45

KATHOLIEKE UNIVERSITEIT BRABANT
TILBURG UNIVERSITY
POSTBOX 90153
5000 LE TILBURG
THE NETHERLANDS

[redacted]

R 11

Simulation Models

DEPARTMENT OF ECONOMICS

RESEARCH MEMORANDUM

VERIFICATION AND VALIDATION OF MODELS

Jack P.C. Kleijnen

FEW 571

Communicated by Prof.Dr. F.A. van der Duyn Schouten

Verification and validation of models

Jack P.C. Kleijnen
Department of Information Systems and Auditing,
Katholieke Universiteit Brabant (Tilburg University),
P.O. Box 90153,
5000 LE Tilburg,
Netherlands.

September 1992

Verification and validation of models

Jack P.C. Kleijnen
Department of Information Systems and Auditing, Katholieke Universiteit Brabant (Tilburg University), P.O. Box 90153, 5000 LE Tilburg, Netherlands.

This paper surveys many issues in the verification and validation of models, especially simulation models in operations research. For verification it discusses (a) checking of intermediate outputs through tracing and statistical testing per module, (b) comparing final outputs with analytical results, using statistical tests, (c) animation, and (d) general good programming practice. For validation it discusses (a) obtaining real-world data, which may be scarce or abundant, (b) some simple tests for comparing model and real data, such as graphical, Schruben-Turing, and t tests and, (c) a new statistical procedure, based on regression analysis, for testing whether model and real responses are positively correlated, (d) sensitivity and risk analyses, (e) system dynamics type of modeling, (f) relationships between simulation and other types of models, including white and black box models, (g) Documentation and credibility assessment. Finally a bibliography with 38 references is given for further study.

Keywords: simulation, statistics, time series, philosophy, practice, econometrics.

1. Introduction

Terminology in the area of verification and validation (or V & V) of computerized models is not standard; see Barlas and Carpenter (1990, p.164, footnote 2). This paper uses the definitions of V & V that are given in the classic simulation textbook by Law and Kelton (1991, p. 299): '*Verification* is determining that a simulation computer program performs as intended, i.e., debugging the computer program.... *Validation* is concerned with determining whether the conceptual simulation model (as opposed to the computer program) is an accurate representation of the system under study'. This paper assumes that verification aims at 'perfect' computer programs, in the sense that they have no programming errors left (they may be made more efficient and more user friendly). Validation, however, does not aim at perfect models, since the perfect model is the real system itself (any model is by definition a simplification of the real system). A quote from another well-known author on V & V in simulation, namely Sargent (1991, p. 38), further clarifies V & V issues in the various *phases* of modeling: 'the *conceptual model* is the mathematical/logical/verbal representation (mimic) of the problem entity developed for a particular study; and the *computerized model* is the conceptual model implemented on a computer. The conceptual model is developed through an *analysis and modelling phase,* the computerized model is developed through a *computer programming and implementation phase,* and inferences about the problem entity are obtained by conducting computer experiments on the computerized model in the *experimentation phase* '.

V & V are important issues in operations research (OR). A computer program with bugs may generate output that is sheer nonsense, or worse, it may generate subtle nonsense that goes unnoticed. A nonvalidated model may lead to decisions that make sense in a fantasy world but not in the real world.

There is no standard theory in the areas of V & V; neither is there a well organized 'box of tools'. There does exist a plethora of philosophical theories, statistical techniques, software practices, and so on. (For system dynamics types of simulation, Barlas 1989 does propose a well organized validation procedure with six steps; see Section 3.5.) The *objective* of this paper is to survey many of these (kaleidoscopic) elements. Moreover, it introduces a new statistical technique for validation (based on familiar regression analysis). Unfortunately, it will turn out that there are no perfect solutions for the problems of V & V: the modeling process has elements of art as well as science, as the title of one of the first books on simulation witnessed; see Tocher (1963). Nevertheless systems analysts and users should be aware of the various options in V & V. It will also be noted that these problems occur not only in simulation models, but also in other types of models (for instance, econometric models) and in other types of computer programs (for example, bookkeeping programs). This paper, however, concentrates on experience in the area of simulation, emphasizing simulation in operations research.

Note that in practice the process of performing a simulation study may give so much insight into the system under study that the study is stopped before the stages of V & V have been reached.

This paper is organized as follows. Section 2 discusses verification. Section 3 examines validation. Section 4 gives conclusions, and is followed by a list of 37 references (to avoid dragging along a cumulative list of everything published on simulation, only those publications are included that either seem to deserve special mentioning or that are not mentioned in the references of this paper, which include two bibliographies, namely Balci and Sargent 1984a, and DeMillo, McCracken, Martin, and Passafiume 1987).

2. Verification

Once the model has been programmed, the analysts/programmers must check if this computer code contains any programming errors. Several techniques are applicable, but none is perfect.

2.1 Verification of intermediate output

The analysts may calculate some intermediate simulation results manually, and compare these results with outputs of the simulation program. Getting all *intermediate* results from a computer program automatically is called *tracing*. Even if the analysts do not wish to calculate intermediate results by hand, they can still 'eyeball' the program's trace and look for 'bugs'. Modern simulation software provides tracing facilities and more advanced 'debuggers'; for details see Pegden, Shannon, and Sadowski (1990, pp. 137-148).

Many simulation programs are very big. Then good programming practice requires that this program be designed *modularly* (no 'spaghetti programming'). Then the analysts can 'divide and conquer', that is, they may try to verify the total computer code, module by module. Different members of the team may check different modules. Some examples now follow.

The analysts may test the *pseudorandom number generator* separately, assuming they had to program that generator themselves or they do not trust the software supplier's expertise. Random numbers are a statistical ideal: they are continuous statistical variables that are uniformly distributed (over the interval from zero to one), and they are statistically independent. In other words, random numbers are a model of reality; that reality is created by the pseudorandom number generators. So the output of the generator must be compared with an ideal. Deviations between reality and ideal may be caused by programming errors (many generators require either machine programming or rather sophisticated programming in a higher language) or by a wrong model for the generator. So testing a generator is a mixture of verification and validation. Such a mixture often occurs in V & V of modeling. Schriber (1991, p. 317) points out that GPSS/H automatically computes chi-square statistics to test the hypothesis that the pseudorandom numbers used are uniformly distributed. Kleijnen and Van Groenendaal (1992) give a detailed discussion of different types of pseudorandom number generators and of many tests to verify their correctness, including tests for various types of statistical dependence among pseudorandom numbers.

An example of pure verification (not validation) is the testing of the subroutines that generate samples from certain *non-uniform distributions*. The analysts may wish to take samples from, say, N(100, 10). They may think that the computer gives them normal variates with expectation 100 and standard deviation 10, whereas actually the variates generated have a *variance* of 10. The cause of this confusion is lack of standard notation: some authors and some software use the notation $N(\mu, \sigma)$, whereas others use $N(\mu, \sigma^2)$. Similar confusion arises for exponential distributions: some authors use the parameter (say) λ to denote the mean number of 'successes', but others use that symbol to denote the rate at which successes occur. The analysts may also specify the wrong unit of measurement, for example, seconds instead of minutes (so the results are wrong by a factor 60). To verify that the subroutine does what it is intended to do, the analysts should first of all read the documentation of the subroutine. Next they may compute the average of the sampled variable (for example, service time in a queuing simulation), and compare that average with its expected value (which is known in a simulation study; for instance, service times are sampled from an exponential distribution with a known mean, namely the mean that is input to the simulation program). Systematic deviations between the observed average and the theoretical mean may be tested through the t test (a slightly more complicated example of a t test will be discussed in equations 1 through 3). Such systematic deviations occur when, for example, the analysts mix up the variance and the standard deviation of the normal distribution. (Random deviations between the sample average and the population mean always occur. The variance reduction technique (VRT) known as control variates can be used to correct the crudely estimated simulation response such that the variance of the resulting estimator is smaller; see Kleijnen and Van Groenendaal 1992, pp. 200-201).

