TECHNICAL ARTICLE

SIMULATION 58:6,410-417

ISSN 0037-5497/92 $3.00 + .10
Printed in the United States of America

Techniques for sensitivity analysis of simulation models: A case study of the CO_2 greenhouse effect

Jack P.C. Kleijnen
Greet van Ham
Katholieke Universiteit Brabant
(Tilburg University)
5000 LE Tilburg
The Netherlands

Jan Rotmans
RIVM
(National Institute of Public Health and Environmental Protection)
The Netherlands

Sensitivity analysis is needed for validation, what-if analysis, and optimization of complicated simulation models. One set of techniques for sensitivity analysis consists of least squares curve fitting, regression analysis, and statistical designs such as factorial designs. In this case study these techniques are applied to several modules of a large integrated assessment model for the greenhouse effect, developed in The Netherlands. The regression models turn out to be valid approximations to the simulation models. Some estimated effects are quite surprising for the simulation users.

Keywords: What-if analysis, validation, optimization, least squares, regression analysis, experimental design, factorial design, ecology, greenhouse effect, simulation, metamodel.

Introduction

Complicated simulation models have been constructed in many disciplines. All these models confront the analysts with the problem of sensitivity analysis; that is, what are the effects of changing the parameters and input variables of the simulation model? That question arises in 'what if' analysis, validation, optimization, and so on. This article introduces and illustrates the application of *simple* techniques that originated in the discipline of mathematical statistics. These techniques are least squares curve-fitting, regression analysis, and statistical designs such as 2^{k-p} designs. The techniques are applied to an integrated assessment model for the greenhouse effect. This model has been developed at a large Dutch institute called National Institute of Public Health and Environmental Protection (or RIVM in Dutch).

One of the major imminent ecological threats of the world is the 'enhanced greenhouse problem': the earth and the lower layers of its atmosphere have shown rising temperatures over the past hundred years. This phenomenon is probably caused by an increase of greenhouse gases (such as carbon dioxyde, methane, and ozone) that absorb the earth's heat radiation, so the global average temperature rises. Mankind is largely responsible for this increased 'greenhouse' gas concentration. Temperatures are expected to rise further, but with different amounts in different regions of the earth (the tropics will be less affected probably). Higher temperatures will cause thermal expansion of the oceans and melting of arctic ice, which raise the sea level. Many more processes, however, are involved; see [9].

One consequence of a higher sea level is the need to raise the dikes' height in the Netherlands. A survey of the effects for society is given in [5].

To gain quantitative insight into the greenhouse problem and develop long-term strategies for coping with climatic changes, RIVM developed the Integrated Model for the Assessment of the Greenhouse Effect or *IMAGE*. This model is a deterministic simulation (but most of the sensitivity techniques applied to this model can also be used in random simulation models). The state of the dynamic biospheric system is computed per half year, starting in the year 1900 and ending in the year 2100. The model is composed of modules, which treat specific parts of the greenhouse problem. Modules get inputs from other modules. Also see Figure 1 and the references [4; 15; 16].

The sensitivity analysis techniques are applied to several modules. This paper concentrates on the carbon-cycle module; the dike raising modules are briefly discussed: see the shaded modules in Figure 1.

Note that there are alternative techniques for sensitivity analysis. *Latin Hypercube* sampling is a Monte Carlo method that is discussed at length in [8] and criticized in [3]; also see [10, pp. 143–145]. This technique was applied to several *IMAGE* modules in [13]; it gave results similar to the results of this paper. More sophisticated techniques do not treat the simulation model as a black box; they use analytical differential analysis; see [7; 14].

This article is organized as follows. First the need for sensitivity analysis is discussed, and the greenhouse case study is introduced. Then metamodels, which explain the input/output behavior of the underlying simulation model, are explained. The coefficients of the metamodel are estimated by least squares regression analysis. The resulting metamodel can be validated. Closely related to the metamodel specification is the selection of an efficient experimental design. All techniques are demonstrated by their application to several modules of the greenhouse simulation model.

