SENSITIVITY ANALYSIS OF A SIMULATION MODEL BY EXPERIMENTAL DESIGN AND METAMODELING

Vonk Noordegraaf, A. and Nielen, M.

Farm Management Group, Department of Social Sciences, Wageningen University, Hollandseweg 1, 6706 KN Wageningen, The Netherlands

Introduction

To support decision makers in the national BHV1-eradication program in The Netherlands, the spatial, dynamic and stochastic simulation model InterIBR-endemic was developed (Vonk Noordegraaf et al., 2000). InterIBR-endemic simulates the spread and control of BHV1 within and between cattle farms in The Netherlands. This model contains many uncertain input factors and for validation purpose we want to evaluate the sensitivity of model-outcome to these factors. For this, we applied the techniques of Design of Experiments (DOE) and metamodeling (Kleijnen, 1998). This paper will present background on this technique and first results of the analysis.

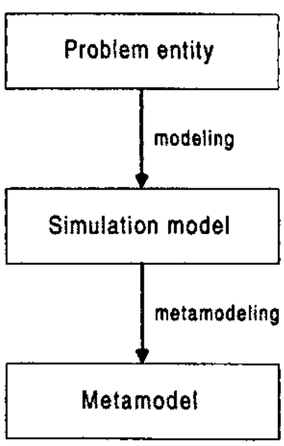

Figure 1. Metamodel, simulation model and problem entity .

In a simulation context, DOE can be defined as selecting out of the great number of possible combinations of factor levels, the set that actually needs to be simulated in an experiment with the simulation model in order to quantify factor effects. The simulation model is run for this set of combinations, and the resulting input-output data are analysed by regression analysis to derive conclusions about the importance (sensitivity) of the factors (Kleijnen, 1998). This analysis is based on a metamodel, which is defined as a model of the simulation model (Kleijnen and Groenendaal, 1992). Figure 1 shows the relationships among metamodel, simulation model and problem entity. This research focuses on the process of metamodeling, and testing of metamodel validity with respect to the simulation model.

Material and methods: the metamodeling process

Kleijnen and Sargent (2000) suggest the following 10 steps for development of a metamodel:
1- Determine the goal of the metamodel
2- Identify the inputs and their characteristics
3- Specify the domain of applicability
4- Identify the output variables and their characteristics
5- Specify the accuracy required of the metamodel
6- Specify the metamodel's validity measures and their required values
7- Specify the metamodel, and review this specification
8- Specify a design including tactical issues, and review the DOE
9- Fit the metamodel
10- Determine the validity of the fitted metamodel
In this paper only a few steps are highlighted.

Step 2: Identify the inputs and their characteristics
A total of 31 factors used in the simulation model are selected as potential independent (X) variables in the metamodel. These factors are related to disease spread, and have in common that they are uncontrollable for decision makers and their estimation contains uncertainty.

Step 3: Specify the domain of applicability
Sensitivity analysis requires that each factor has at least 2 levels, therefore a lower (0 and upper (1) level is determined for each factor. This determines the experimental frame for which the metamodel is to be valid, assuming linearity between these points.

Step 4: Identify the output variables and their characteristics
We are interested in multiple outputs, for each of these outputs a metamodel is specified:
Y1: mean number of weeks to reach a prevalence level of 5% in dairy cattle population
Y2: mean total discounted costs (1000 Dfl.) in this period
Y3: mean number of outbreaks per year on certified dairy farms
Y4: mean prevalence level in the dairy cattle population after 4 years of control (%)

Step 7: Specify the metamodel, and review this specification
In the analysis, initially each metamodel is specified as a simple first-order polynomial in which the independent variables (X) are standardised at either 0 or 1:

$$y_i = \beta_0 + \sum_{h=1}^{k} \beta_h x_{i,h} + e_i$$

In step 9 (model fit), simulation I/O data is also checked for the presence of interactions between some pre-specified independent variables. We assume that only interactions between factors with significant main effects should be included.
Step 8: Specify a design including tactical issues, and review the DOE
To obtain a resolution-3 design, giving unbiased estimators of the k=31 main effects and overall mean of the first-order polynomial regression model, a minimum of 32 (n = k + 4 – [k modulo 4]) factor combinations is required. A 2^{k-p} fractional factorial design matrix D was constructed with k=31 and p=26, by making a full factorial design for the first 5 columns, and using 26 generators to obtain the other columns. The resulting design matrix is orthogonal, thereby minimising the variance of the estimated factor effects. However, if there are interactions between factors, estimators of main effects based on the resolution-3 design are biased. In this case, unbiased estimators can be achieved by applying the foldover theorem: adding the mirror image –D to the original resolution-design matrix D (Van Groenendaal and

Kleijnen, 1997). This design is called a resolution-4 design, containing in our case a total of 64 scenarios. Because we are dealing with random simulation, multiple replications for each scenario are desired. We performed the minimum of 2 replications for each scenario.