Instead of testing only the mean, the analysts may test the whole distribution of the random variable. Then they can apply a goodness-of-fit test such as the well-known chi-square and Kolmogorov-Smirnov tests; see Kleijnen (1987, pp. 94-95) for a survey of goodness of fit tests.

Besides the modules for generating uniform and non-uniform variates there are other modules, at least a module that generates the final simulation response, which is of real interest to the users of the simulation program. Their verification is discussed next.

2.2 Comparing final model outputs with analytical results

The final output of (say) a queuing simulation program may result only after thousands of customers have been processed, so the steady state has been reached and its mean waiting time can be estimated. Indeed, if the traffic intensity of the system is high, millions of customers must be simulated. Extremely long simulation runs are also necessary when estimating rare events like blocking or breakdowns in highly reliable systems. Verifying such types of simulation responses by hand or by eyeballing the trace is practically impossible. Restricting attention to short time series may be misleading (as many commentators on animation have pointed out; also see Section 2.3). The analysts may then verify the simulation response by running a *simplified* version of the simulation program with a known *analytical solution*. This approach assumes that the analysts can indeed find such a version, but this is not such an unrealistic assumption. Simulation without any theoretical background means that the analysts are 'simulating their own ignorance'! For example, in the study of factories seen as queuing systems the analysts can and should use some of the many textbooks on queuing theory, which offer formulas for the steady state expectations of several types of response (mean waiting time of jobs and mean utilization of machines) for models with Markovian arrival and service times with (say) n servers: M/M/n models. So first the analysts may run the simulation program with exponential (Markovian) arrival and service times, only to verify the correctness of the computer program. If that response 'agrees with' the known mean response (see the statistical test in equation 1 later on), then they run the simulation program with non-Markovian input variables to find the responses that are of real interest to the users. They must then hope that this minor change in the computer program does not introduce new bugs.

There is also much literature for more complicated queuing systems. These systems consist of servers, not only in parallel but also in sequence, and with customers who can follow different paths through the queuing network. For certain queuing networks (for example, without finite buffers for work in process) the analysts can compute steady state solutions numerically. There is also software that gives these analytical, numerical, and simulation results; see Kleijnen and Van Groenendaal (1992, p. 127). Much research is going on in queuing theory with applications in computer, communications, and manufacturing systems. In other areas (for example, inventory management and econometrics) there is also a substantial body of theory available. So a stream of publications and software may help the simulation analysts to find models that are related to their simulation models and that have analytical or numerical solutions.

General systems theory emphasizes that the scope of a study can be reduced by either studying a subsystem only (say, queuing at one specific machine in the factory) or by restricting the type of variables (for example, financial variables only). In this way the analysts may find simplified models with known responses, and may verify the total queuing simulation program or certain modules.

The variance of (say) $\underline{x}$, the response of the simulation model that is really of interest and has no known solution (random variables are underlined), can be decreased by correcting that response for the deviation between $E(\underline{y})$, the known response of the related analytical model, and $\underline{y}$, the simulated response of the simplified model. This yields another form of control variates (estimate the simulation response through $\underline{x} - \{\underline{y} - E(\underline{y})\}$; for more details see Kleijnen 1974, pp. 162-163). So the effort of simulating a model with a known solution pays of in two ways: it helps detecting programming errors in the simulation model without a known solution (verification), and it helps improve the statistical accuracy of the response of the latter model (variance reduction technique).

Analysts should understand that in the steady state the system is still stochastic (but the probability law that governs the stochastic process no longer depends on the initial state). So the use of mathematical statistics is necessary. Analysts may test, for example, that the expected value of the response $\underline{x}$ of the simplified simulation program equals the steady state expected value μ (which is computed analytically or numerically, as discussed in the preceding paragraph):

$$H_0\colon E(\underline{x}) = \mu. \tag{1}$$

A simple test of this hypothesis is possible under the usual assumptions of a normally and independently distributed (NID) simulation response $\underline{x}$ with mean μ and constant variance σ_x^2. To estimate this variance, the analysts may partition the simulation run into (say) m subruns and compute the estimated variance s_x^2 of $\underline{x}$:

$$\underline{s}_x^2 = \sum_{i=1}^{m} (\underline{x}_i - \overline{\underline{x}})^2/(m - 1), \tag{2}$$

where $\overline{\underline{x}}$ denotes the average of the m subrun averages, which is identical to the average of the whole run. Then the test statistic becomes

$$\underline{t}_{m-1} = (\underline{\overline{x}} - \mu)/(\underline{s}_x/\sqrt{m}). \tag{3}$$

Kleijnen and Van Groenendaal (1992, pp.190-195) present several alternative approaches (such as renewal analysis) to the estimation of the variance of the simulation response in the steady state.

In practice, however, most simulation studies concern the behavior of the real system in the *transient* state, not the steady state ('in the long run we all are dead', said Keynes). For example, the users may be interested in the waiting times during the next day, under various scheduling algorithms (or priority rules), so the simulation run stops as soon as the end of that simulated day is reached. Such types of simulation are called 'terminating' simulations. When verifying such a simulation, there are usually no analytical or numerical solutions known: most solutions hold in the steady state only. The analysts may then first simulate a non-terminating variant of their simulation model, for verification purposes only. For validation purposes they must next change the simulation program, that is, they must introduce the terminating event (in the example this event is the 'arrival' of the end of the working day). As pointed out above, they must then hope that this minor change in the computer program does not introduce new bugs. There is no guarantee: the introduction in Section 1 has already mentioned that there are no perfect solutions for the problems of V & V.

There is a technical complication, as virtually all simulation programs have *multiple* responses: the computer program transforms (say) S inputs into T outputs with $S \geq 1$ and $T \geq 1$. Of course that transformation must be correct for all simulation responses of interest to the users. Consequently the probability of rejecting a null-hypothesis like (1) for one or more response types increases as the number of responses increases, even if the program is correct (this property follows from the definition of the type I or α error of a statistical test; also see Section 3.2). Fortunately there is a very simple solution: when comparing the t value that results from (3) with the critical value in the table for the t statistic (denoted as $t_{m;\,\alpha/2}$ in a two-sided t test with type I error rate probability fixed at α), the analysts should replace the traditional value for α (for example, 0.05) by α/T (for instance, if average waiting time of customers and machine utilization are the two responses, then $T = 2$). Note that this implies that bigger discrepancies between the theoretical mean and the simulation response are accepted. This solution is based on *Bonferroni's inequality*, and it can be proved that it keeps the overall 'experimentwise' error rate below the value α. Alternatives to this combination of univariate techniques (like the t test) and Bonferroni's inequality are multivariate techniques, which are more sophisticated and not always more powerful; see Balci and Sargent (1984b) and Kleijnen and Van Groenendaal (1992, pp. 144, 155).

In some situations the analysts know the theoretical outputs, provided the inputs are *deterministic*. One class of examples is provided by inventory models: for a constant demand per period (and under certain other assumptions) the classic 'economic order quantity' (EOQ) model holds, which has a known analytical solution. For a single server queuing model, constant arrival and service times (say) λ and μ respectively with $\lambda > \mu$ give a known utilization rate μ/λ for the single server and zero waiting times for all customers. Examples of economic models with deterministic inputs and known outputs are given in Kleijnen and Van Groenendaal (1992, pp.58-64). In all these examples, no mathematical statistics is needed: the simulation response must be identical to the theoretical response, except for numerical inaccuracies.