Metamodeling through regression analysis

A simulation model maps its inputs into one or more outputs; hence a simulation model is a mathematical function (say) s( ). The inputs are parameters, input variables, and behavioral relationships (or submodules); see [10, p. 136]. These inputs are called *factors* in the statistical design of experiments. They may be represented by z_j with $j = 1, \ldots, k$ and k 1. In the greenhouse model all factors are quantitative, but the techniques also apply to qualitative factors. The case study concentrates on a single output variable, namely the global average atmospheric CO_2 concentration in the year 2100, which is denoted by y. (If there were several outputs, the technique could be applied per output.) This yields

$$y = s(z_1, \ldots, z_j, \ldots, z_k). \tag{1}$$

A mathematical function may be approximated by a Taylor series, under certain mathematical conditions. Suppose the initial approximation is

$$\hat{y} = \gamma_0 + \sum_{j=1}^{k-1} \sum_{h=j+1}^{k} \gamma_{jh} z_j z_h, \tag{2}$$

where γ_0 is the overall or grand mean, γ_j is the first-order or main effect of factor j, and γ_{jh} is the interaction between the factors j and h (that is, the effect of factor j depends on the level of factor h).

Note that the variables in approximation (2) may be functions of the variables in simulation model (1); for example, in (2) z_j may be replaced by $\log(z_j)$ or $1/z_j$. Then (2) remains linear in the parameters c, so linear regression analysis still applies; see [10, pp. 160–161].

The approximation in (2) is called a *metamodel* because it is a model of the input/output behavior of the underlying simulation model. The Taylor series argument may be one inspiration for the specification of such a metamodel. Because the mathematical conditions of the Taylor series do not hold in complicated simulation models, the validity of the metamodel must be checked. In other words, the metamodel is only an approximation. Before that model can be validated, it must be calibrated, that is, its coefficients or parameters γ must be estimated. Moreover there is a scaling problem. These issues are discussed now.

For simplicity's sake the interactions in the metamodel (2) are ignored temporarily. If the input variable z_j increases by one unit, then the output changes by γ_j units. We assume, however, that sensitivity analysis is meant to quantify the effect of a change of the input over its *whole* experimental area. (Next those 'important' factors are further investigated to validate and optimize the simulation model; if only optimization were the goal, then local marginal effects would suffice.) So the importance of factor j is measured by the difference between the outputs at the lowest and the highest value of that factor. Denoting those two extreme factor values by L_j and H_j respectively (so the 'experimental area' is a k-dimensional rectangle), the original variables z_j yield the *standardized variables* x_j, which range between –1 and +1:

$$z_j = a_j x_j + b_j \text{ with } a_j = \frac{H_j - L_j}{2} \text{ and } b_j = \frac{H_j + L_j}{2}. \tag{3}$$

The simple transformation (3) together with the original metamodel (2) yields the standardized metamodel

$$\hat{y} = \beta_0 + \textstyle\sum_{j=1}^{k} \beta_j x_j \displaystyle\sum_{j=1}^{k-1} \sum_{h=j+1}^{k} \beta_{jh} x_j x_h. \tag{4}$$