Step 9: Fit the metamodel
A total of 64 scenarios (combinations of factor levels), each consisting of 2 replications, were run with the simulation model. Resulting I/O data were used to select and fit each metamodel, applying the techniques of Ordinary Least Squares (OLS).We applied a backwards elimination procedure to select significant ($p<0.05$) main effects in each regression model. Then, each regression model was tested for the significance of interactions between factors that had significant main effects.

Results

Only results for the metamodel related to dependent variable 'mean number of outbreaks per year on certified dairy farms' will be shown here. Table 1 shows the factors and coefficients of each factor that were included in the final metamodel for this dependent variable.

Table 1. Factors in the final metamodel for dependent variable 'mean number of outbreaks per year on certified dairy farms' ($R^2_{adjusted} = 0.78$).

Factor	Coefficient	SE	p-value
Intercept	-59	32.8	0.078
Local spread	220	40.2	0.000
Reactivation transport	96	23.2	0.000
Professional contact	125	23.2	0.000
R_0 non-vaccinated herds	29	32.8	0.386
Hygiene certified farms	-20	32.8	0.545
Local spread x hygiene	-151	46.5	0.002
Local spread x R_0 non-vaccinated herds	191	46.5	0.000

For each scenario, this fitted regression metamodel can be used to predict the value of the dependent variable and compare with the simulation realisation. A plot of the prediction and simulation result is given in Figure 2. Pearson's correlation coefficient between metamodel prediction and simulation result was 0.90.

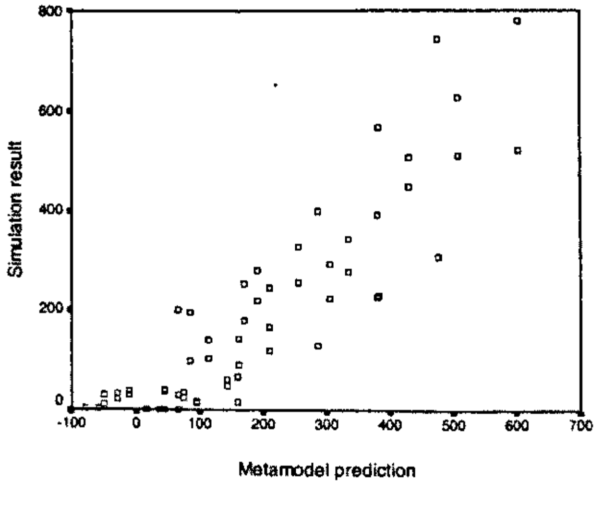

Figure 2. Plot of metamodel predictions and simulation results for dependent variable 'mean number of outbreaks per year on certified dairy farms' dependent variable.

Conclusions and discussion

Data generated by the computer experiment described in this study are still being analysed, using regression techniques as logistic, multivariate and tobit regression. The techniques op Experimental Design and metamodeling are considered very useful in the sensitivity analysis of complex simulation models, and certainly should be applied to more case studies in the area of economic modelling of animal health. In comparison with simple sensitivity analysis (changing one factor at a time), the use of experimental design supports a structural approach, providing more accurate estimators of factor main effects and enabling the estimation of interactions among factors.

References

Kleijnen, J.P.C., Sargent, R.G., 2000. A methodology for fitting and validating metamodels in simulation. European Journal of Operational Research, 120, 14-29.

Kleijnen, J.P.C., 1998. Experimental design for sensitivity analysis, optimization, and validation of simulation models. In: Banks, J. (Ed.). Handbook of simulation: principles, methodology, advances, applications, and practice. New York, Wiley, 173-223.

Kleijnen, J.P.C., 1999. Validation of models: statistical techniques and data availability. In: P.A. Farrington, H.B. Nembhard, D.T. Sturrock and G.W. Evans (Eds.), Proceedings of the 1999 Winter Simulation Conference, 647-654.

Van Groenendaal, W.J.H., Kleijnen, J.P.C., 1997. On the assessment of economic risk: Factorial design versus Monte Carlo methods. Journal of Reliability Engineering and System Safety, 57, 91-102.

Vonk Noordegraaf, A., Nielen, M., De Jong, M.C.M., Frankena, K., Lam, T.J.G.M. Franken, P. & Dijkhuizen, A.A., 2000. Simulation and monitoring to support the BHV1 eradication program in The Netherlands. In: Salman, M.D., Morley, P.S., Ruch-Gallie, R. (Eds.), Proceeding of the 9th ISVEE Symposium, Colorado, 514-516.