2.3 Animation

To verify the computer program of a dynamic system, the analysts may use *animation*. They then present moving pictures of the simulated system, to the users. Since the users are most familiar with the corresponding real system, they are well qualified to detect programming errors (and conceptual errors too, but that concerns validation). Well-known examples are simulations that show how vehicles defy the laws of nature and cross through each other, and simulations that have customers who get lost ('evaporate') during the simulation run. Most simulation researchers agree that animation may be dangerous too, as the analysts and users tend to concentrate on very short simulation runs so the problems that tend to occur in long runs only go unnoticed.

2.4 General good programming practice

Software engineers have developed several procedures for writing good computer programs and for verifying software, in general (not specifically for mathematical models, including simulation programs). This is such a vast area of research that only a few key terms are given here: modular programming, chief programmer's approach, structured walk-throughs, correctness proofs. Details are given in Baber (1987), DeMillo et al. (1987), and Whitner and Balci (1989). The book by DeMillo et al. has a comprehensive bibliography.

3. Validation

Once the analysts believe that the model is programmed correctly, they must face the next question: does this computer program give a valid model? The introduction in Section 1 stated: 'validation is concerned with determining whether the conceptual model (as opposed to the computer program) is an accurate representation of the system under study'. This raises many questions.

A very old *philosophical* question is: do humans have accurate knowledge of reality or do they have only flickering images of reality, as Plato stated? In the sixteenth through the current centuries this question was again raised by philosophers such as Descartes, Kant, Popper, and Kuhn; see Barlas and Carpenter (1990) and Naylor, Balintfy, Burdick, and Chu (1966, pp.310-320). This paper, however, does not discuss general, abstract, philosophical questions like 'can a human being know the world?', but specific, practical, operational questions like 'should one machine be added to this factory?'.

3.1 Obtaining real-world data

The system analysts must explicitly formulate the laws that they think govern the 'system under study', that is the system that exist or is planned to exist in the real world. The *system concept*, however, implies that the analysts must subjectively decide on the boundary of that system and on the attributes to be quantified. So to obtain a valid model, they should try to measure the inputs and outputs of the real system, and the attributes of intermediate variables. In practice, data are available in different degrees, as will be explained next.

Sometimes it is difficult or impossible to obtain relevant data. For example, in a simulation study of the recovery of the US economy after a nuclear attack, it was (fortunately) impossible to get the necessary data. In the simulation of whale population dynamics, a major problem is that data on whale behavior are hard to obtain. In the latter example more effort is needed for data collection. In the former example the analysts may try to show that the exact values of the input data are not critical. These problems will be further analyzed in Section 3.4 on sensitivity analysis.

Usually, however, it is possible to get some data. For example, decision makers want to choose among different options for the real system. So the analysts simulate different scenarios, in order to enable the decision makers to select a 'good' scenario. Usually the analysts have data only on the existing system variant or on a few historical variants; for example, the existing manufacturing system with its current scheduling rule.

In the military, however, it is common to conduct *field tests* in order to obtain data on *future* variants. Kleijnen and Alink (1992) present such a military case study, namely mine hunting at sea by means of a sonar system: mine fields are created (not by the enemy but by the friendly navy) and a mine hunt is executed in this field, to collect data. Fossett, Harrison, Weintrob, and Gass (1991) also discuss several field tests for military simulations. Shannon (1975, pp. 231-233) briefly discusses military field tests, too.

In some applications there is an *overload* of input data, namely if these data are collected electronically. For example, in the simulation of the performance of computer systems, the analysts use hardware and software monitors to collect data on the system state at regular time points (say, each nanosecond) or at each system state change (event). These data can be used to drive the simulation. Another example is provided by point-of-sale (POS) systems:

all transactions at the supermarket check-out are recorded electronically (data capture at the source). In the near future more applications will be realized; for example, the geographical positions of trucks and railroad cars will be determined and communicated electronically, and electronic data interchange (EDI) among companies will generate lots of data.

The real-world data (either scarce or abundant) may show *observation error*, which complicates the comparison of real-world and model time series. Barlas (1989, p. 72) and Kleijnen and Alink (1992) discuss observation errors in a theoretical and a practical situation respectively.

Sometimes the model is meant to predict, not relative responses (which correspond to different scenarios), but *absolute* responses. For example, in the mine-hunting case study one of the questions does not concern different naval tactics and technical sonar parameters, but asks for the probability of detecting mines in a certain area (the 'huntability' of the area: is the probability of success such that it makes sense to do a mine sweep?). Validation is more difficult when absolute instead of relative answers are needed; also see the discussion leading to (12).

Suppose the analysts wish to validate the model as a whole (validation of individual modules with observable inputs and outputs proceeds similarly; modules without such inputs and outputs can be subjected to sensitivity analyses; see Section 3.4). Then the analysts feed the model with real-world input data in historical order. This is sometimes called *trace driven* simulation. They run the simulation program, obtain the simulation output, and compare that time series with the historical time series for the output of the existing system. So they do not sample the simulation input from the (raw or smoothed) histogram of real-world input values (after the simulation model has been validated, different scenarios should be compared for sampled inputs, not historical inputs, because it is 'certain' that in the future this history will never be repeated exactly). Instead they use the historical input values in historical order: $(z_{-T}, z_{-T+1}, \ldots, z_{-1}, z_0)$ where z denotes the historical input and $T + 1$ denotes the size of the historical sample. The further the analysts go back into the past, the more data they get and the more powerful the validation test will be, unless they go so far back that different laws governed the system. For example, many econometric models do not use data prior to 1945, because the economic infrastructure changed drastically during World War II.

3.2 Some simple techniques for comparing model and real-world data

The output data of the real system and the simulated system can be plotted such that the horizontal axis denotes time ($t = -T, -T + 1, \ldots, -1, 0$) and the vertical axis denotes the real and simulated values respectively. The users may *eyeball timepaths* to decide whether the model 'accurately' reflects the phenomena of interest. For example, do the simulation data in a business cycle study indicate an economic downturn at the time such a slump occurred in practice? And do the simulation data in a queuing study show the same saturation behavior (such as exploding queuelengths and blocking) as happened in the real system? Barlas (1989, p. 68) gives a system dynamics example that seems to allow subjective graphical analysis only, since the time series (simulated and real) show 'highly transient, non-stationary behavior'.

Another simple technique is the *Schruben-Turing* test. The analysts then present a mixture of simulated and real time series to their clients, and challenge them to identify (say) the data that were generated by computer. Of course, these clients may correctly identify some of the data by mere chance; this, however, the analysts can test statistically. Turing introduced such an idea to validate Artificial Intelligence computer programs: users are challenged to identify which data (say, chess moves) are generated by computer, and which data are results of human reasoning? Schruben (1980) applies this concept to the validation of simulation models. He adds several statistical tests and presents some case studies.

Instead of subjectively eyeballing the simulated and the real time series, the analysts can use *mathematical statistics* and obtain reproducible, objective, quantitative data about the quality of the model. The problem, however, with the statistical analysis of simulation output data is that these data form a time series, whereas elementary statistical procedures assume identically and independently distributed (i.i.d.) observations. Nevertheless it is possible to derive independent observations, so that elementary statistical theory can be applied, as the next example will demonstrate.

Let $\underline{w}_i$ denote the *average* waiting time on day i in the simulation and $\underline{v}_i$ the similar response of the real system, with $i = 1,\ldots,n$ (so n days are simulated and observed in reality respectively). These averages do not need to be computed from a steady state time series of individual waiting times, but may be calculated from (say) the individual waiting times of all customers who arrive on day i between 8 a.m. and 5 p.m., which includes the start-up, transient phase. Then these n averages $\underline{w}_i$ are i.i.d. Suppose further that the historical arrival and service times are used to drive the simulation model. Statistically this trace driven simulation means that there are n (i.i.d.) 'paired' differences $\underline{d}_i = \underline{w}_i - \underline{v}_i$ (difference between simulated and real average waiting time on day i when using the same arrival and service times). Then the t statistic is analogous to (3):

$$\underline{t}_{n-1} = (\overline{\underline{d}} - \delta)/(\underline{s}_d/\sqrt{n}), \tag{4}$$

where $\overline{\underline{d}}$ denotes the average and $\underline{s}_d$ represents the estimated standard deviation of $\underline{d}$ (so $\overline{\underline{d}}$ is the average of the n differences between the n average simulated and n average real waiting times per day). Suppose that for $\delta = 0$, (4) gives a value t_{n-1} that is significant (the t value exceeds the critical value $t_{n-1;\,\alpha/2}$). Then the simulation model is rejected, since the simulation model gives average waiting times per day that deviate significantly from reality. If $\delta = 0$ gives a non-significant t_{n-1}, then the conclusion is that the simulated and the real means are 'practically' the same so the simulation is 'valid enough'. This interpretation, however, deserves some comments.