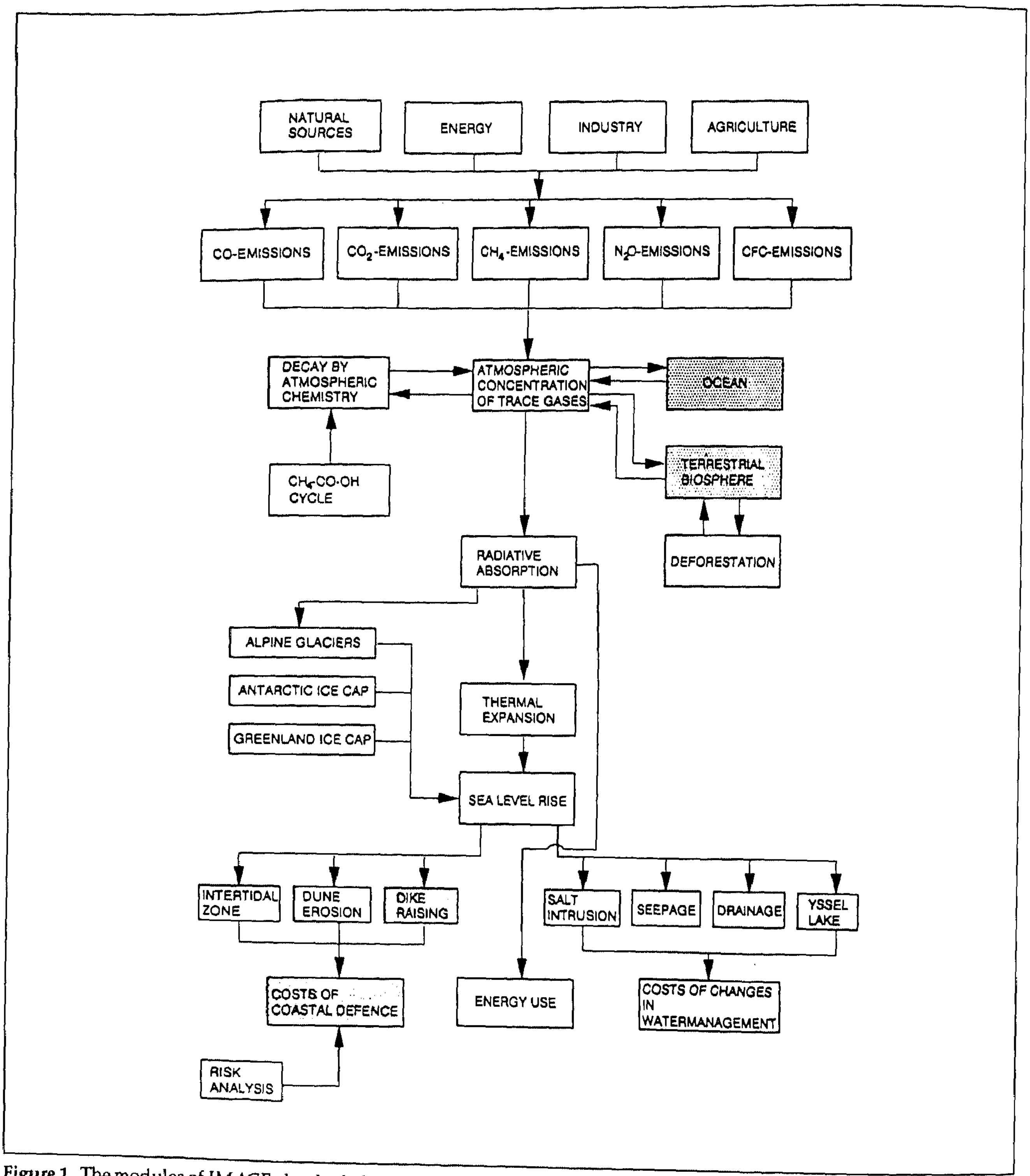

Figure 1. The modules of *IMAGE*; the shaded modules are submitted to sensitivity analyses

It is simple to prove that β_j reflects the *importance* of factor j: $\beta_j = \gamma_j (H_j - L_j)/2$, ignoring interactions. See [2].

Note that in its search for the optimum combination of the input factors, Response Surface Methodology (RSM) combines a first-order metamodel with the steepest ascent technique. That search should use neither the original nor the standardized model but the centered model

$$\hat{y} = \delta_0 + \sum_{j=1}^{k} \delta_j (z_j - \bar{z}_j) + \sum_{j=1}^{k-1} \sum_{h=j+1}^{k} \delta_{jh} (z_j - \bar{z}_j)(z_h - \bar{z}_h) , \quad (5)$$

where $\bar{z}_j = \sum_{i=1}^{n} z_{ij}/n$, which assumes that n combinations of input factors are simulated. See [2].

Calibration means that the parameters of the model are quantified. So the metamodel parameters β in (4) are estimated. Therefore the metamodel is fitted to the simulation data. Let q denote the number of parameters in the metamodel; in (4) q equals $1 + k + k(k-1)/2$. To get estimated parameter values $\hat{\beta}$, n combinations of the factor values are simulated. That set of simulated combinations yields the n x q matrix of independent variables X corresponding to the metamodel in (4):

$$X = \begin{bmatrix} 1, x_{11}, \dots x_{1k}, x_{11}x_{12}, \dots, x_{1,k-1}x_{1k} \\ 1, x_{i1}, \dots x_{ik}, x_{i1}x_{i2}, \dots, x_{i,k-1}x_{ik} \\ 1, x_{n1}, \dots x_{nk}, x_{n1}x_{n2}, \dots, x_{n,k-1}x_{nk} \end{bmatrix} \quad (6)$$

