE can account for interactions among factors. However, the regression analysis with a classical α-level of 10% does not give significant interactions.

The second part of this article dealt with the optimization of Policy III's two inputs, namely conveyor belt capacity and bunker capacity. A heuristic was derived to obtain the input combination with minimum cost and 100% efficiency. The resulting minimum cost is 15.12% lower than Wolstenholme's optimum.

General principles

This case study illustrates the following general principles. In practice there is prior knowledge about the System Dynamics model and the underlying real system. This knowledge can be formalized in a *tentative* regression (meta)model; in other words, this model must be tested later on to check its validity. The regression model specifies *which inputs* seem important, which *interactions* among these inputs may be important, and which *experimental domain* seems appropriate.

The purpose of the regression model is to guide the design of the simulation experiment and to interpret the resulting simulation data.

In all experiments, analysts use metamodels, explicitly or implicitly. For example, if they change one factor at a time, then (implicitly) they assume that all interactions $\beta_{k,k'}$ are zero. Of course it is better to make the metamodel explicit and to find a design that fits that model.

The tentative regression metamodel guides the DOE. The design 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.

Classic DOE gives designs that are both 'efficient' and 'effective'. *Efficiency* means that the number of input combinations is 'small'. When there are Q effects in the regression metamodel, the number of runs *n* should satisfy the condition $n \geq Q$; for example, $K + 1$ runs suffice if there are no interactions among the *K* inputs. Designs are *effective* if they permit the estimation of *interactions*.

Experimental design and regression analysis are statistical techniques that have already been widely applied to data obtained by *real life* experimentation and observation in agriculture, chemistry, social sciences, and so on. The techniques need certain adaptations to account for the peculiarities of simulation. In discrete-event simulation, these techniques have also gained popularity: many case studies and much research have been published. Their application to System Dynamics models is straightforward, as this article has demonstrated, hopefully.

References

Barlas, Y. and S. Carpenter. 1990, Philosophical roots of model validation: two paradigms. *System Dynamics Review*, 6, 2, 148-166.

Barton, R.R. 1992. Metamodels for simulation input-output relations. *Proceedings of the 1992 Winter Simulation Conference*, ed. J.J. Swain, D. Goldsman, R.C. Crain, and J.R. Wilson: 289-299.

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

Bettonvil, B. and J.P.C. Kleijnen. 1994. *Identifying the Important Factors in Simulation Models with Many Factors*. Tilburg: Tilburg University.

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

Fu, M. 1994. Optimization via simulation: a review. *Annals of Operations Research* 53 (to appear).

Gustafsson, L. and M. Wiechowski. 1986, Coupling DYNAMO and optimization software. *System Dynamics Review*, 2, 1, 62-66.

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

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

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

Kleijnen, J.P.C. 1992. Regression metamodels for simulation with common random numbers: comparison of validation tests and confidence intervals. *Management Science* 38, no. 8: 1164-1185.

Kleijnen, J.P.C. 1993. Simulation and optimization in production planning: a case study. *Decision Support Systems* 9: 269-280.

Kleijnen, J.P.C. 1994. Sensitivity analysis and optimization of simulation models. *Proceedings of the 1994 European Simulation Symposium*, La Jolla: The Society for Computer Simulation, 3-8. Kleijnen, J.P.C. (1995), Verification and validation of simulation models. *European Journal of Operational Research* (in press).

Kleijnen, J.P.C. and W. Van Groenendaal. 1992. *Simulation: a Statistical Perspective*. Chichester: John Wiley & Sons.

Kleijnen, J.P.C., G. van Ham and J. Rotmans. 1992. Techniques for sensitivity analysis of simulation models: a case study of the CO_2 greenhouse effect. *Simulation* 58, no. 5: 410-417.

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

Merkuryev, Y. and V.L. Visipkov. 1994. A survey of optimization methods in discrete systems simulation. *CISS First Joint Conference of International Simulation Societies Proceedings*, ed. J. Jalin, W. Karplus, and R. Rimane, La Jolla: The Society for Computer Simulation, 104-110.

Richardson, G.H. and A. Pugh. 1981, *Introduction to System Dynamics Modeling with DYNAMO*, Cambridge: MIT Press.

Sacks, J., W.J. Welch, T.J. Mitchell, and H.P. Wynn 1989. Design and analysis of computer experiments (including Comments). *Statistical Science* 4 (4): 409-435.

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

Wolstenholme, E.F. and A. Al-Alusi. 1987. System Dynamics and heuristic optimisation in defence analysis. *System Dynamics Review*, 3, 2, 102-115.

Wolstenholme, E.F., S. Henderson, and A. Gavine. 1993. *The Evaluation of Management Information Systems; a Dynamic and Holistic Approach*. Chichester (United Kingdom): John Wiley & Sons.

Acknowledgment

Patrick Bosman, a student in the Department of Econometrics of Tilburg University, programmed the System Dynamics model, and executed the experimental design, the regression analysis, and the optimization heuristic. Bert Bettonvil, Willem van Groenendaal, and Martin Smits of the Department of Information Systems gave valuable comments on a preliminary version of this article.