Strictly speaking, the simulation is only a model, so simulating 'very many' days ('large' n) would show that $\delta = 0$ gives a significant t_{n-1}. When testing the validity of a model through statistics like (4), the analysts can make either a 'type I' or a 'type II' error. So they may reject the model while the model is valid: type I or α error. Or they may accept the model while the model is not valid: type II or β error (the probability of a β error is the complement of the 'power' of the test, which is the probability of rejecting the model when the model is wrong indeed). The probability of a type I error in simulation is also called the *model builder's risk*; the type II error probability is the *model user's risk*. The power of the statistical test increases as the model specification error δ increases (this power can be computed through the 'noncentral' t statistic, which is the t statistic with nonzero mean). A significance or 'critical' level α means that the type I error probability equals α. Obviously the probability of a β error increases as α decreases, given a fixed number of simulated days (sample size n). To decrease both error probabilities the analysts may increase the number of simulated days. They may also try to decrease var($\underline{w}$), the variance of the simulated system response, through VRTs like control variates and antithetics. If those techniques work, they decrease the variance of $\underline{w}$ and hence the variance of $\underline{d}$, the difference between the simulated and the real responses. Then the denominator of (4) has a smaller expected value and the t test becomes more powerful. Running the simulation with the same inputs as observed in the real world is a form of using common random numbers, which is the simplest and most popular VRT. Balci and Sargent (1984b) analyze the theoretical tradeoffs among α and β error probabilities, sample size, and so on.

The selection of a value for α is problematic. Popular values are 0.10 and 0.05. Theoretically speaking, the analysts should determine these values by accounting for the financial consequences (or more generally, the disutilities) of making type I and type II errors respectively. Such an approach is indeed followed in decision theory and in Bayesian analysis. This paper, however, is based on the classic statistical theory.

3.3 A new statistical technique for comparing model and real-world data

A *most stringent* validation test requires not only that the means of the simulated and the historical responses are identical as in (4), but also that if a historical observation exceeds its mean then the corresponding simulated observation (that is the simulation response that uses the same inputs as the historical observation did) tends to exceed its mean, too. For example, $\underline{v}$ and $\underline{w}$ (defined above equation 4) should not only have the same mean ($\delta = 0$) but also be *positively correlated*: $\rho > 0$ where ρ denotes the linear correlation coefficient. To investigate this correlation, analysts may plot w versus v. This paper,however, formalizes that graphical approach through the use of least squares and a statistical test. This test presumes that certain statistical assumptions hold and -to make the test powerful- that enough observations are available. Testing the hypothesis of positively correlated $\underline{v}$ and $\underline{w}$ is simple if $\underline{v}$ and $\underline{w}$ are bivariate normally distributed (which is a realistic assumption in the example that lead to (4), because of a central limit theorem that can be applied to autocorrelated variables like waiting times of successive customers). It can be proved that such a bivariate normal distribution implies

$$E(\underline{w} | \underline{v} = v) = \beta_o + \beta_1 v. \tag{5}$$

So analysts can plot w as a function of v, and use ordinary least squares (OLS) to estimate the intercept and slope of the straight line that passes through the 'cloud' of points (v_i, w_i); the OLS formulas can be found in any statistics text. The proposed stringent test calls the model valid if the following composite hypothesis holds:

$$H_o: \beta_o = 0 \text{ and } \beta_1 = 1, \tag{6}$$

which implies $E(\underline{w}) = E(\underline{v})$ (which was also tested through equation 4). Moreover, it can be found in any statistics text that

$$\beta_1 = \rho\sigma_w / \sigma_v. \tag{7}$$

So if $\beta_1 = 1$ and $\rho < 1$ then $\sigma_w > \sigma_v$, that is, if the model is not perfect ($\rho < 1$) then its variance exceeds the real variance. (Further, if $\beta_1 = 1$ and $\sigma_w = \sigma_v$ then $\rho = 1$, which is an unrealistic case because it means that the simulation model gives responses that are identical to the real world responses; also see the introduction in Section 1; if $\beta_1 =$ 1 and $\sigma_w < \sigma_v$ then $\rho > 1$, which violates the statistical model.) To test the hypothesis (6), the analysts should compute the Sum of Squared Errors (SSE) with and without that hypothesis (which correspond with the 'reduced' and the 'full' regression model respectively), and compare these two values, as follows. Based on (5) the analysts can compute

$$\underline{\hat{w}}_i = \hat{\beta}_0 + \hat{\beta}_1 \underline{v}_i, \tag{8}$$

which yields

$$\underline{SSE}_{full} = \sum_i^n (\underline{w}_i - \underline{\hat{w}}_i)^2. \tag{9}$$

Next they should compute the SSE under the composite hypothesis of (6); obviously a restricted regression model gives a higher SSE. That hypothesis implies $\hat{w} = v$, so

$$\underline{SSE}_{reduced} = \sum_1^n (\underline{w}_i - \underline{v}_i)^2. \tag{10}$$

It can be proved that the following expression is an F statistic with degrees of freedom 2 (the number of parameters in the hypothesis of equation 6) and n - 2 (the degrees of freedom of the SSE for the full regression model with n observations and two estimated parameters):

$$\underline{F}_{2,n-2} = \{(\underline{SSE}_{reduced} - \underline{SSE}_{full})/2\}/\{\underline{SSE}_{full}/(n-2)\}. \tag{11}$$

If the computed F statistic is significantly high, the analysts should reject the hypothesis in (6) and conclude that the simulation model is not valid. Details on this F test can be found in Kleijnen (1987, pp.156-157).

The analysts may formulate a *less stringent* validation requirement: the means are not necessarily equal, but the simulated and the real responses are positively correlated. This requirement makes sense if the simulation is used to predict relative responses (as in sensitivity analysis), not absolute responses (also see Section 3.1). To test this hypothesis the analysts should formulate the null-hypothesis

$$H_o: \beta_1 \leq 0. \tag{12}$$

To test this null-hypothesis, a t statistic can be used, as any textbook on regression analysis shows. This test means that analysts reject the null-hypothesis of (12) and accept the simulation model if there is strong evidence that the simulated and the real responses are *positively* correlated.

The two tests developed above use familiar statistical techniques, namely regression analysis and testing, but combine these techniques with the issues of validation, in a novel way.

Note that statistical analyses as in (4) through (12) require many observations to make the tests powerful, as has been stated at several places above. In validation, however, there are often not many observations on the real system; see Section 3.1. If there are very many observations, then not only the means of the simulated and the real time series can be compared but also the various *autocorrelations* (lag 1, 2, etc.). Spectral analysis is a sophisticated technique that estimates the autocorrelation structure of the simulated and the historical time series respectively, and compares these two structures. Unfortunately, that analysis is rather difficult, and (as stated) requires long time series. Barlas (1989, p. 61) criticizes Box-Jenkins models for the same reasons. So as the introduction in Section 1 has already mentioned, there are no perfect solutions for the problems of V & V.