Example: Suppose there are three factors (k=3), which in combination i have the values +1, –1, and –1 respectively. (Remember that standardization means that in this combination factor 1 is at its highest level H_1, factor 2 is at its lowest level L_2, and so on; see equation (3). Then the interaction variable x_1x_2 has the value (+1) (–1) = –1 in this combination, and so on. Obviously β_0 corresponds to the 'variable' that is +1 in all combinations. So row i of X equals

$$x'_i = (+1, +1, -1, -1, -1, -1, +1).$$

The simulated output of combination i is y_i; see (1). Fitting the metamodel to the simulation data, using the *least squares* criterion, yields the estimated parameters

$$\hat{\beta} = (X'X)^{-1} X' y . \quad (7)$$

The least squares criterion yields unique estimates only if X is non-singular so that the inverse of (X'X) exists. A necessary condition is n q (the number of simulated factor combinations is not smaller than the number of parameters in the metamodel). This condition, however, is not sufficient. For example, if the factors 1 and 2 are changed simultaneously in the n combinations, then their two columns are identical ($x_{i1} = x_{i2}$ for i = 1 ,..., n) and X is singular. Obviously X is not singular if all its columns are *orthogonal* $\left(\sum_{i=1}^{n} x_{ij} x_{ih} = 0\right)$. Under certain statistical assumptions, an orthogonal matrix X is optimal; see the next section.

The calibrated metamodel can now be *validated*. One aspect is how well this metamodel fits the simulation data. One overall criterion is

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \quad (8)$$

where $\bar{y} = \sum_{i=1}^{n} y_i/n$. A 'perfect' fit means that $y_i = \hat{y}_i$ for all i, so the upper limit for R^2 is 1. Unfortunately, a lower threshold for R^2 is hard to give. Therefore we propose to compute the relative errors $(y_i - \hat{y}_i)/y_i$, which can be 'eyeballed' by the user.

Validating a model, however, usually means that the model is used to forecast the output; next that forecast is confronted with the true output. Therefore *cross-validation* should be used. So delete one combination (x'_i, y_i) from the old data set (X,y); denote the remaining set by (X_{-i}, y_{-i}). Reestimate the metamodel parameters β, analogous to (7):

$$\hat{\beta}_{-i} = (X'_{-i} X_{-i})^{-1} X'_{-i} y_{-i} . \quad (9)$$

Predict the output of combination i, not using the data of combination i:

$$\hat{y}_i = x'_i \hat{\beta}_{-i} . \quad (10)$$

Compute the forecast errors

$$e_{-i} = \hat{y}_{-i} - y_i . \quad (11)$$

The user may again evaluate the relative errors e_{-i}/y_i. This procedure is repeated for all i (i=1,...,n).

The errors e_i can be computed without applying the least squares criterion n times (to n–1 combinations). First no data are eliminated; see (7). Next the so-called 'hat' matrix H is computed:

$$H = X(X'X)^{-1} X' . \quad (12)$$

H has diagonal elements h_{ii}, which yield

$$e_{-i} = e_i / (1 - h_{ii}) . \quad (13)$$

Many modern regression analysis packages give those 'leave one out residuals'. See [10, p. 178] and [1, p. 13].

The mathematical analysis can be refined if a *statistical* (sub)model is added for the *fitting errors* e. Kleijnen (10, p. 164) assumes that these errors are normally and independently distributed with common variance (say) σ^2. Then the least squares algorithm yields the Best Linear Unbiased Estimators (BLUE); that is, the estimators have minimum variances and correct expected values. Those variances var $(\widehat{\beta}_j)$ are given by the main-diagonal elements of the variance-covariance matrix of $\widehat{\underline{\beta}}$:

$$\mathrm{cov}(\widehat{\underline{\beta}}) = (X'X)^{-1}\sigma^2. \tag{14}$$

The parameter σ^2 in (14) is estimated through the Mean Squared Error:

$$\widehat{\sigma}^2 = \sum_{i=1}^{n}(\widehat{y}_i - y_i)^2 / (n-q). \tag{15}$$

The estimated variances (or standard errors) of $\widehat{\beta}_j$ yield a t statistic with n – q degrees of freedom:

$$t_{n-q} = (\widehat{\beta}_j - \beta_j) / \widehat{\sigma}_j \quad (j=1, ..., q), \tag{16}$$

where $\widehat{\sigma}_j$ denotes $\{\widehat{var}(\widehat{\beta}_j)\}^{\frac{1}{2}}$ and β_j is the j[th] element of $\underline{\beta}$ (so β_1 in (16) is identical to β_0 in (4), β_2 in (16) is β_1 in (4),..., β_q in (16) is $\beta_{k-1,k}$ in (4)). The *significance* of $\widehat{\beta}_j$ can be tested statistically, using the t table for a given significance level or type-I error (say) α. For example, $\alpha = 0.05$ and $v = 12$ give the critical t value 2.18 in a two-sided test, which considers the absolute value of t_{n-q}.

Note that a more sophisticated model for the fitting errors is used by Sachs et al. [17]. They assume that the errors are not independent but form a stationary process with a specific correlation function. Also see [11].

Statistical design of experiments

The metamodel determines the experimental design. For example, a model with interactions such as (4) cannot be calibrated through a design that changes one factor at a time; see [10, pp. 266–267]. If purely quadratic terms $\beta_{jj}x_j^2$ are added to (4), then the variable x_j cannot be observed at only two levels (–1 and +1).

Given the metamodel, there is more than one design to calibrate that model. A necessary condition for the design is that the resulting matrix of independent variables X is non-singular; see (6) and (7). Consider, for example, a first-order model: in (4) the double summation term vanishes. That model has q = 1 + k effects, so a necessary condition is that the number of combinations satisfies: n k + 1. For k = 3, Table 1 gives two designs that give a non-singular X. Intuitively the 2^{3-1} design is attractive because it is *balanced:* each column of X has an equal number of plus and minus signs, and each pair of columns has an equal number of the four combinations (–, –), (–, +), (+, –), (+, +). If the classical statistical model for the errors is assumed, then the covariance matrix of $\widehat{\underline{\beta}}$ is given by (14). An orthogonal X *minimizes* the variances of $\widehat{\beta}_j$; see [10, p. 335].

If the metamodel includes interactions, then the number of effects increases considerably. To keep the number of combinations relatively small, the user may specify which interactions may be important; the remaining interactions are assumed to be negligible. Examples will be presented later.

The metamodel may be expanded with *purely quadratic* effects: $\sum_{j=1}^{k}\beta_{jj}x_j^2$ is added to (4). These quadratic effects quantify the curvature of the response surface. Then more than two values per factor must be simulated (otherwise all columns for x_j^2 are identical to the column for x_0). A classical design is the *central composite*

Table 1. Two designs for a first order model with k = 3

Combination	One factor at a time design			2^{3-1} design		
1	–1	–1	–1	–1	–1	+1
2	+1	–1	–1	+1	–1	–1
3	–1	+1	–1	–1	+1	–1
4	–1	–1	+1	+1	+1	+1

design: each factor is observed not only at –1 and +1 but also at the 'center point' 0 and at two other values, for example, –2 and +2 (together five values). The 2^{k-p} designs that is used to estimate main effects and interactions, are augmented with the center point plus 2k combinations:

$$\begin{array}{l} (0,\ 0,\ ...,\ 0) \\ (-2,\ 0,\ ...,\ 0) \\ (+2,\ 0,\ ...,\ 0) \\ (0,\ -2,\ ...,\ 0) \\ (0\ +2,\ ...,\ 0) \\ \quad . \\ \quad . \\ \quad . \\ (0,\ 0,\ ...,\ -2) \\ (0,\ 0,\ ...,+2). \end{array} \tag{17}$$

Next applications of metamodeling and experimental design will be presented. First the results for a relatively simple module of IMAGE will be discussed; then results and technical details for a more complicated module will be presented.

Dike raising in IMAGE

One module of IMAGE estimates the magnitude of the necessary dike raise and the resulting costs; see the lower part of Figure 1. Eleven factors are examined (k=11). An example of a factor is the "unit dike raising cost", which is the cost of increasing the dike level by one meter. Finding a valid metamodel takes several iterations; altogether nine different models (and their concomitant designs) are tried. In an early iteration the metamodel helped to detect a serious error in the underlying simulation model: the original module needed to be split into two modules such that the first submodule yields the dike raise necessary to keep the flooding probability under a fixed safety level, while the second submodule takes that raise as input and yields the costs as output. So metamodeling may serve *verification* of the simulation model. Moreover, metamodeling may show in which area the simulation model is valid; for factor combinations outside that experimental area the simulation is not a correct model.

To obtain a valid metamodel for the dike raising costs module, the ranges of the original input variables must be decreased. This makes sense mathematically, since a Taylor approximation is better in a smaller area. The final metamodel yields relative forecast errors smaller than 10%, which is acceptable for the IMAGE analysts. Nine of the eleven factors are significant, and so is one interaction. The most important factor is the "unit dike raising cost", as the analysts expected. The order of importance of the other factors was surprising, and gave more insight into the simulation model; for details see [18].

The carbon-dioxide cycle in IMAGE: ocean module

There are two modules for the CO_2 cycle in IMAGE: one for the oceans and one for the terrestrial biosphere; also see the upper part of Figure 1. This section covers the first module; the next section will discuss the second module.