3.4 Sensitivity analysis and risk analysis

Some authors (for example, Banks 1989 and Barlas 1989) claim that a model should remain valid under *extreme* conditions. But Zeigler (1976, p. 30), who wrote a fundamental book on the theory of modeling and simulation, emphasizes the concept of *experimental frame,* which he defines as 'a limited set of circumstances under which the real system is to be observed or experimented with'. He observes that 'a model may be valid in one experimental frame but invalid in another'. This paper has already mentioned (see Section 3.1) that going far back into the past may yield historical data that are not representative of the current system; that is, the old system was ruled by different laws. Similarly, a model is accurate only if the values of its input data remain within a certain area. For example, Bettonvil and Kleijnen's (1991) sensitivity analysis shows that a (deterministic) simulation model of the greenhouse effect of carbon dioxide (CO_2) and other gases is valid only if the simulation input values range over a relatively small area.

Readers familiar with regression metamodeling in simulation will appreciate the following example: a regression model of *first* order is a good approximation of the input/output behavior of a simulated M/M/1 system, only if the traffic load is 'low'; see Kleijnen and Van Groenendaal (1992).

Most simulation models have *many* input variables. To detect the important inputs the analysts may use *experimental designs* combined with regression analysis (or Analysis of Variance, ANOVA). For example, Kleijnen, Rotmans, and Van Ham (1992) apply experimental designs and regression analysis to the same simulation model of the greenhouse effect that was mentioned above. These designs and analysis give estimates of the effects of the various inputs. These estimated effects should have the right *signs*: some inputs are known to increase the global temperature. Wrong signs indicate computer errors or conceptual errors. Indeed both Kleijnen et al. (1992) and Kleijnen and Alink (1992) give examples of estimated sensitivity estimates with the wrong signs, which lead to corrections of the simulation models.

The *magnitudes* of the sensitivity estimates show which inputs are important. For important inputs the analysts must collect accurate information on the input values that may occur in practice. If the inputs are under the decision makers' control, these inputs should be steered in the right direction; for example, in the greenhouse case the governments should restrict emissions of the gases concerned.

Classic experimental designs, however, are not adequate to identify the important inputs of a simulation model, when the simulation study is still in its early phase so *very many inputs* (say, hundreds or thousands of inputs) may be potentially important. Such a situation calls for *screening*. Bettonvil and Kleijnen (1991) derive a special technique based on sequential experimentation with the simulation model and splitting up the aggregated inputs as the experiment proceeds, until finally the important individual inputs are identified and their effects are estimated. They apply this technique to the ecological simulation mentioned above.

Inputs may be *qualitative;* examples are the queuing or priority rules in a production planning simulation and the emission control scenarios in an ecological simulation. Then the analysts can estimate the effects of the quantitative inputs *per* policy or scenario. If these estimates do not vary with the policy, then there are no interactions between the quantitative and the qualitative inputs.

So this paper emphasizes that sensitivity analysis should be applied to find out which inputs are really important, and that collecting information on those inputs is worth the effort. Nevertheless it may be impossible or impractical to collect reliable information on those inputs, as the whale and the nuclear attack examples in Section 3.1 have already demonstrated. Then the analysts may apply *risk analysis*. So first they derive a probability distribution of input values, using the clients' expert knowledge. Next they use Monte Carlo sampling to generate input values that are fed into the simulation model, which yields a probability distribution of output values. For technical details on risk analysis the reader is referred to Kleijnen and Van Groenendaal (1992). The relationship between sensitivity and risk analyses requires more research; see Kleijnen (1990) and (1992).

Some authors use model *calibration*, that is, they adjust the simulation model's parameters (using some algorithm for function minimization) such that the simulated output deviates minimally from the real output. Obviously, those latter data can *not* be used to validate the model. Examples of calibration can be found in ecological modeling; see Beck (1987). Another example is provided by the mine hunting simulation in Kleijnen and Alink(1992), which uses an artificial parameter to steer the simulation response into the direction of the observed real responses. Obviously calibration is a last resort that must be employed if the system is a black box (as in econometrics and other social sciences).

This type of calibration must be distinguished from *cross-validation* in metamodeling, which was mentioned above. This cross-validation uses some data to estimate (or calibrate) the regression metamodel, whereas it employs some other data to validate the resulting metamodel. For details see Kleijnen and Van Groenendaal 1992, pp. 156-157).

3.5 System dynamics type of simulation

System dynamics is more than a technique: it is a world view. The resulting models are solved by means of simulation. This type of simulation is not of the discrete event type (queuing models are). System dynamics models consist of sets of nonlinear difference equations, which are evaluated at equidistant points of time (in operations research, inventory models and corporate models have the same characteristic). System dynamics models are highly aggregated, and aim not at exact estimates of system responses (say, the inventory costs per year), but at dynamic characteristics (oscillating and exploding time series). System dynamics is explained and compared with other types of simulation in Kleijnen and Van Groenendaal (1992, pp. 78-91, 130-131).

The *validation* of system dynamics models is discussed in Barlas and Carpenter (1990). This article deserves one critical comment. They state (on page 149) 'If a critic can show that one of the model equations does not make sense (does not agree with an obvious causality), then the model is refuted...'. But system dynamics relies on exponential delay functions, which may be said to be 'model equations' that do *not* agree with 'obvious causality'; that is, these functions are hard to understand and require training. Queuing simulation, however, does use intuitive knowledge about the real system: a job arrives, looks for an idle machine, and so on.

Barlas (1989) presents several *statistical techniques* for the validation of system dynamics models; for example, trend estimation and comparison between the simulated and real output time series, autocorrelation tests (which are related to spectral analysis, discussed in Section 3.3). One technical comment on Barlas (1989, p.63) is that he seems unaware of Bonferroni's inequality, which can be used to obtain a simple solution to the problem of testing estimated autocorrelations for *several* lag values (lag 1, 2, ...); Bonferroni's inequality was discussed in Section 2.2.

More references on system dynamics models and their validation are given in Barlas and Carpenter (1990); see also Kleijnen (1980, pp.137) and Wolstenhome (1990, pp. 58-60).

3.6 Relationships between simulation and other types of models: white versus black boxes

Karplus (1983) discusses a whole spectrum of mathematical models, ranging from black box (noncausal) models in the social sciences through gray models in ecology to white box (causal) models in physics and astronomy. The validation of simulation models in operations research is closely related to the validation of other mathematical models in OR, such as the models used in regression analysis, inventory control, and linear programming.

This paper has already mentioned some typical aspects of simulation models; for example, the *time series* character of its inputs and outputs (because simulation is dynamic), and the *random* noise in stochastic simulation. Other models share some of these characteristics with simulation models. For example, inventory and econometric models may also be dynamic and stochastic. Another typical aspect of many simulation studies is that their *conceptual* models are based on common sense and on direct observation of the real system; that is, the latter system is a *white box*. For example, a simulation model of a factory modeled as a queuing network represents intuitive knowledge about the real system: a job arrives, looks for an idle machine in the first stage of the production process, leaves the machine upon completion of the required service, goes to the second stage of its fabrication sequence, and so on. Connecting the models for subsystems gives the total simulation model, which grows in complexity and - hopefully - realism. The analysts cannot apply such a *bottom-up* approach in other types of mathematical models. Animation may help to obtain *face validity* of white box simulation models.

In some application areas, however, the simulation model are *black box* models. Examples are plentiful in aggegrated econometric modeling, which is performed by national planning agencies and large corporations. These models use macro-economic consumption functions, which relate total, aggregated consumption of a certain country to Gross National Product (GNP); see the examples in Kleijnen and Van Groenendaal (1992, pp. 57-69). The validation of black box models is more difficult, since the analysts can measure input and output data but not the internal relationships and the internal data. This problem is also known as the *observability* of systems; also see Zeigler (1976). In black box models the emphasis in validation is on prediction, not explanation; also see Barlas and Carpenter (1990).