The oceans show three CO_2 processes, described in [6]. For these processes ten factors are investigated; for example, factor 5 is thickness of ocean layers. For each factor a range is specified by the analysts; for example, factor 5 varies between 3,000 and 4,000 meters but factor 2 (diffusion coefficient) ranges between 3,716 and 5,984 cm^2/second. These variables are standardized, as described by (3). The analysts list eleven specific interactions that they think might be important; the remaining 34 interactions are neglected. To verify the design the reader should know that the following eleven interactions may be important: 1 3, 1 5, 2 3, 2 5, 3 6, 3 7, 5 6, 5 7, 5 8, 6 8 and 7 8, where 1 3 stands for β_{13}, and so on. So the metamodel is given by (4) with k = 10 and only eleven specific interactions β_{jh}.

A 'full factorial' design requires 2^{10} combinations, which takes too much computer time. The number of effects in the tentative metamodel is: q = 1 + 10 + 11 = 22. Hence a classical 2^{k-p} design with enough combinations requires: $n = 2^{10-p}$ 22 or p 5 (least squares applied to the whole data set requires n q, whereas cross validation requires n > q). There are many 2^{10-5} designs. Accounting for the eleven specific interactions, the following design is selected; details are given in [10, pp. 295–300]. Write down all $2^{k-p} = 2^5$ combinations of the factors 1,2,4,9, and 10. Write down element i of the column for factor 3 as the product of the elements i in the columns for the factors 9 and 10; that is, $x_{i3} = x_{i9} x_{i10}$ with i = 1,..., n and $n = 2^{k-p} = 2^5$ or in short-hand: 3 = 9 10. This is called a 'generator' of the design. The 2^{k-p} design is fully specified by its p = 5 generators

$$3 = 9\ 10\quad 5 = 4\ 10\quad 6 = 1\ 9\quad 7 = 2\ 9\quad 8 = 1\ 2\ 4, \tag{18}$$

where 8 = 1 2 4 stands for $x_{i8} = x_{i1} x_{i2} x_{i4}$. The generator 3 = 9 10 means that the main effect of factor 3 is *confounded* or *aliased* with the interaction between the factors 9 and 10; that is, $\widehat{\beta}_3 = \widehat{\beta}_{9\,10}$ and $E(\widehat{\beta}_3) = \beta_3 + \beta_{9\,10}$. If indeed the interaction $\beta_{9\,10}$ is negligible, then obviously this confounding is acceptable. Analogously, the generator 8 = 1 2 4 implies $E(\widehat{\beta}_8) = \beta_8 + \beta_{124}$ where b_{124} is a 'three factor' interaction. This interaction was not yet defined in this paper, but such high-order interactions are assumed negligible in metamodeling.

The 2^{10-5} combinations of this design are simulated, and the outputs are compared with the predictions of the calibrated metamodel. This results in relative errors exceeding 10% in eight combinations, which is considered unacceptable. Shrinking the ranges of the original variables does not help. Next the metamodel is expanded with purely quadratic effects. The central composite design of (17) requires 1 + 2k extra combinations. To save computer time, only five of the ten factors are investigated, namely those five factors that are significant in the previous metamodel. Because that metamodel is not valid, it is dangerous to use it for the selection of factors; the resulting new metamodel, however, will be validated again.

The quadratic model is used for the five factors 3,4,5,7, and 10. Only four (not all ten) interactions between these factors are conjectured to be important. So the number of effects excluding purely quadratic effects, is 1 + 5 + 4 = 10. Hence the 2^{k-p} design, which is part of the central composite design, must satisfy $n = 2^{5-p}$ 10 or p ≤ 1. So a single generator is selected, namely 4 = 3 7 10. These sixteen combinations are augmented with the eleven combinations that follow from (17). This experiment yields a calibrated metamodel, which is cross-validated. Moreover, six additional combinations are selected randomly, simulated, and compared with the predicted outcomes. Finally, a 'base' combination is examined; this combination is not the center combination (0,0,. . .,0) of (17), but is close to it; it is a combination intuitively specified by the analysts. All validation results are acceptable: the errors are smaller than 10%. The individual effects of this accepted metamodel are discussed next.

Because a statistical model for the fitting errors e is assumed, (16) gives the relevant t statistic with degrees of freedom n–q = (16+11) – (10+5) = 12. For $\alpha = 0.05$ the critical t value is $t_{12}^{\alpha/2} = 2.18$. Table 2 shows significant effects only, in decreasing order of importance.