3.7 Documentation and credibility assessment

A model's validity is determined by its assumptions. Therefore the analysts should state these assumptions in the model's documentation. (It might be claimed that being explicit about one's assumptions is the difference between a scientist and a politician.) In practice, however, many assumptions are left implicit, deliberately or accidently.

The importance of documentation is discussed at length by Fossett et al. (1991). On page 711 they define *assessment* as 'a process by which interested parties (who were not involved in a model's origins, development, and implementation) can determine, with some level of confidence, whether or not a model's result can be used in decision making'. Obviously documentation is necessary to enable users to assess a simulation model. The authors mention V & V as other important components of assessment.

Fossett et al. (1991, p. 712) further define *credibility* as 'the level of confidence in [a simulation's] results'. They present a framework for assessing the credibility of a simulation. Their framework comprises 14 inputs, but these inputs have also been discussed in this paper, explicitly or implicitly. They apply their framework to three military weapon simulations.

This paper has already shown that V & V has many aspects, involves different parties in the modeling process, and requires good documentation. Hence it is not strange that Gass (1984) proposes to produce four manuals, namely for analysts, users, programmers, and managers respectively.

Note that the lack of good documentation is a problem, not only with simulation programs but also with other types of mathematical models and with software in general; see Davis and Olson (1985). Documentation is not further discussed in this paper, as this essay focusses on V & V.

3.8 Supplementary literature

V & V of simulation models have been discussed in many textbooks on simulation. Examples are Banks and Carson (1984), Law and Kelton (1991, pp. 298-324), and Pegden et al. (1990, pp. 133-162). These books give many additional references.

Some case studies were mentioned above. In addition, Kleijnen (1988) gives a production-planning case study, and Carson (1989) presents a cigarette fabrication case study.

Dekker, Groenendijk, and Sliggers (1990) discuss the verification and validation of models that are used to compute air pollution. These models are employed to issue permits for building new factories and the like.

Banks (1989) proposes control charts, which have already gained popularity in quality control. Reckhow (1989) discusses several more statistical techniques.

Hodges (1991) gives a more polemical discussion of validation.

Balci and Sargent (1984a) give a detailed bibliography. The references of this paper update and augment that bibliography.

4. Conclusion

This paper surveyed a variety of techniques that seem useful in verification and validation (V & V) of models, especially simulation models in operations research. It emphasized techniques that yield reproducible, objective, quantitative data about the quality of the model.

For *verification* it discussed (a) checking of intermediate model outputs through tracing and statistical testing per module, (b) comparing final model outputs with analytical results, using statistical tests, (c) animation, and (d) general good programming practice. For *validation* it discussed (a) obtaining real-world data for trace driven simulation, which may be scarce or abundant, (b) some simple tests for comparing model and real data, such as graphical, Schruben-Turing, and t tests and, (c) a new statistical procedure, based on regression analysis, for testing whether model and real responses are positively correlated, (d) sensitivity and risk analyses, (e) system dynamics type of simulation, (f) relationships between simulation and other types of models, including white and black box models, (g) Documentation and credibility assessment. (h) supplementary literature; in total 37 references were given for further study.

The essay has demonstrated the usefulness of mathematical statistics in V & V, as the equations (1) through (12) illustrated. Nevertheless it is indisputable that analysts and users of a model should be convinced of its validity, not only by statistics but also in other ways. Some ways were presented above; for example, face validity of a simulation model can be established by animation (not statistics).

It seems hard to prescribe a fixed order in which the various V & V techniques should be applied (in some applications certain techniques do not apply at all). Practice shows that these techniques are applied in a haphazard way. It may be hoped that in the future, analysts and users will pay more attention to the various aspects of V & V and will apply some of the available V & V techniques, which were surveyed in this paper. Nevertheless, modeling will remain both an art as well as a science.

References

Baber R., *The Spine of Software; Designing Provable Correct Software: Theory and Practice*, John Wiley & Sons, Inc. Chichester, England, 1987.

Balci O. and R.G. Sargent, A bibliography on the credibili ty, assessment and validation of simulation and mathematical models, *Simuletter*, 15, 3, 1984(a), pp. 15-27.

Balci O. and R.G. Sargent, Validation of simulation models via simultaneous confidence intervals, *American Journal of Mathematical and Management Science*, 4, 3 and 4, 1984(b), pp. 375-406.

Banks J., Testing, understanding and validating complex simulation models, *Proceedings of the 1989 Winter Simulation Conference*, 1989.

Banks J. and J.S. Carson, *Discrete-event System Simulation*, Prentice-Hall, Englewood Cliffs, New Jersey, 1984.

Barlas Y., Multiple tests for validation of system dynamics type of simulation models, *European Journal of Operational Research*, 42, no. 1, September 1989, pp. 59-87.

Barlas Y. and S. Carpenter, Philosophical roots of model validation: two paradigms. *System Dynamics Review*, 6, no. 2, Summer 1990, pp. 148-166.

Beck, M.B., Water quality modeling: a review of the analy sis of uncertainty, *Water Resources Research*, 23, no. 8,August 1987, pp. 1393-1442.

Bettonvil B. and J.P.C. Kleijnen, Identifying the important factors in simulation models with many factors, Tilburg University, 1991.

Carson J.S., Verification and validation: a consultant's perspective, *Proceedings of the 1989 Winter Simulation Conference*, 1989.

Davis G.B. and M.H. Olson, *Management Information Systems: Conceptual Foundations, Structure and Development*, Mc-Graw-Hill, New York, 1985.

Dekker C.M., A. Groenendijk and C.J. Sliggers, Kwaliteits riteria voor modellen om luchtverontreiniging te berekenen (Quality criteria for models to compute air pollution), Report 90, VROM, Leidschendam, Netherlands, 1990.

DeMillo R.A., W.M. McCracken, R.J. Martin and J.F.Passafi ume, *Software Testing and Evaluation*, The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, 1987.

Fossett C.A., D. Harrison, H. Weintrob and S.I. Gass, An assessment procedure for simulation models: a case study, *Operations Research*, 39, 5, Sept.-Oct. 1991, pp. 710-723.

Gass S.I., Documenting a computer-based model, *Interfaces*, 14, 3, May-June 1984, pp. 84-93.

Hodges, J.S., Six (or so) things you can do with a bad model. *Operations Research*, 39, no. 3, May-June 1991, pp. 355-365.

Karplus, W.J., The spectrum of mathematical models, *Per spectives in Computing*, 3, no.2, May 1983, pp. 4-13.16.

Kleijnen J.P.C., *Statistical Techniques in Simulation, Part I*. Marcel Dekker, Inc., New York, 1974.

Kleijnen J.P.C., *Computers and Profits: Quantifying Finan cial Benefits of Information,* Addison-Wesley, Reading, Massachusetts, 1980.

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

Kleijnen J.P.C., Simulation and optimization in production planning: a case study, Tilburg, 1988. (To appear in DSS Journal, 1992/1993.)

Kleijnen J.P.C., Statistics and deterministic simulation models: Why not?, Proceedings of the 1990 Winter Simulation Conference, 1990.

Kleijnen J.P.C., A note on Wild and Pignatiello's experi mental design strategy, *Simulation, 58*, no. 6, June 1992, pp.393-394.

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

Kleijnen J.P.C. and G.A. Alink, Validation of simulation models: Mine-hunting case study, Tilburg University, 1992.

Kleijnen J.P.C. and W. van Groenendaal, *Simulation: a Sta tistical Perspective*, Wiley, Chichester, 1992.

Law A.M. and W.D. Kelton, Simulation Modeling and Analysis, second edition, McGraw-Hill Book Company, New York, 1991.

Naylor, T.H., J.L. Balintfy, D.S. Burdick and K. Chu, *Com puter Simulation Techniques*. John Wiley & Sons, Inc., New York, 1966.

Pegden C.P., R.E. Shannon and R.P. Sadowski, *Introduction to Simulation using SIMAN,* McGraw-Hill, Inc. New York, 1990.