Note that if no statistical model for the fitting errors were assumed, then the last column should be ignored. If the design were orthogonal, then 'significance' and importance would coincide: (14) through (16) imply that $t_{n-q} = \hat{\beta}_j / (\hat{\sigma}/\sqrt{n})$, so if effects are sorted in order of magnitude $(|\beta_j|)$, , they are sorted in order of significance $(|t_{n-q}|)$, . The central composite design, however, is not orthogonal, because the quadratic effects and the overall mean are not orthogonal.

Table 2. Significant effects of ocean module

Effect	Estimate	t Statistic
β_0	1074.66	154.08
β_5	–244.95	–139.35
β_4	158.37	98.10
β_3	51.77	32.07
β_{55}	27.79	14.05
β_7	18.78	11.63
β_{45}	–17.65	–8.93
β_{47}	–12.26	–6.20
β_{35}	–8.35	–4.22

Summarizing, originally ten factors are investigated for the ocean module. Because the metamodel without quadratic effects cannot be accepted, a model including such effects is specified. That model, however, is restricted to five factors. The latter model is accepted, and yields only four important factors. These factors have significant main effects, one significant quadratic effect, and three significant interactions.

The carbon-dioxide cycle in IMAGE: terrestrial biosphere module

The terrestrial biosphere module is described in [6]. The analysis of this module is presented, because the module contains many input variables, namely k = 62. These variables are described in [18]. There are designs that yield estimators of main effects without bias by possible interactions; moreover, these designs yield estimators of confounded interactions. These designs are called *resolution-IV* designs (see [10], p. 301). It can be proven that such designs require at least 2k combinations (so $k + 1 << n << 1 + k + k(k-1)/2$). So for k = 62 a 2^{k-p} design of resolution IV satisfies 2^{62-p} 124 or p = 55. Hence 55 generators must be selected. Each selection yields a specific confounding pattern of estimated effects. The analysts give 26 interactions that they think might be important. Based on that list, 55 generators are selected; see [18].

One estimated effect turns out to have the wrong sign: the effect is significantly positive whereas the analysts expect a negative effect. So the ranges of the input variables are decreased. Now the results become acceptable: the relative forecast errors are small, all 26 confounded interactions are non-significant, and all significant (unbiased) main effects have correct signs. There are only 13 significant main effects (significance is measured by the t statistic with 128 – (1+62+26) = 39 degrees of freedom). Finally the metamodel is validated through an experiment with twelve randomly selected extra combinations; its relative forecast errors vary between –5.5% and –0.10%.

Conclusion

Any simulation model requires sensitivity analysis. The simulation model can be treated as a black box, if the techniques of regression analysis and experimental design are applied. The regression model is a metamodel of the simulation model, and guides the experimental design. The design leads to efficient and effective experimentation.

The case study demonstrates that application of these statistical techniques requires knowledge of the underlying simulation model and real world system. For example, potentially important factors and their ranges must be given by the analysts. Some statistical expertise is needed to select the generators for the design.

The case study was a success. The metamodels give acceptable forecast errors. The significance of certain effects surprises the analysts. For example, quadratic effects in the ocean module were not expected; the major importance of the 'biotic growth' factor in the terrestrial module is also surprising. Another surprise is the 'bug' in the dike raising module; the metamodel shows that this module must be split into two submodules.

The sensitivity analysis of IMAGE took quite some time and effort, but this investment in metamodeling is judged to be profitable. The conclusions of this analysis will guide the development of an interactive version of IMAGE.

Summarizing, regression metamodels and experimental designs are useful in the sensitivity analysis of simulation models, as the case study demonstrates. Details on the techniques can be found in [10] and [12].

References

1. Atkinson, A. C., 1985. *Plots, Transformations, and Regression*. Clarendon Press, Oxford.
2. Bettonvil, B. and J. P. C. Kleijnen, 1990. "Measurement Scales and Resolution IV Designs: a Note", *American Journal of Mathematical and Management Sciences* 10: 309-322.
3. Easterling, R. G., 1986. " Letter to the editor." *Technometrics*, 28: 91–92.
4. Den Elzen, M. G. J. and J. Rotmans, 1988. *Simulatiemodel voor een Aantal Maatschappelijke Gevolgen van het Broeikaseffect voor Nederland*, (Simulation model for a number of societal consequences of the greenhouse effect for the Netherlands), RIVM Report No. 758471008, Bilthoven, Netherlands.
5. Gezondheidsraad, 1987. *2e Deeladvies inzake de CO_2-problematiek: Wetenschappelijke Inzichten en Maatschappelijke Gevolgen*, (2nd Report on the CO_2 problem: scientific insights and societal consequences), Staatsdrukkerij, The Hague, Netherlands.
6. Goudriaan J. and P. Ketner, 1984. "A Simulation Study for the Global Carbon Cycle, Including Man's Impact on the Biosphere", *Climatic Change* 6: 167–192.
7. Ho, Y. and X. Cao, 1991. *Perturbation Analysis of Discrete Event Dynamic Systems*, Kluwer, Dordrecht (Netherlands).
8. Iman, R. L. and J. C. Helton, 1988. "An investigation of uncertainty and sensitivity analysis techniques for computer codes." *Risk Analysis*, 8: 71–90.
9. Intergovernmental Panel on Climate Change (IPCC), 1990. "*Scientific Assessment of Climate Change*". Report of Working Group I, WMO/UNEP.
10. Kleijnen, J. P. C., 1987. *Statistical Tools for Simulation Practitioners*, Marcel Dekker, Inc., New York.
11. Kleijnen, J. P. C., 1990. *Statistics and Deterministic Simulation: Why Not?* Katholieke Universiteit Brabant (Tilburg University), Tilburg, Netherlands.
12. Kleijnen, J. P. C. and W. van Groenendaal, 1992. *Simulation: a Statistical Perspective*, Wiley, Chichester (United Kingdom).
13. Lammerts, I., 1989. *Onzekerheidsanalyse Toegepast op het Broeikassimulatiemodel IMAGE*, Uncertainty analysis applied to the greenhouse simulation model IMAGE), RIVM, Bilthoven, Netherlands.
14. McRae, G., 1989. *Designing Air Quality Control Strategies*. Chemical Engineering and Engineering and Public Policy, Carnegie-Mellon University, Pittsburgh (Pennsylvania).
15. Rotmans, J., 1990. *IMAGE: An Integrated Model to Assess the Greenhouse Effect*, Kluwer, Dordrecht (The Netherlands).
16. Rotmans, J., H. de Boois and R. J. Swart, 1990. "IMAGE: An Integrated Model for the Assessment of the Greenhouse Effect: the Dutch Approach", *Climatic Change*, 16: 331–356.
17. Sachs, J., W. J. Welch, T. J. Mitchell and H. P. Wynn, 1989. "Design and analysis of computer experiments." *Statistical Science*, 4: 409–435.
18. Van Ham, G., J. Rotmans and J. P. C. Kleijnen, 1990. *Een Gevoeligheidsanalyse via Metamodellen en Experimentele Proefopzetten, toegepast op een Simulatiemodel voor het Broeikaseffect*, (Sensitivity analysis through metamodels and experimental design, applied to a simulation model for the greenhouse effect), RIVM Report, No. 758471010, Bilthoven, Netherlands.

JACK P. C. KLEIJNEN is Professor of Simulation and Information Systems at the 'Katholieke Universiteit Brabant' (Tilburg University) in Tilburg, The Netherlands. He received his doctor's and master's degrees in Management Science, in 1971 and 1964 repectively, at that same university. He has published four books and more than 100 articles in international journals on simulation, statistics, operations research, computer science, etc. His first book entitled *Statistical Techniques in Simulation* (Dekker, N.Y., 1974/1975) received a Lanchester Prize Honorable Mention from the Operations Research Society of America, and was translated into Russian. He received a number of fellowships and awards, both nationally and internationally.

GREET VAN HAM graduated in 1982 for her Doctorandus-degree (Drs.) in Ecometrics at the Katholieke Universiteit Brabant (Tilburg University) in The Netherlands. She graduated on a sensitivity analysis, through metamodels and experimental design, applied to a simulation model for the greenhouse effect, performed at the RIVM Institute in Bilthoven. She's currently employed as a software engineer by VOLMAC Ltd., a software bureau in The Netherlands.

JAN ROTHMANS graduated in 1986 with an M.Sc. in Mathematics from the University of Delft, and a Ph.D., in Mathematics in 1990 from the University of Limburg (Maastricht) The Netherlands. He is currently a Professor in Global Change Modelling at the University of Limburg. His current research interests are in operations research, environmental modelling, and computer/simulation modeling of global atmospheric phenomena.