Reckhow K.H., Validation of simulation models: philosophi cal and statistical methods of confirmation, in: *Systems & Control Encyclopedia,* M.D. Singh (ed.), Pergamon Press, 1989.

Sargent R.G., Simulation model verification and validation, *Proceedings of the 1991 Winter Simulation Conference*, 1991.

Schriber, T.J., *An Introduction to Simulation Using GPSS/H*. John Wiley & Sons, New York, 1991.

Schruben L.W., Establishing the credibility of simulations, *Simulation,* 34, 3, 1980, pp. 101-105.

Shannon R.E., *Systems Simulation: the Art and Science*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
Tocher, K.D., *The Art of Simulation*. The English Universi ties Press Ltd., London, 1963
Whitner R.B. and O. Balci, Guidelines for selecting and using simulation model verification techniques, *Proceedings of the 1989 Winter Simulation Conference*, 1989.
Wolstenhome E.F., *Systems Enquiry: A System Dynamics Appro ach*, John Wiley & Sons, Inc., Chichester, England, 1990.
Zeigler B., *Theory of Modelling and Simulation*, Wiley Int erscience, New York, 1976.

Acknowledgements

This paper is based on Chapter 11 of Kleijnen and Van Groenendaal (1992), which benefitted from comments made by Jerry Banks (Georgia Institute of Technology), David Kelton (University of Minnesota), and Bob Sargent (Syracuse Unoversity). The resulting version benefitted from discussions with Willem van Groenendaal (Tilburg University).

IN 1991 REEDS VERSCHENEN

466 Prof.Dr. Th.C.M.J. van de Klundert - Prof.Dr. A.B.T.M. van Schaik
Economische groei in Nederland in een internationaal perspectief

467 Dr. Sylvester C.W. Eijffinger
The convergence of monetary policy - Germany and France as an example

468 E. Nijssen
Strategisch gedrag, planning en prestatie. Een inductieve studie binnen de computerbranche

469 Anne van den Nouweland, Peter Borm, Guillermo Owen and Stef Tijs
Cost allocation and communication

470 Drs. J. Grazell en Drs. C.H. Veld
Motieven voor de uitgifte van converteerbare obligatieleningen en warrant-obligatieleningen: een agency-theoretische benadering

471 P.C. van Batenburg, J. Kriens, W.M. Lammerts van Bueren and R.H. Veenstra
Audit Assurance Model and Bayesian Discovery Sampling

472 Marcel Kerkhofs
Identification and Estimation of Household Production Models

473 Robert P. Gilles, Guillermo Owen, René van den Brink
Games with Permission Structures: The Conjunctive Approach

474 Jack P.C. Kleijnen
Sensitivity Analysis of Simulation Experiments: Tutorial on Regression Analysis and Statistical Design

475 C.P.M. van Hoesel
An $O(nlogn)$ algorithm for the two-machine flow shop problem with controllable machine speeds

476 Stephan G. Vanneste
A Markov Model for Opportunity Maintenance

477 F.A. van der Duyn Schouten, M.J.G. van Eijs, R.M.J. Heuts
Coordinated replenishment systems with discount opportunities

478 A. van den Nouweland, J. Potters, S. Tijs and J. Zarzuelo
Cores and related solution concepts for multi-choice games

479 Drs. C.H. Veld
Warrant pricing: a review of theoretical and empirical research

480 E. Nijssen
De Miles and Snow-typologie: Een exploratieve studie in de meubelbranche

481 Harry G. Barkema
Are managers indeed motivated by their bonuses?

482 Jacob C. Engwerda, André C.M. Ran, Arie L. Rijkeboer
Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^{T}X^{-1}A = I$

483 Peter M. Kort
A dynamic model of the firm with uncertain earnings and adjustment costs

484 Raymond H.J.M. Gradus, Peter M. Kort
Optimal taxation on profit and pollution within a macroeconomic framework

485 René van den Brink, Robert P. Gilles
Axiomatizations of the Conjunctive Permission Value for Games with Permission Structures

486 A.E. Brouwer & W.H. Haemers
The Gewirtz graph - an exercise in the theory of graph spectra

487 Pim Adang, Bertrand Melenberg
Intratemporal uncertainty in the multi-good life cycle consumption model: motivation and application

488 J.H.J. Roemen
The long term elasticity of the milk supply with respect to the milk price in the Netherlands in the period 1969-1984

489 Herbert Hamers
The Shapley-Entrance Game

490 Rezaul Kabir and Theo Vermaelen
Insider trading restrictions and the stock market

491 Piet A. Verheyen
The economic explanation of the jump of the co-state variable

492 Drs. F.L.J.W. Manders en Dr. J.A.C. de Haan
De organisatorische aspecten bij systeemontwikkeling
een beschouwing op besturing en verandering

493 Paul C. van Batenburg and J. Kriens
Applications of statistical methods and techniques to auditing and accounting

494 Ruud T. Frambach
The diffusion of innovations: the influence of supply-side factors

495 J.H.J. Roemen
A decision rule for the (des)investments in the dairy cow stock

496 Hans Kremers and Dolf Talman
An SLSPP-algorithm to compute an equilibrium in an economy with linear production technologies

497 L.W.G. Strijbosch and R.M.J. Heuts
Investigating several alternatives for estimating the compound lead time demand in an (s,Q) inventory model

498 Bert Bettonvil and Jack P.C. Kleijnen
Identifying the important factors in simulation models with many factors

499 Drs. H.C.A. Roest, Drs. F.L. Tijssen
Beheersing van het kwaliteitsperceptieproces bij diensten door middel van keurmerken

500 B.B. van der Genugten
Density of the F-statistic in the linear model with arbitrarily normal distributed errors

501 Harry Barkema and Sytse Douma
The direction, mode and location of corporate expansions

502 Gert Nieuwenhuis
Bridging the gap between a stationary point process and its Palm distribution

503 Chris Veld
Motives for the use of equity-warrants by Dutch companies

504 Pieter K. Jagersma
Een etiologie van horizontale internationale ondernemingsexpansie

505 B. Kaper
On M-functions and their application to input-output models

506 A.B.T.M. van Schaik
Produktiviteit en Arbeidsparticipatie

507 Peter Borm, Anne van den Nouweland and Stef Tijs
Cooperation and communication restrictions: a survey

508 Willy Spanjers, Robert P. Gilles, Pieter H.M. Ruys
Hierarchical trade and downstream information

509 Martijn P. Tummers
The Effect of Systematic Misperception of Income on the Subjective Poverty Line

510 A.G. de Kok
Basics of Inventory Management: Part 1
Renewal theoretic background

511 J.P.C. Blanc, F.A. van der Duyn Schouten, B. Pourbabai
Optimizing flow rates in a queueing network with side constraints

512 R. Peeters
On Coloring j-Unit Sphere Graphs

513 Drs. J. Dagevos, Drs. L. Oerlemans, Dr. F. Boekema
Regional economic policy, economic technological innovation and networks

514 Erwin van der Krabben
Het functioneren van stedelijke onroerend-goed-markten in Nederland - een theoretisch kader

515 Drs. E. Schaling
European central bank independence and inflation persistence

516 Peter M. Kort
Optimal abatement policies within a stochastic dynamic model of the firm

517 Pim Adang
Expenditure versus consumption in the multi-good life cycle consumption model

518 Pim Adang
Large, infrequent consumption in the multi-good life cycle consumption model

519 Raymond Gradus, Sjak Smulders
Pollution and Endogenous Growth

520 Raymond Gradus en Hugo Keuzenkamp
Arbeidsongeschiktheid, subjectief ziektegevoel en collectief belang

521 A.G. de Kok
Basics of inventory management: Part 2
The (R,S)-model

522 A.G. de Kok
Basics of inventory management: Part 3
The (b,Q)-model

523 A.G. de Kok
Basics of inventory management: Part 4
The (s,S)-model

524 A.G. de Kok
Basics of inventory management: Part 5
The (R,b,Q)-model

525 A.G. de Kok
Basics of inventory management: Part 6
The (R,s,S)-model

526 Rob de Groof and Martin van Tuijl
Financial integration and fiscal policy in interdependent two-sector economies with real and nominal wage rigidity

527 A.G.M. van Eijs, M.J.G. van Eijs, R.M.J. Heuts
Gecoördineerde bestelsystemen
een management-georiënteerde benadering

528 M.J.G. van Eijs
Multi-item inventory systems with joint ordering and transportation decisions

529 Stephan G. Vanneste
Maintenance optimization of a production system with buffercapacity

530 Michel R.R. van Bremen, Jeroen C.G. Zijlstra
Het stochastische variantie optiewaarderingsmodel

531 Willy Spanjers
Arbitrage and Walrasian Equilibrium in Economies with Limited Information

IN 1992 REEDS VERSCHENEN

532 F.G. van den Heuvel en M.R.M. Turlings
Privatisering van arbeidsongeschiktheidsregelingen
Refereed by Prof.Dr. H. Verbon

533 J.C. Engwerda, L.G. van Willigenburg
LQ-control of sampled continuous-time systems
Refereed by Prof.dr. J.M. Schumacher

534 J.C. Engwerda, A.C.M. Ran & A.L. Rijkeboer
Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^*X^{-1}A = Q$.
Refereed by Prof.dr. J.M. Schumacher

535 Jacob C. Engwerda
The indefinite LQ-problem: the finite planning horizon case
Refereed by Prof.dr. J.M. Schumacher

536 Gert-Jan Otten, Peter Borm, Ton Storcken, Stef Tijs
Effectivity functions and associated claim game correspondences
Refereed by Prof.dr. P.H.M. Ruys

537 Jack P.C. Kleijnen, Gustav A. Alink
Validation of simulation models: mine-hunting case-study
Refereed by Prof.dr.ir. C.A.T. Takkenberg

538 V. Feltkamp and A. van den Nouweland
Controlled Communication Networks
Refereed by Prof.dr. S.H. Tijs

539 A. van Schaik
Productivity, Labour Force Participation and the Solow Growth Model
Refereed by Prof.dr. Th.C.M.J. van de Klundert

540 J.J.G. Lemmen and S.C.W. Eijffinger
The Degree of Financial Integration in the European Community
Refereed by Prof.dr. A.B.T.M. van Schaik

541 J. Bell, P.K. Jagersma
Internationale Joint Ventures
Refereed by Prof.dr. H.G. Barkema

542 Jack P.C. Kleijnen
Verification and validation of simulation models
Refereed by Prof.dr.ir. C.A.T. Takkenberg

543 Gert Nieuwenhuis
Uniform Approximations of the Stationary and Palm Distributions of Marked Point Processes
Refereed by Prof.dr. B.B. van der Genugten

544 R. Heuts, P. Nederstigt, W. Roebroek, W. Selen
Multi-Product Cycling with Packaging in the Process Industry
Refereed by Prof.dr. F.A. van der Duyn Schouten

545 J.C. Engwerda
Calculation of an approximate solution of the infinite time-varying LQ-problem
Refereed by Prof.dr. J.M. Schumacher

546 Raymond H.J.M. Gradus and Peter M. Kort
On time-inconsistency and pollution control: a macroeconomic approach
Refereed by Prof.dr. A.J. de Zeeuw

547 Drs. Dolph Cantrijn en Dr. Rezaul Kabir
De Invloed van de Invoering van Preferente Beschermingsaandelen op Aandelenkoersen van Nederlandse Beursgenoteerde Ondernemingen
Refereed by Prof.dr. P.W. Moerland

548 Sylvester Eijffinger and Eric Schaling
Central bank independence: criteria and indices
Refereed by Prof.dr. J.J. Sijben

549 Drs. A. Schmeits
Geïntegreerde investerings- en financieringsbeslissingen; Implicaties voor Capital Budgeting
Refereed by Prof.dr. P.W. Moerland

550 Peter M. Kort
Standards versus standards: the effects of different pollution restrictions on the firm's dynamic investment policy
Refereed by Prof.dr. F.A. van der Duyn Schouten

551 Niels G. Noorderhaven, Bart Nooteboom and Johannes Berger
Temporal, cognitive and behavioral dimensions of transaction costs; to an understanding of hybrid vertical inter-firm relations
Refereed by Prof.dr. S.W. Douma

552 Ton Storcken and Harrie de Swart
Towards an axiomatization of orderings
Refereed by Prof.dr. P.H.M. Ruys

553 J.H.J. Roemen
The derivation of a long term milk supply model from an optimization model
Refereed by Prof.dr. F.A. van der Duyn Schouten

554 Geert J. Almekinders and Sylvester C.W. Eijffinger
Daily Bundesbank and Federal Reserve Intervention and the Conditional Variance Tale in DM/$-Returns
Refereed by Prof.dr. A.B.T.M. van Schaik

555 Dr. M. Hetebrij, Drs. B.F.L. Jonker, Prof.dr. W.H.J. de Freytas
"Tussen achterstand en voorsprong" de scholings- en personeelsvoorzieningsproblematiek van bedrijven in de procesindustrie
Refereed by Prof.dr. Th.M.M. Verhallen

556 Ton Geerts
Regularity and singularity in linear-quadratic control subject to implicit continuous-time systems
Communicated by Prof.dr. J. Schumacher

557 Ton Geerts
Invariant subspaces and invertibility properties for singular systems: the general case
Communicated by Prof.dr. J. Schumacher

558 Ton Geerts
Solvability conditions, consistency and weak consistency for linear differential-algebraic equations and time-invariant singular systems: the general case
Communicated by Prof.dr. J. Schumacher

559 C. Fricker and M.R. Jaïbi
Monotonicity and stability of periodic polling models
Communicated by Prof.dr.ir. O.J. Boxma

560 Ton Geerts
Free end-point linear-quadratic control subject to implicit continuous-time systems: necessary and sufficient conditions for solvability
Communicated by Prof.dr. J. Schumacher

561 Paul G.H. Mulder and Anton L. Hempenius
Expected Utility of Life Time in the Presence of a Chronic Noncommunicable Disease State
Communicated by Prof.dr. B.B. van der Genugten

562 Jan van der Leeuw
The covariance matrix of ARMA-errors in closed form
Communicated by Dr. H.H. Tigelaar

563 J.P.C. Blanc and R.D. van der Mei
Optimization of polling systems with Bernoulli schedules
Communicated by Prof.dr.ir. O.J. Boxma

564 B.B. van der Genugten
Density of the least squares estimator in the multivariate linear model with arbitrarily normal variables
Communicated by Prof.dr. M.H.C. Paardekooper

565 René van den Brink, Robert P. Gilles
Measuring Domination in Directed Graphs
Communicated by Prof.dr. P.H.M. Ruys

566 Harry G. Barkema
The significance of work incentives from bonuses: some new evidence
Communicated by Dr. Th.E. Nijman

567 Rob de Groof and Martin van Tuijl
Commercial integration and fiscal policy in interdependent, financially integrated two-sector economies with real and nominal wage rigidity.
Communicated by Prof.dr. A.L. Bovenberg

568 F.A. van der Duyn Schouten, M.J.G. van Eijs, R.M.J. Heuts
The value of information in a fixed order quantity inventory system
Communicated by Prof.dr. A.J.J. Talman

569 E.N. Kertzman
Begrotingsnormering en EMU
Communicated by Prof.dr. J.W. van der Dussen

570 A. van den Elzen, D. Talman
Finding a Nash-equilibrium in noncooperative N-person games by solving a sequence of linear stationary point problems
Communicated by Prof.dr. S.H. Tijs